

Relative Navigation for Satellite Formation Flying
based on Radio Frequency Metrology

Rui SUN

Relative Navigation for Satellite Formation Flying based on Radio Frequency Metrology

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. ir. K.C.A.M. Luyben,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op donderdag 9 oktober 2014 om 12:30 uur

door

Rui SUN
Information and Communication Engineering,
M.Sc., Harbin Institute of Technology,
Harbin, China

geboren te Harbin, China

Dit proefschrift is goedgekeurd door de promotor:
Prof. dr. E.K.A. Gill

Copromotor:
Dr. J. Guo

Samenstelling promotiecommissie:

Rector Magnificus	voorzitter
Prof. dr. E.K.A. Gill	Technische Universiteit Delft, promotor
Dr. J. Guo	Technische Universiteit Delft, copromotor
Prof. ir. B.A.C. Ambrosius	Technische Universiteit Delft
Prof. ir. P. Hoogeboom	Technische Universiteit Delft
Prof. dr. G.W. Hein	European Space Agency, Universität der Bundeswehr München
Prof. dr. P. Axelrad	University of Colorado
Dr. ir. A.A. Verhagen	Technische Universiteit Delft

Rui SUN
Group of Space System Engineering,
Department of Space Engineering,
Aerospace Engineering,
Delft University of Technology

ISBN: 978-94-6186-358-4

Copyright ©2014 by Rui SUN

All rights reserved. No part of the material protected by this copyright may be reproduced, or utilized in any other form or by any means, electronic or mechanical, including photocopying, recording or by any other information storage and retrieval system, without the prior permission of the author.

Printed by WÖHRMANN PRINT SERVICE B.V., Zutphen, the Netherlands

Summary

To increase mission return, utilizing two or more spacecraft instead of one may sometimes be superior. This is especially true when a large spaceborne instrument needs to be created through larger and configurable baselines, such as telescopes and interferometers. However, coordinating the alignment of the individual components of such a spaceborne instrument on separate spacecraft (involving the estimation and control of baselines) will require a high level of accuracy for relative navigation and control. The increasing demand of such science missions or challenges on complex functions such as rendezvous and docking, calls for high accuracy levels of ranging at centimeter or even millimeter levels.

The objective of this research is to develop a relative navigation system based on radio-frequency (RF) metrology for future formation flying missions. This RF-based system inherits Global Navigation Satellite System (GNSS) technologies through transmission and reception of locally generated GNSS-like pseudo random noise (PRN) ranging codes and carrier phases via inter-satellite links. This enables operation, e.g., in high Earth orbits where GNSS constellations are poorly visible. The RF-based navigation system is designed to comprise of one transmitter, one receiver and several antennas in order to enable the coarse-mode inter-satellite distance estimation (meter level) based on pseudorange measurements and fine-mode distance (centimeter level) and line-of-sight (LOS) estimation (sub-degree level) based on carrier phases in addition to pseudorange.

A benchmarking system, called the Formation Flying Radio Frequency (FFRF) sensor, has been successfully flown and demonstrated on PRISMA mission. This research improves the performance of FFRF with respect to the technologies

- 1) to deal with errors and uncertainties, especially multipath;
- 2) to perform an unaided, fast and reliable carrier phase integer ambiguity resolution (IAR); and
- 3) to share channels among multiple spacecraft.

Multipath In space applications, receivers on space vehicles may suffer from very-short-delay multipath (< 4 m) that is reflected from the vehicle itself or from other vehicles during the operations of rendezvous and docking.

The thesis proposes a novel method, named as “Multipath Envelope Curve Fitting”, to mitigate very-short-delay-multipath on pseudorange measurements by approximately 50%. It also exhibits a promising performance for medium or large delayed multipath as compared to state-of-the-art methods. The method is based on the fact that the signal strength information, reported by early or late correlators inside the receiver, has an in-phase correlation with the pseudorange multipath error. By linearly combining multiple signal strength estimators from multiple correlators,

the pseudorange multipath error has been accurately estimated. The weights for the linear combination were obtained by curve fitting based on the least-squares adjustment. A simple implementation strategy was also proposed that enables a receiver-internal multipath estimation process operated in conjunction with the tracking loop with a minimal additional computational overhead.

Compared to the pseudorange multipath, the carrier phase multipath has more significant impacts on high precision navigation, especially when it is coupled with the carrier phase IAR. By making use of the signal to noise ratio (SNR) data of multiple antennas, this thesis proposes a novel cascaded extended Kalman filter (EKF) to mitigate carrier phase multipath. This method accelerates the IAR process significantly and guarantees an achievement of sub-degree LOS accuracy. Both real-valued and complex-valued EKF are proposed and evaluated. The complex-valued EKF has been found to be insensitive to poorly defined initial conditions, when the real-valued EKF has difficulties converging. Moreover, the complex-valued EKF has shown better convergence properties for SNR observations with a large amount of noise.

Integer Ambiguity Resolution The second challenge of this research is to perform an unaided, fast and reliable carrier phase IAR. Single-epoch IAR algorithms are proposed in this thesis by making use of a nonlinear quadratic LOS length constraint and taking advantages of antenna arrays. Two methods, namely, the validation method and the subset ambiguity bounding method, are proposed. They replace the equality quadratic constraint by inequality boundaries such that the well-known Least-squares AMBiguity Decorrelation Adjustment (LAMBDA) integer ambiguity resolution process is implemented within a pre-defined threshold to increase the integer search fidelity. Numerical simulations and field tests demonstrated that both the validation method and the subset ambiguity bounding method provided remarkable improvements with up to 80% higher success rates than the original LAMBDA method based on single-epoch measurements. The validation method showed a slightly better performance than the subset ambiguity bounding method as they differ in utilizing all-ambiguity-set and subset-ambiguity, respectively. Better IAR robustness against multipath can also be observed as compared to the original LAMBDA method. An Ambiguity Dilution of Precision (ADOP) measure under the LOS constraint is derived, which is an easy-to-use and insightful indicator of the ambiguity resolution capability. A rule-of-thumb for the pre-defined threshold has also been derived in the closed-form expression, providing guidance on how to choose boundaries according to the noise level and antenna geometry.

Multiple Access Technology Enabling multiple access capability is of critical importance for future missions with four or more spacecraft. The Code Division Multiple Access (CDMA) technology is recommended to be used in combination with a flexible roles rotating topology in this research. This allows coping with time-critical relative navigation requirements and enables flexible operations during various mission phases. Through realistic formation case studies, the limitation of CDMA was extensively investigated in terms of the multiple access interference (MAI) which could result in a ranging error of several meters and is highly dependent on the Doppler offset. Recommendations are given in this thesis to reduce corresponding MAI errors.

Samenvatting

Voor sommige missies heeft het voordelen om twee of meerdere satellieten in te zetten, omdat dit soms meer bruikbare data oplevert. Dit geldt in het bijzonder wanneer men een groot orbitaal instrument wil maken door grote en aanpasbare zogenaamde “baselines” te creëren, bijvoorbeeld voor telescopen en interferometers. Het coördineren van de elementen van zulke instrumenten echter vereist een hoge mate van nauwkeurigheid om relatieve navigatie, en daarmee controle over de onderlinge afstanden mogelijk te maken. De groeiende vraag naar zulke wetenschappelijke missies, alsmede de uitdagingen voor complexe bewegingen in de ruimte, zoals bij rendezvous en docking, roept om nauwkeurige manieren om onderlinge afstanden te bepalen met nauwkeurigheden van centimeters of zelfs millimeters.

Het doel van dit onderzoek was dan ook om een relatief navigatiesysteem te ontwikkelen, gebaseerd op hoogfrequente (HF) metrologie voor toekomstige formatievluchten in de ruimte. Deze HF-gebaseerde methode maakt gebruik van technologieën oorspronkelijk ontwikkeld voor satelliet-navigatie voor de verschillende Global Navigation Satellite Systems (GNSS), door het verzenden en ontvangen van lokaal gecreëerde GNSS-achtige pseudo-willekeurige ruis afstandsbepalings-codes en zelfs draaggolf-fases over een inter-satelliet-verbinding. Dit staat het systeem toe om bijvoorbeeld in zeer hoge banen om de aarde, waar er een slechte zichtbaarheid is van de verschillende GNSS constellaties te opereren. Het HF-systeem is ontworpen rondom een enkele ontvanger en een enkele zender, aangevuld met verschillende antennes om grove afstandsbepalings (met nauwkeurigheden in de orde van een meter) op basis van pseudo-range metingen te verrichten, alsmede precisie-metingen (op centimeter-niveau) te verrichten. Ook Line-of-sight (LOS) schattingen op basis van de draaggolf-fases zijn mogelijk met een nauwkeurigheid van minder dan 1° .

Een eerste testsysteem, genaamd de Formation Flying Radio Frequency (FFRF) sensor heeft zijn nut inmiddels bewezen op de PRISMA missie. Dit onderzoek verbetert de prestaties van dit FFRF-systeem in de volgende aspecten:

- 1) Rekening houdend met fouten en onzekerheden, voornamelijk veroorzaakt door zogenaamde multipath-effecten;
- 2) Het uitvoeren van autonome, snelle en betrouwbare draaggolf-fase Integer Ambiguity Resolution (IAR); en
- 3) Het delen van kanalen over meerdere satellieten.

Multipath, oftewel het verschijnsel dat optreedt wanneer radiogolven op verschillende oppervlakken op de satelliet weerkaatsen, zorgt op satellieten voornamelijk voor zeer kleine afwijkingen (< 4 meter) door het reflecteren hetzij op de ontvangende satelliet, of op de satelliet die probeert aan te meren.

Deze dissertatie stelt een innovatieve methode voor, genaamd “Multipath Envelope Curve Fitting” om de fouten veroorzaakt door dit verschijnsel met zo’n 50%

te verminderen. Deze methode toont ook een verbetering voor middelmatige en lange-afstands multipath effecten in vergelijking met de nieuwste methodes. Deze methode is gebaseerd op het feit dat de signaalsterkteinformatie die wordt gemeld door de vroege-of laat correlatoren in de ontvanger een in-fase correlatie heeft met de multipath-fout. Door het lineair combineren van de verschillende signaalsterkte-schattingen van de verschillende correlatoren kan de multipath-fout precies geschat worden. De weegfactoren voor de lineaire combinatie zijn verkregen door een trendlijn door de kleinste-kwadraten-aanpassing te passen. Een eenvoudige strategie om deze methode te implementeren is ook voorgesteld, hetgeen een ontvanger in staat stelt om intern de multipath-effecten af te schatten, simultaan met de tracking-loop, met toch weinig additionele processorbelasting.

In vergelijking met de pseudo-range multipath effecten hebben de draaggolf-fase multipath-effecten een grotere invloed op de precisie van de navigatieoplossing, vooral wanneer deze effecten gekoppeld zijn aan de IAR van de carrier fase. In deze thesis wordt een methode voorgesteld die gebruik maakt van informatie over de signaalsterkte over verschillende antennes, op basis van een nieuwe zogenaamde cascaded extended Kalman filter (EKF) om multipath-effecten op de draaggolf-fase te mitigeren. Deze methode versnelt het IAR proces aanzienlijk en garandeert het bereiken van een Line-Of-Sight precisie van minder dan 1 graad. Zowel reële als complexe EKF's worden aangedragen en geanalyseerd. De complexe EKF is aantoonbaar ongevoeliger voor slecht geformuleerde beginwaarden, in vergelijking met de reële EKF, die dan moeite heeft om te convergeren. Ook in omstandigheden met veel achtergrondruis convergeert de complexe EKF beter.

Integer Ambiguity Resolution: De tweede grote uitdaging tijdens dit onderzoek was om een snelle, autonome en betrouwbare draaggolf-fase IAR uit te voeren. Enkele epoch IAR algoritmes worden aangedragen in deze thesis door gebruik te maken van een niet-lineaire kwadratische LOS lengte-beperking, die terwijl gebruik maakt van de voordelen van antenne-arrays. Twee methodieken, met name de validatiemethode alsmede de subset ambiguity bounding methode worden voorgedragen. Ze vervangen de kwadratische gelijkheids-beperking door ongelijkheidsgrenzen opdat de bekende kleinste-kwadraten ambiguïteit-decorrelatie aanpassing-oplossingsproces (Engels: LAMBDA, of Least-squares AMBiguity Decorrelation Adjustment) wordt geïmplementeerd met een vooraf gedefinieerde drempel om de precisie te verbeteren. In numerieke simulaties en proeven in het veld werd aangetoond dat zowel de validatiemethode als de subset ambiguity bounding-methode de slagingspercentages met tot wel 80% verbeterden in vergelijking met de originele LAMBDA-methode, gebaseerd om enkele epoch-metingen. De validatiemethode toonde aan iets beter te presteren dan de subset ambiguity bounding-methode doordat de validatiemethode de gehele ambiguïteit-set gebruikt, in tegenstelling tot een sub-set. Een verbeterde IAR tolerantie tegen multipath-effecten werd ook geobserveerd. Er wordt ook een Ambiguity-Dilution of Precision (ADOP) maat onder een LOS-beperking afgeleid, die dienst doet als een gemakkelijk inzetbare en heldere indicator is van de ambiguïteits-oplossings-prestaties van een methode. De vuistregel voor een vooraf gedefinieerde drempel werd ook afgeleid in een gesloten vergelijking, hetgeen aangeeft hoe begrenzingen gekozen dienen te worden gegeven een bepaald ruisniveau en een bepaalde antenne-geometrie.

Multiple Access Technology: Het mogelijk maken van simultane toegang door meerdere satellieten wordt gezien als een belangrijke stap in de richting van toekomstige

stige satellietmissies met vier of meer satellieten. De Code-Division Multiple Access (CDMA) technologie hier gebruikt wordt daarom ook aangeraden om te gebruiken in combinatie met een roterende topologie met een flexibele rolverdeling. Dit staat het gebruik ervan in situaties waar tijd-kritische relatieve-navigatie van belang is toe en staat verder ook verschillende rollen gedurende verschillende missie-fases toe. Door enkele realistische casussen te bestuderen op het gebied van de limiteringen opgelegd door CDMA, werden de effecten veroorzaakt door Multiple Access Interferentie (MAI) bestudeerd. Deze kunnen zorgen voor afstandsmetingsfouten in de orde van verscheidene meters, en deze blijkt sterk afhankelijk te zijn van de Dopler bias. Enkele aanbevelingen worden in deze thesis gegeven omtrent het verminderen van zulke MAI-geïnduceerde fouten.

Acknowledgement

First of all, I would love to express my sincere gratefulness to my promoter Prof. Dr. Eberhard Gill. Thank you for offering me such a great opportunity to work in the Space System Engineering group at Delft University of Technology. Your invaluable guidance, accurate and critical feedback greatly improve the quality of the research. I cannot thank you enough for your valuable advice, not only about academic performance but also the way of working. My special thanks extend to my supervisor Dr. Jian Guo. You were always supportive and caring. It has been a great pleasure and a learning experience working with you. I appreciate also very much that you and your family always welcomed me for the traditional Chinese festivals, which makes my life in the Netherlands like home.

Secondly, my appreciation goes to Prof. Gérard Lachapelle and Associate Prof. Kyle O’Keefe, who invited me to the Positioning, Location and Navigation group at University of Calgary in Canada. Thanks a lot, Kyle, for helping me arranging field experiments and providing me with valuable advice on our joint publications. My gratefulness also deeply goes to Mr. Marios Smyrnaiois of Leibniz Universität Hannover, who was kind enough to provide the antenna model for the verification of my work.

I would also like to thank all my colleagues in the SSE group of provided me an admirable environment to not only explore science and technology, but also to explore my mind and personality. The first two names I would like to mention are Prem Sundaramoorthy and Jing Chu. You guys are the best office mates. We shared a lot of insightful ideas about research and social life in the office. To my other colleagues, Arash Noroozi, Daan Maessen, Steven Engelen, Adolfo Chaves Jimenez, Rouzbeh Amini, Barry Zandbergen, Chris Verhoeven, Daniel Choukroun, Angelo Cervone, Jasper Bouwmeester, Hans Kuiper, Luca Guadagni and Nuno Baltazar dos Santos, I am proud of working with all of you, and will keep the good memory of interesting conversation in the coffee corner and the space bar. Needless to say, my appreciation goes to our secretaries, Debby van der Sande and Relly van Wingaarden who showed your smiles every time I needed help in administrative tasks. Thank you both for such great assistance.

Also, the committee members of my PhD defence are greatly acknowledged. I thank you for devoting your precious time to read and provide feedback to my thesis. I am looking forward to welcoming you at the defence ceremony.

Special thanks goes to Steven Engelen, who helped me in translating the summary and propositions of this thesis from English to Dutch. Appreciation is also given to Lauri Koponen, who helped me in revising the cover picture of this thesis.

To my dear friends in Delft, Jinglang Feng, Xuedong Zhang, Bin Zhao, Qikai Zhuang, Yanqing Hou, Davide Imperato, and my friends in Calgary, Bei Huang, Jingjing Dou, Peng Xie, Yihe Li, Erin Kahr, Da Wang, I thank you all for making

my PhD period more colourful and lively. For the ups and downs during the PhD period, I can always count on you guys.

Last but not least, my whole heart belongs to my family, my parents Limei Sun, Xiuqing Sun, my husband Zongyu Liu, and my parents in-law Zhilin Liu and Qinghua Zhao. Thanks for always standing beside me with your unconditional love, encouragement and understanding during all these years. Finally I have the chance to thank you in the preface of my book. I love you all.

Rui Sun

Delft, May 6, 2014

Abbreviations

ADC	Analog to Digital Converter
ADOP	Ambiguity Dilution of Precision
AFF	Autonomous Formation Flying
AGC	Automatic Gain Control
APME	A-Posteriori Multipath Estimation
BOC	Binary offset carrier
BPF	Bandpass Filter
BPSK	Binary Phase Shift Keying
BPSK-R	Binary Phase Shift Keying signalling with Rectangular chips
C/A	Coarse/Acquisition
CDMA	Code Division Multiple Access
DLL	Delay Lock Loop
ECKF	Extended Complex-valued Kalman Filter
EKF	Extended Kalman Filter
ELS	Early Late Slope
ESA	European Space Agency
FAS	Formation Acquisition Sensor
FDMA	Frequency Division Multiple Access
FFRF	Formation Flying Radio Frequency
FFT	Fast Fourier Transform
FMCW	Frequency Modulated Continuous Wave
GEO	Geostationary Orbit
GNC	Guidance, Navigation and Control
GNSS	Global Navigational Satellite System
GPS	Global Positioning System
GRACE	Gravity Recovery And Climate Experiment mission
HEO	Highly Elliptical Orbit
IAR	Integer Ambiguity Resolution
IF	Intermediate Frequency
ILS	Integer Least Squares
IMU	Inertial Measurement Unit
INS	Inertial Navigation System
IRAS	Inter-satellite Ranging and Alarm System
ITU	International Telecommunication Union
ISL	Inter-Satellite Link
ISS	International Space Station
KF	Kalman Filter

KRS	K/Ka-band Ranging System
LAMBDA	Least-squares AMBIGUITY Decorrelation Adjustment
LEO	Low Earth Orbit
LHCP	Left-hand Circular Polarization
LNA	Low Noise Amplifier
LOS	Line of Sight
MAI	Multiple Access Interference
MEDLL	Multipath Estimation Delay Lock Loop
MMS	Magnetospheric Multiscale
MMT	Multipath Mitigation Technique
NASA	National Aeronautics and Space Administration
NCO	Numerically Controlled Oscillator
nEML	narrow Early-Minus-Late correlator
OEXO	Oven-controlled Crystal Oscillator
PA	Power Amplifier
PLL	Phase Lock Loop
PPS	Pulse Per Second
PRISMA	PRecursore IperSpettrale della Missione Applicativa
PRN	Pseudo Random Noise
PSD	Power Spectrum Density
QPSK	Quadrature Phase Shift Keying
RF	Radio Frequency
RHCP	Right-hand Circular Polarization
Rx	Receiver
SD	Single Differenced
SNR	Signal to Noise Ratio
SPARCE	SPaceborne Active Ranging and Communication System
SSE	Space Systems Engineering
TDMA	Time Division Multiple Access
TPF	Terrestrial Planet Finder
TT&C	Telemetry, tracking, and command
TU Delft	Technische Universiteit Delft
Tx	Transmitter
UWB	Ultra-Wideband
VBS	Vision-based Sensor

Symbols and Notations

Mathematical and Statistical Notation

$\sum(\cdot)$	summation
\otimes	Kronecker product
$\Delta(\cdot)$	single differencing operator
$\ \cdot\ $	norm of a vector
$\ \cdot\ _{\mathbf{Q}}^2$	weighted squared norm $(\cdot)\mathbf{Q}^{-1}(\cdot)$
$E(\cdot)$	mathematical expectation operator
$D(\cdot)$	mathematical dispersion operator
$tr(\cdot)$	trace, sum of elements on the main diagonal of a square matrix
$diag(\cdot)$	Diagonal matrix
$sgn(\cdot)$	signum function (1 if the argument is positive, -1 if the argument is negative)
\mathbb{R}^p	real space of dimension p
\mathbb{Z}^p	integer space of dimension p

Transmitter and Receiver Symbols

$c(t)$	PRN sequence
T_c	chip period
f_c	chipping rate, the reciprocal of the chip period
$R(\tau)$	auto-correlation with time shift of τ
$R_c(\tau, F)$	generalized auto-correlation with time and frequency shifts of τ and F
T_{sc}	half-period of a square wave
f_{sc}	subcarrier frequency of a BOC code
T_s	sampling period
f_s	sampling frequency
$G(f)$	power spectrum density
f_d	Doppler frequency
P_{fa}	false alarm probability in signal acquisition
P_d	detection probability in signal acquisition
T_h	Threshold in signal acquisition

$\delta\tau, \delta f, \delta\phi$	code, frequency and phase misalignment in tracking
$D(\delta\tau)$	DLL discriminator function
$D(\delta\phi)$	PLL discriminator function
I_E, I_P, I_L	early, prompt and late correlator
d	early-late spacing
Δd	spacing between adjacent correlators
B_L	tracking noise bandwidth
T	integration time
β_r	front-end bandwidth
C/N_0	carrier to noise ratio
A_0, A_1	signal amplitude for line-of-sight signal and multipath
τ_1, ψ_1	multipath delay and multipath phase
$\delta\hat{\tau}_{mp}$	pseudorange multipath estimation
$\delta\hat{\phi}_{mp}$	carrier phase multipath estimation

Estimation Algorithm Symbols

ρ, ϕ	pseudorange and carrier phase observation
$\mathbf{P}, \mathbf{\Phi}$	column vectors containing pseudorange and carrier phase observations
σ_ρ, σ_ϕ	code and phase thermal noise
lb	instrumental delays, mainly including the line bias
I, T	ionospheric and tropospheric delays
dt_r, dt_s	receiver r and transmitter s clock errors
$\theta_r(t_0), \theta_s(t_0)$	initial phases of the generated replica carrier signal and the transmitter carrier signal
el	elevation
az	azimuth
\mathbf{a}	integer ambiguities
$\mathbf{Q}_{\hat{\mathbf{x}}\hat{\mathbf{x}}}$	covariance matrix of $\hat{\mathbf{x}}$
$\hat{\mathbf{x}}$	float solution of \mathbf{x}
$\tilde{\mathbf{x}}$	fixed solution of \mathbf{x} (after ambiguities are fixed)
$\tilde{\mathbf{x}}(\mathbf{a})$	\mathbf{x} conditioned on \mathbf{a}
e_n	$n \times 1$ column vector with all elements equal to one
λ_f	signal wavelength at frequency f
\mathbf{g}_{ij}	antenna baseline vector between reference antenna j and auxiliary antenna i
\mathbf{G}	antenna baseline coordinate matrix for n baselines, $\mathbf{G} = [\mathbf{g}_{1j}^T; \mathbf{g}_{2j}^T; \dots; \mathbf{g}_{nj}^T]$
P_{LB}	bootstrapped lower bound IAR success rate

P_{ADOP}	approximated value of IAR success rate
δl	line-of-sight length constraining threshold
$ADOP_{\infty}$	unconstrained ambiguity dilution of precision
$ADOP_{\delta l}$	constrained ambiguity dilution of precision

Contents

Summary	ii
Samenvatting	iv
Acknowledgement	vii
List of Abbreviations	x
List of Symbols and Notations	xi
1 Introduction	1
1.1 Background and motivation	1
1.1.1 Formation flying	1
1.1.2 Formation flying metrology	5
1.1.3 Formation flying radio frequency metrology	6
1.2 Research questions, objectives and methodologies	9
1.3 Structure of the thesis	11
2 RF-based Relative Navigation System Design and Analysis	13
2.1 RF-based relative navigation system design	13
2.1.1 Architecture	13
2.1.2 Frequency allocation	16
2.1.3 PRN code structure	18
2.2 Transmitter architecture	21
2.3 Receiver architecture and analysis	22
2.3.1 Signal conditioning in the front-end	22
2.3.2 Acquisition	24
2.3.3 Tracking	27
2.3.4 Lower bound of code tracking accuracy	35
2.3.5 Multipath effects	36
2.3.6 Evaluation of BPSK-R and BOC codes	40
2.4 Code and phase observations	42
2.4.1 Undifferenced observation model	42
2.4.2 Single-differenced model between receivers/antennas	43
2.4.3 Bias analysis	44
2.5 Relative navigation model	46
2.5.1 Line-of-sight estimation model	46
2.5.2 Inter-satellite distance estimation model	48
2.6 Chapter summary	50

3	Line-of-sight Estimation	53
3.1	Problem statement and existing methods	53
3.1.1	Integer ambiguity resolution	53
3.1.2	Benchmarking solution: LAMBDA	57
3.1.3	Constrained LAMBDA	61
3.2	Theory of LOS estimation and associated constrained LAMBDA	65
3.2.1	Single-epoch LOS estimation model	65
3.2.2	Bias calibration	67
3.2.3	Constraint on the float solution	67
3.2.4	Constraint on the integer mapping process	70
3.2.5	Threshold	72
3.3	Antenna geometry aspects	77
3.3.1	LOS dilution of precision	77
3.3.2	Constrained ambiguity dilution of precision	78
3.3.3	Antenna geometry	80
3.4	Verification	82
3.4.1	Numerical simulations	82
3.4.2	Field tests	83
3.5	Chapter summary	91
4	Code Multipath Effects and Mitigation Method	93
4.1	Problem statements and existing methods	93
4.1.1	Multipath in space	94
4.1.2	Multipath mitigation method categorisation	94
4.1.3	Characterizing multipath envelope	96
4.1.4	Several receiver-internal multipath mitigation techniques	99
4.2	Theory of the signal strength-based multipath envelope curve fitting	102
4.2.1	Characterizing the relation between the multipath error and the signal strength	102
4.2.2	Principle of the multipath envelope curve fitting	105
4.2.3	Variance	108
4.2.4	Discussions on the amount and locations of correlators	111
4.2.5	Applications on the BPSK-R code	113
4.2.6	Implementation	117
4.2.7	Limitations	119
4.3	Verification	119
4.3.1	Software-defined signal simulator and receiver	119
4.3.2	Simulation settings	120
4.3.3	Performance	122
4.4	Chapter summary	124
5	Carrier Phase Multipath Effects and Mitigation Methods	125
5.1	Problem statement and existing Methods	125
5.1.1	Charactering phase multipath and SNR	126
5.1.2	SNR based multipath estimation	127
5.1.3	Multi-antenna based multipath estimation	129
5.1.4	Multipath mapping	130
5.2	Theory of multi-antenna based multipath estimation on the fly	130
5.2.1	Kalman filter and extended Kalman filter	130
5.2.2	Model of the satellite-antenna-reflector geometry	134

5.2.3	Multipath correction procedure	136
5.2.4	Noise filtering for ratios of SNRs	139
5.2.5	Multipath correction before fixing integer ambiguities	141
5.2.6	Combined multipath correction and LOS estimation after fixing integer ambiguities	145
5.3	Performance evaluation	148
5.3.1	Sensitivity to initial conditions	149
5.3.2	Tolerance to large noise observations	152
5.3.3	Robustness in multi-reflection conditions	153
5.4	Multipath effects on the integer ambiguity resolution	153
5.4.1	IAR acceleration	154
5.4.2	Multipath robustness in single-epoch IAR	155
5.5	Chapter summary	158
6	Network Architecture	161
6.1	Dedicated network architecture requirements	161
6.1.1	Time-critical requirement	162
6.1.2	Flexible operations across all mission phases	163
6.2	Candidates for network architectures	164
6.2.1	TDMA with deterministic time slot	167
6.2.2	Roles rotating CDMA with flexible time slot	168
6.3	CDMA limitations: multiple access interference and near-far problem	168
6.3.1	Cross correlation without Doppler effect	168
6.3.2	Cross correlation at high Doppler offset	171
6.3.3	Near-far problem at Doppler crossover	172
6.4	Case-studies	173
6.4.1	Case-study set-up	173
6.4.2	Circular low earth orbit formation scenario	174
6.4.3	Highly elliptical orbit scenario: MMS mission	177
6.4.4	Case-study summary	179
6.5	Chapter summary	181
7	Conclusions and Outlook	183
7.1	Summary	183
7.2	Conclusions	184
7.3	Outlook	187
A	Covariance Matrices of $\mathbf{Q}_{\hat{a}\hat{a}}$, $\mathbf{Q}_{\hat{a}\hat{a}}$, $\mathbf{Q}_{\hat{x}(\mathbf{a})\hat{x}(\mathbf{a})}$ and $\mathbf{Q}_{\hat{x}(\mathbf{a}_p)\hat{x}(\mathbf{a}_p)}$	189
B	The Determinant of $\mathbf{Q}_{\hat{a}\hat{a}}$	193
C	Correlation Coefficient between Early/Late Correlators	195
	Bibliography	196
	List of Author's Publications	208
	Curriculum Vitae	210

Chapter 1

Introduction

As mission requirements advance, satellite formation flying with multiple satellites in a coordinated manner has become of greater importance. This chapter introduces the definition of formation flying, overviews different enabling metrologies for variable types of formation flying, and investigates the increasing needs for a radio-frequency (RF) based metrology. The thesis will then focus on key technologies of developing RF-based system and associated algorithms.

1.1 Background and motivation

To increase mission return, utilizing two or more small satellites in a coordinated formation can be beneficial or even necessary compared to a single one. This is especially true for creating a large spaceborne scientific instrument such as telescope and interferometer through large and configurable baselines between/among satellites. The motivation of this thesis is to investigate advanced technologies to enable the alignment of baselines, e.g., estimating and controlling baselines, for enhanced scientific research and experiments.

1.1.1 Formation flying

Satellite formation flying allows for multiple satellites working together to accomplish the objective of one larger, usually more expensive, satellite. Formation flying is a subset of a more general category that is defined as distributed spacecraft systems (DSS), which include also, e.g., constellation, cluster of satellites in a less coordinated manner. Across the formation flying community there exists a wide range of definitions for formation flying and related terms. The most distinct differences in definition occur between the science (or instrument/sensor) community and the engineering (or technology) community (Leitner, 2004):

Engineering definition: the tracking or maintenance of a desired separation between/among two or more satellites;

Science definition: the collective use of multiple satellites to perform the function of a single, large, virtual instrument.

From an engineer point of view, coordinating smaller satellites in a formation

has many benefits over a single satellite including simpler designs, faster build times and cheaper replacement creating higher redundancy. From a scientist perspective, formation flying allows for viewing targets from multiple points and/or at multiple times (Wikipedia, 2013), and offers the possibility for unprecedented high resolution by creating a large spaceborne instrument such as telescope and interferometer through the distribution of functions and payloads among fleets of coordinated small satellites. The science return can be dramatically enhanced through observations made with larger and configurable baselines (D'Amico, 2010). However, this requires a challenging technology for coordinating the alignment of baselines, e.g., estimating and controlling baselines with a high navigation and control accuracy requirement.

Four main development lines have been identified for the current proposed or flown formation flying missions (Grelier et al., 2009; D'Amico, 2010) in the science community, as listed in Table 1.1. These mission concepts drive an increasing level of complexity for engineers, mainly dictated by the payload metrology and actuation needs, and require a high level of accuracy of relative navigation and control.

Earth Observation: These missions, in Low Earth Orbit (LEO), respond to the demand for highly accurate Earth models on a global space and time scale. Key examples are the SAR interferometry missions, e.g., TerraSAR-X/TanDEM-X (Krieger et al., 2007) and gravity recovery missions, e.g., GRACE (the gravity recovery and climate experiment) mission (Bertiger et al., 2002). The TerraSAR-X/TanDEM-X mission consists of two satellites, launched in 2007 and 2010, respectively, to perform a precisely controlled radar interferometer in 500 km altitude with typical baselines of 1 km. The GRACE mission, launched in 2002, was used to make detailed measurements of Earth's gravity field. The mission uses an inter-satellite microwave ranging system to accurately measure changes in the speed and distance between two identical satellites flying in a polar orbit about 220 km apart. This ranging system is sensitive enough to detect separation changes as small as 10 micrometres (Bertiger et al., 2002).

The typical relative orbit control accuracy required for Earth observation formations is relatively coarse (~ 100 m) and may drive the need for real-time onboard relative navigation accuracy at a 1-10 m level. High precision (submillimeter) post-facto reconstruction of the three-dimensional relative motion may be needed for some missions (D'Amico, 2010).

Apart from science missions, demonstration missions in LEO are also important in terms of advanced technology validation. The PRISMA dual satellite mission, launched in 2010, is such a formation flying demonstration mission (Gill et al., 2007; Thevenet and Grelier, 2012) in LEO. Key navigation sensors on PRISMA comprise GPS receivers, formation flying radio frequency sensors (FFRF) and vision-based sensors (VBS) to demonstrate a fully autonomous, robust and accurate formation flying through experiments in autonomous formation flying, homing, rendezvous scenarios as well as close-range proximity and final approach and receding operations. Autonomous formation flying performs on-board guidance, navigation and control tasks without ground intervention in-the-loop (Gill et al., 2007). Full autonomy with real-time relative navigation accuracy at centimeter level has been achieved on PRISMA. The success of PRISMA boosts the autonomous formation flying being utilized for future Earth observation missions, e.g., PostGRACE, PostGOCE, as well as the potentials employing multiple baselines.

Dual Spacecraft Telescopes: These instruments aim at the detailed spectral investigation of sources which are too faint for the current generation of observatories

Table 1.1: Formation flying mission overview (Grellet et al., 2009; D’Amico, 2010)

Applications	Earth Observation (e.g. SAR interferometer and Gravimeter)	Dual Spacecraft Telescope	Multi Spacecraft Telescope	Long-range and RdV
Missions	PRISMA ¹ , GRACE PostGOCE, PostGRACE TerraSAR/TanDEMx	PROBA-3, Simbol-X Xeus, MAX	Darwin, TPF New Millennium	MMS, MagCON, MAXIM NextMars, ATV, MSR CSTS-ISS(LEO)
Orbit	Low Earth Orbit (LEO)	High Earth Orbit Highly Elliptical Orbit (HEO) (or Lagrange point)	Lagrange point (or HEO)	HEO, moon, Mars LEO(ISS) Geostationary Orbit (GEO)
Number of spacecraft	≥ 2	2	≥ 3	≥ 3 or 2 (RdV)
Typical separation	100 m - 1000 km	30-250 m	10-1500 m	Long-range: 100 m-3000 km
Control accuracy	10-100 m	0.1-10 cm	1-100 cm	Long-range: 1 km or larger
Navigation accuracy	1-10 m (1 mm post-facto)	0.1-10 mm	1-100 mm	Long-range: 10-100 m
Navigation technology	GNSS space receivers (or integrated with RF metrology)	(or integrated with optical metrology)	RF metrology (or integrated with GNSS in LEO, or optical technology during operations of docking)	RF metrology (or integrated with GNSS in LEO, or optical technology during operations of docking)

¹ Strictly speaking, PRISMA is not an Earth observation mission, but a demonstration mission for future formation flying technologies. It has been categorized into “Earth observation” column as it flies in LEO and it relies on GNSS and RF metrologies.

like the Chandra X-ray Observatory (Weisskopf et al., 2000) and the XMM-Newton (Jansen et al., 2001). The typical mission profile seeks orbits characterized by a low level of perturbations, stable thermal environment, lack of eclipses, and wide sky visibility (Grelier et al., 2009; D’Amico, 2010). In contrast to the unfavorable LEO environment, optimum conditions are offered by highly elliptical orbits (HEO) or geostationary orbits (GEO).

Distributed telescopes are space systems composed of a detector and a mirror spacecraft flying as a formation during science operations (Rupp et al., 2007). Typical separations aim at focal lengths of the order of 30-100 m. Autonomous formation flying needs are driven by the telescope optical design and should allow uninterrupted science observations. This translates into combined attitude/orbit control systems with required relative navigation accuracies at (sub-)centimeter level (Grelier et al., 2009; D’Amico, 2010).

Relevant examples of dual spacecraft telescopes include e.g. PROBA-3, XEUS and Simbol-X missions. The two satellites in PROBA-3 mission will together form a 150 m long solar coronagraph, one carrying the detector and the other carrying the Sun Occulter disc, to study the Sun’s faint corona close to the solar rim. Both satellites are flying in precise formation in a Highly Elliptical Orbit (HEO) with orbital period in the order of one day. Besides the scientific interest, the PROBA-3 is also a formation flying technology demonstration mission. The demonstration will exercise generic formation configurations valid for multiple types of target missions, including a wide range of formation acquisition and maintenance, formation autonomy, formation reconfigurations, manoeuvres, information and commands exchanged between satellites, etc (Llorente et al., 2013; Landgraf and Mestreau-Garreau, 2013). The launch of PROBA-3 mission is planned in 2017 (Llorente et al., 2013). PROBA-3 is also being designed with focus on verification of the requirements coming from the European Space Agency’s (ESA) XEUS (X-Ray Evolving Universe Spectroscopy) studies, a formation flying X-ray astronomy mission. XEUS consists of a mirror spacecraft that carried a large X-ray telescope with a mirror area of about 5 m². The detector spacecraft in XEUS mission will fly in formation at a distance of approximately 35 m to the telescope, in the focus of the telescope. Maintaining the baseline alignment between the mirror and detector via relative navigation and control is crucially important for the success of such mission. The XEUS has been merged with National Aeronautics and Space Administration’s (NASA) Constellation-X mission and renamed as International X-Ray Observatory (IXO), scheduled to launch in 2020 (Centrella and Reddy, 2011).

Multi-spacecraft Telescope: The third type of application addresses the usage of multiple spacecraft telescopes. Interferometry in the infrared and visible wavelength regions has been identified as key technology to new astrophysics discoveries and to the direct search for terrestrial exoplanets. To that purpose, clusters of three or more units need to fly in millimeter precision close formations with inter-satellite navigation accuracies at the sub-millimeter level (D’Amico, 2010), as listed in Table 1.1. Examples of this type of missions in Europe include the infrared space interferometer DARWIN (Bourga et al., 2002) and in USA include the NASA’s Terrestrial Planet Finder (TPF) (Tien et al., 2004). However, the study of DARWIN mission ended in 2007 with no further activities planned (ESA, 2007), while the TPF mission was also deferred “indefinitely” by NASA in 2007 due to budget constraints (Mullen, 2011).

Long Range and RdV Missions: Finally, long range and rendezvous (RdV) missions have been proposed requiring relative navigation and control. The long

range operation phase indicates that satellites are far apart. The chaser satellite shall be able to detect, acquire, and track the relative position of the target satellite to close on, and then perform the final approach, rendezvous and docking (Grelier et al., 2009). Examples include Post-ATV (Crew Space Transportation System or CSTS), Next-Mars and Mars Sample Return (MSR) missions. Long-range metrology used in HEO activities include the Magnetospheric Multiscale (MMS) (Heckler et al., 2008), Magnetosphere Constellation (MagCON) and Cross-Scale mission, all studying the Earth's magnetosphere. Taking the MMS mission for example, a long-range operation across thousands of kilometers will perform in this four satellite formation as satellites progress through portions of a highly elliptical orbit. Weak-signal GPS receivers and an integrated S-band inter-satellite transceiver are used for relative navigation. The MMS mission is currently at the test stage, and is planned to launch in 2015.

1.1.2 Formation flying metrology

Formation flying creates large spaceborne instruments by coordinating two or more smaller satellites through observations made with larger and configurable baselines. This requires a high level accuracy of relative navigation and control.

The common way to perform relative navigation for formation flying missions is to utilize differential Global Navigational Satellite System (GNSS) measurements. This configuration could enable an accuracy better than one centimeter in certain cases, but is limited to formation flying missions in LEO.

For these LEO formation flying applications, the GNSS-based relative navigation is a standard metrology due to its accuracy, availability, flexibility and robustness. As opposed to the GNSS-based relative navigation that relies on the visibility of GNSS constellations, self-contained relative (inter-satellite) navigation metrology, i.e., through the transmission/reception of radio frequency (RF) and optical signals via inter-satellite links (ISLs), attract much attention recently. The availability of locally generated inter-satellite ranging links can augment the GNSS-based metrology for a more rapid and stable usage of the navigation filter for LEO applications, especially when the separations among spacecraft are highly variable (Renga et al., 2013).

Another clear reason to apply the self-contained relative navigation metrology arises from the need to implement a navigation system at altitudes above the GNSS constellations (e.g. GPS satellite are orbiting at an altitude of approximately 20,200 km). Only the very weak GNSS signals from the sidelobes in the opposite side of the Earth may be discontinuously available. More importantly, it is difficult to receive signals from four or more satellites simultaneously in such scenarios. Furthermore, poor geometric dilution of precision (GDOP) and slow line-of-sight (LOS) vector dynamics between the receivers and GNSS satellites make the precise GNSS carrier phase based solution difficult, especially make the carrier phase integer ambiguities weakly observable, hindering the estimator's ability to resolve the values on-the-fly (Mohiuddin and Psiaki, 2008). Therefore, self-contained relative navigation sensors, i.e., RF or optical, shall be used to fulfil the requirements of high altitude formation flying missions.

Optical metrology can enable higher ranging accuracy than the RF metrology. However, one problem of optical sensors is that they tend to have a relatively small field of view. To obtain global coverage for initial formation acquisition, one needs to have a large number of them or to extend their field of view by scanning them

(Tien et al., 2004). Another problem that militates against optical sensors is the Sun. When the sensing direction approaches the line of sight of the Sun, the sunlight may saturate or blind the optical sensor.

RF sensors have problems of their own, but they appear more manageable (Tien et al., 2004). RF metrology has been officially selected by ESA and CNES (Centre National d'Etudes Spatiales) as a coarse metrology for the first-stage formation acquisition on future European non-LEO formation flying missions (Grelier et al., 2008). A subsequent optical metrology subsystem can be further applied for a higher accuracy. This RF-based relative navigation does not rely on any a-priori information and can functionally enable omni-directional coverage to assure the initial formation acquisition.

This thesis investigates key technologies of a self-contained RF metrology system that can be used as an augmentation or a substitution to GNSS for future formation flying relative navigation.

1.1.3 Formation flying radio frequency metrology

The RF metrology via ISLs relies on local transmission/reception of ranging signals. In this context, different technologies can be exploited potentially:

- Ultra-WideBand (UWB) ranging technology;
- Radar transponder technology;
- GNSS-like RF technology.

UWB ranging technology involves the transmission of very short electromagnetic pulses at a very low energy level. These short pulse width signals (less than 1 ns pulse width) have very large signal bandwidths (minimum of 500 MHz, up to 7.5 GHz), which should, in theory and under proper circumstances, allow to share spectrum with other users. In contrast to most continuous wave ranging techniques, UWB signals have no carrier and use inherently spread spectrum¹, making them attractive in precision locating and tracking applications with centimeter level accuracy. UWB based ranging systems are already available commercially (MacGougan et al., 2008). However, important limitations still remain which prevent the implementation of such a technology in satellite formation flying applications. A key limitation stems from the nature that the UWB signal is spread over an extremely large spectrum, thus enables only low energy level emissions and limited operational range to prevent spectrum pollution.

Radar transponder technology has been recognized as a potential method for inter-satellite ranging and communication in space. The Dutch scientific research organisation TNO recently proposed a SPaceborne Active Ranging and Communication System (SPARCS) (Busking et al., 2011; Elferink and Hoogeboom, 2013), which utilizes a well-known Frequency Modulated Continuous Wave (FMCW) radar for inter-satellite two-way ranging. The basic principle is to transmit a FMCW signal, frequency modulated by, e.g., a saw tooth of a triangular signal as a function of time, to the transponder on the target spacecraft. The target in turn needs to gen-

¹In telecommunication and radio communication, spread spectrum techniques are methods by which a signal generated with a particular bandwidth is deliberately spread in the frequency domain, resulting in a signal with a wider bandwidth (Wikipedia, 2014b). The GNSS technology is a good application of spread spectrum, which generally makes use of a sequential noise-like signal structure to spread the normally narrowband information signal over a relatively wideband of frequencies.

Table 1.2: GNSS-like RF metrology overview

Radio Frequency (RF) inter-satellite system	Designer	Mission	Frequency band	PRN code	Data rate	Ranging accuracy
Star Ranger	AeroAstro	TechSat21	Ku	2 PRN code: 1023 chips (short), 1e8 chips (long)	128 kbps	sub-cm
AAF (Autonomous Formation Flying sensor)	JPL	St-3	Ka	100 Mcps ¹	-	1 cm
CCNT (Constellation Communication and Navigation Transceiver)	JPL	ST-5	S	100 Mcps	1 kbps	1cm
FAS (Formation Acquisition Sensor)	JPL	TPF	S	Ultra-BOC 10 Mcps	-	0.5 m, 1° LOS
SPTC (Stanford Pseudolite Transceiver Crosslink)	Stanford	-	L	Pulsed C/A code	38.4 kbps	2-5 m
FFRF (Formation Flying Radio Frequency Sensor)	CNES	PRISMA	S	C/A	4 kbps (500 m-30 km), 12 kbps (10 m-500 m)	1 m, 20° LOS (coarse mode) 1 cm, 1° LOS (fine model)
IRAS (Inter-satellite Ranging and Alarm System)	NASA & GSFC	MMS	S	C/A	128 bps (1800-3500 km), 512 bps (640-1800 km), 4kbps (250 m-640 km)	36 km (1800-3500 km), 9 km (640-1800 km), 30 m (250 m-640 km)
KRS (K/Ka-band Ranging System)	NASA	GRACE	K/Ka	-	-	range change: micron-level bias removal by integration with GPS

¹ Mcps: Mega chips per second.

erate a local FMCW signal, synchronized with the incoming FMCW signal, to transmit it back to the location of origin. It is not practical for the transponder to reflect back the received signal directly, as a signal propagation loss ratio of $1/d^2$ over the path d between the two spacecraft applies for both the path from the transmitter to the transponder, and vice versa, resulting in an signal attenuation of $1/d^4$. The synchronization step is of critical importance in this technique in order to compensate for the drift of the local FMCW generator (or oscillator). This radar transponder technology was only tested on ground. Some limitations have been found due to signal reflections (multipath effect).

Compared to the UWB-based method and the radar transponder technology, GNSS-like RF ranging technology is undoubtedly the most mature, taking advantages of well-assessed experiences on GNSS hardware and software in space applications. Table 1.2 summarizes the existing or proposed GNSS-like RF inter-satellite systems. Key examples include the K/Ka-band Ranging System (KRS) on the GRACE mission with micrometer-level ranging rate accuracy (Bertiger et al., 2002), the S-band Formation Flying Radio Frequency sensor (FFRF) on PRISMA mission with centimeter-level ranging accuracy (Thevenet and Grelier, 2012), and the Inter-satellite Ranging and Alarm System (IRAS) on the MMS mission with up to 30 m level accuracy (Heckler et al., 2008). For some missions like NASA's New Millennium Program missions ST-3 (Starlight) (Aung et al., 2002), ST-5 (BarSever et al., 2001), and Techsat-21 (Zenick and Kohlhepp, 2000), although aborted or heavily modified, developed technologies regarding RF ISLs are still valuable and inspiring. The Autonomous Formation Flying (AFF) sensor developed for ST-3 has been modified to a version called Formation Acquisition Sensor (FAS) with the intention to reuse it for the Terrestrial Planet Finder (TPF) mission (Tien et al., 2004).

These GNSS-like systems rely, in principle, on one-way ranging signals whose structure can be the same as conventional GPS pseudo random noise (PRN) C/A^2 transmissions. Some of the proposed systems modify the C/A code signal from a chipping rate of 1.023 Mcps to higher chipping rate (e.g. 100 Mcps), or to utilize other waveforms such as the binary offset carrier (BOC) signal (which is used for Galileo constellations (Hein et al., 2004)). The BOC signal modulates additional sub-carriers onto the conventional PRN code and has demonstrated an enhanced navigation performance as compared to the C/A code in terms of improved ranging accuracy and better immunity to signal reflections (multipath).

The GNSS-like systems support also inter-satellite communication. The supported data rate is mainly determined by the data type of key traffic according to mission requirements. Typical key traffic carried by ISLs include navigation measurements, housekeeping, timing and formation control commands, while a large amount of scientific data has not been considered to be transmitted between spacecraft in most proposed or flown formation missions provided a bandwidth limitation and low power consumption. Power constraints and a wide range of inter-satellite distances can require variable data rate communication, as well as variable ranging accuracy requirements at different mission phases (e.g., FFRF and IRAS systems in Table 1.2).

Besides the communication and relative distance estimation, the GNSS-like inter-

²The Coarse/Acquisition (C/A) code is a 1023 bit deterministic sequence which, when transmitted at 1.023 Megabits per second by GPS satellites, repeats every millisecond (Kaplan and Hegarty, 2006). Each bit in the C/A sequence is called a chip. The reciprocal of the chip period is known as the chipping rate. The C/A code has a chipping rate of 1.023 Mega chips per second (Mcps).

satellite systems FAS and FFRF provide also line-of-sight (LOS) bearing angles (e.g. elevation and azimuth) using carrier phase differences arriving at multiple antennas. The FFRF system has also been divided into coarse and fine modes. The coarse mode provides an accuracy of 1 m for distance and 20° for LOS with omni-directional coverage, while the fine mode provides an accuracy is 1 cm for distance and 1° for LOS, respectively.

This GNSS-like RF metrology, taking advantages of existing GNSS hardware and software, is the most mature, robust and cost effective technology as compared to the UWB-based system or Radar Transponder. Like the GNSS system, this GNSS-like system transmits two types of ranging measurements - pseudorange and carrier phase. The pseudorange measurement conveys information about the “apparent” distance between transmitter and receiver antennas, and is thus unambiguous while imprecise (on the order of meters). On the contrary, the carrier phase measurement is a range measure in units of cycles of the carrier frequency, thus can be made with very high precision (on the order of millimeters), but contains an unknown integer number of cycles (called integer ambiguity). This integer ambiguity has to be resolved before reaching millimeter level accuracy.

1.2 Research questions, objectives and methodologies

This thesis aims at investigating key technologies of a self-contained GNSS-like RF metrology for future formation flying missions. The following specific research questions (RQs) are addressed in this thesis.

RQ1: What is the architecture and functionality of an inter-satellite RF system?

RQ2: What algorithms shall be developed to enable relative navigation?

RQ3: How to improve the relative navigation performance in terms of accuracy, efficiency and reliability?

RQ4: How to apply relative navigation in a large-scale formation with four or more satellites?

The research starts with the investigation of system architecture and functionality. Although such a GNSS-like system can inherit mature GNSS hardware and software, questions are still present with respect to the frequency allocation, the antenna arrangement (e.g., the number of antennas and their relative orientation) and ranging code structure selection. From the experience of the benchmarking system - FFRF on the PRISMA mission, this RF system is expected to integrate inter-satellite communication, inter-satellite distance estimation as well as line-of-sight (LOS) estimation. The research in this thesis focuses more on the inter-satellite communication and LOS estimation, while the inter-satellite distance estimation algorithms are not specifically included in this thesis.

Both pseudorange and carrier phase measurements shall be used to allow for coarse-mode and fine-mode LOS estimation. The associated carrier phase integer ambiguity resolution (IAR) and error reduction in the fine-mode are elaborately discussed. Dominating error sources in the LOS model include mainly multipath, caused by signal reflections from structures in the surrounding of antennas. Novel methods for mitigating multipath in pseudorange and carrier phase measurements

will be proposed so as to improve LOS accuracy and accelerate IAR process. The thesis also includes discussions on potential network architectures to allow for relative navigation among a large-scale formation with four or more satellites.

More specifically, the research in this thesis has the following objectives:

1. Design of GNSS-like RF system architecture with functionalities of inter-satellite communication, inter-satellite distance estimation and LOS estimation. The inter-satellite distance needs to be estimated with meter-level accuracy in the course-mode and centimeter-level accuracy in the fine-mode. The LOS estimation is required at sub-degree accuracy in the fine-mode.
2. Investigation of transmitter and receiver architectures, frequency allocation, ranging code structure and signal processing strategies, as well as development of a software-defined radio for simulations and testing.
3. Development of an un-aided, fast and reliable integer ambiguity resolution for the LOS estimation.
4. Development of pseudorange multipath mitigation methods for improved accuracy.
5. Development of carrier phase multipath mitigation methods for improved accuracy as well as for IAR acceleration.
6. Investigation of formation network architecture to support various mission phases, from initial deployment to formation acquisition, maintenance and/or reconfiguration.

The research in this thesis is addressed and validated by different methods such as numerical simulations, software-defined simulator and receiver, case studies as well as field experiments. The software development environment is MATLAB.

Numerical simulations were used in Chapter 3 and 5, covering a large number of different measurement scenarios, where the impact of measurement precision and antenna geometry was analysed. Apart from the classical pure software simulations which make use of emulated measurements, the research in Chapter 3 also presents results from field tests for the demonstration of LOS estimation and associated IAR performances. Field tests were implemented in open-sky to simulate a two-spacecraft formation scenario where the GPS receiver on the ground represents one spacecraft while one of the GPS satellites is treated as the other spacecraft. IAR performance was tested with different receiver-satellite geometries.

The RF system functionality, architecture and performance were investigated by establishing of the software-defined simulator and receiver, in Chapter 2, 4 and 6. Implementing the software-defined simulator and receiver is a convenient starting point as they are easy and transparent to reconfigure and control. To demonstrate the proposed multipath mitigation performance, some unwanted error sources, e.g., the atmospheric errors, can be avoided in the signal generation process in the software-defined simulator, so that the isolated multipath effects are highlighted. The simulator and receiver in this research is stimulated by the work in Borre et al. (2007).

The research in this thesis also conducts case studies in Chapter 6 for the analysis of relative navigation errors due to multiple access interference in the formation with four or more spacecraft. Two realistic mission scenarios, one of a circular LEO

mission with centralized chief-deputy satellite topology and another for a highly elliptical orbit with four identical satellites, were investigated to demonstrate the severe multiple access interferences.

1.3 Structure of the thesis

The thesis is structured as follows.

Chapter 2 presents specific features in the design of a RF-based relative navigation transceiver. The transceiver architecture, functionality, performance, and the associated frequency allocation and ranging code structure are elaborately discussed. The chapter also introduces the outputs of the transceiver, which are the pseudorange and carrier phase observables. Following an analysis of various error sources, the basic principles for the inter-satellite distance and LOS estimation are derived using these observables.

Chapter 3 focuses on elaborating the LOS estimation model, resolving the associated carrier phase integer ambiguities timely, efficiently and reliably, evaluating the antenna geometry impacts, deriving ambiguity dilution of precision analytically, and characterizing the estimation performance by both numerical simulations and field tests.

The proposed ambiguity resolution methods in Chapter 3 are based on single-epoch measurements. Only random noise is assumed in the model. Small multipath can be tolerated when it is lumped together with the thermal noise in a single epoch. For large multipath, the multi-epoch processing has to be applied when multipath is treated as a coloured noise with time correlations.

Chapter 4 aims at mitigating multipath on pseudorange measurements. The chapter explores correlations between the multipath and signal strength. A multipath envelope curve fitting method is then proposed that provides the best fit to the multipath error in the least-squares sense by using the combination of multiple signal strength estimators. Both the estimation performance and the noise induced in the estimation process are discussed. The software multipath simulator and receiver are designed to demonstrate this new method.

Chapter 5 is devoted to the carrier phase multipath mitigation solutions. A promising multiple antenna-based method is proposed using the signal-to-noise ratio (SNR) data in cascaded extended Kalman filters. A cascaded procedure is used in order to split the multipath correction process into cascaded filters before and after fixing integer ambiguities. The filter can be either real-valued or complex-valued. The filter performance is evaluated and the multipath effects on the integer ambiguity resolution are also examined.

As a successor to previous chapters on the RF-based relative navigation system and algorithms designed for a two-spacecraft formation, Chapter 6 aims at extending previous scenarios and results for a large scale formation with four or more spacecraft. The chapter includes a discussion on potential formation network architectures and an investigation of limitations in implementing specific architectures. CDMA is emphasized in this chapter with its limitations in terms of multiple access interference and near-far problem. Two realistic mission scenarios in the Low Earth Orbit (LEO) and in the Highly Elliptical Orbit (HEO) are analysed to address the effects of the multiple access interference on the communication performance as well as on the navigation accuracy.

Finally, Chapter 7 summarizes the thesis, draws conclusions and provides recommendations for future work.

Chapter 2

RF-based Relative Navigation System Design and Analysis

This chapter presents specific features in the design of a radio-frequency (RF)-based relative navigation transceiver. The transceiver architecture, functionality, performance, and the associated ranging code structure are all elaborately discussed. The chapter also introduces the outputs of the transceiver, which are the pseudorange and carrier phase observables. Following an analysis of various error sources, the basic principles for the inter-satellite distance and line-of-sight (LOS) estimation are derived using these observables.

Chapter 3 will then further elaborate the LOS estimation and the associated carrier phase integer ambiguity resolution. A dominating error source, multipath, will be specifically discussed in chapter 4 and 5, together with innovative mitigation methods. In addition, based on the proposed system architecture of this chapter, a software transceiver has been designed and will be used for performance verifications in chapter 4 and 6.

2.1 RF-based relative navigation system design

2.1.1 Architecture

A RF-based relative navigation functionality can be achieved by utilizing locally generated RF ranging signals. A cost effective manner to generate these signals is to modify an existing GNSS receiver such that it can operate as a transceiver.

The transceiver terminal is suggested to consist of one transmitter (Tx), one receiver (Rx) and several antennas (see Figure 2.1), enabling a joint inter-satellite communication and relative navigation. Two types of antennas, Tx/Rx antennas and Rx-only antennas, are utilized. While the Tx/Rx antenna enables the exchange of communication data and ranging measurements between two spacecraft, the Rx-only antenna is only used for navigation purposes, i.e. to assist in the estimation of the relative line-of-sight (LOS) bearing angles (elevation and azimuth) using carrier phase differences arriving at multiple antennas. Multiple channels are allocated for

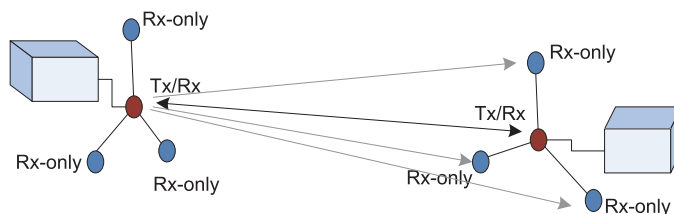


Figure 2.1: Inter-satellite relative navigation system configuration (Thevenet and Grelier, 2012)

Table 2.1: Proposed signal structure (Justifications on frequency allocations can be found in section 2.1.2)

Carrier frequency	S1: 2271.06 MHz (222×10.23 MHz)	S2: 2107.38 MHz (206×10.23 MHz)	S3: 2056.23 MHz (201×10.23 MHz)
Modulation	QPSK		BPSK
Channel	I (pilot)	Q (data)	I (data)
Ranging code	PRN code	-	-
Data rate	-	12 kb/s	tbd*

*to be determined.

signal reception from multiple antennas.

In Figure 2.1, the transceiver is identical on each spacecraft, supporting a distributed formation topology with an equal navigational capability. For a master-slave formation, some functionalities on the slave satellite (i.e. the LOS estimation) may not be required and the associated hardware (i.e. Rx-only antennas) can then be removed for energy saving.

The signal structure is proposed in Table 2.1 according to support both coarse-mode and fine-mode navigation. In the course-mode, i.e. for collision avoidance, the pseudorange measurement can be solely used to achieve a meter-level accuracy. In the fine-mode, much more precise carrier phase measurements have to be included to enable the distance estimation to centimeter-level accuracy and the LOS bearing angle estimation to sub-degree level accuracy. The carrier phase is measured modulus 2π . An integer number of cycles thus remains unknown and must be resolved before the carrier phase reaches its nominal precision. Three carrier frequencies S1, S2 and S3 are suggested in order to enable fast and reliable integer ambiguity resolution (IAR) (Teunissen et al., 2002; O’Keefe et al., 2009). It is also possible to fix ambiguities using only dual-frequency or even single-frequency measurements. However, in this case, either the success rate is lower or other resources, e.g., from inertial sensors are required. The IAR process will be elaborated in chapter 3. Specific values for S1, S2 and S3 shall meet the regulations of the International Telecommunication Union (ITU). The frequency intervals between S1 and S2 as well as S2 and S3 are chosen to enable the potential of building widelane and extra-widelane measurements, which can be used to facilitate the IAR process (O’Keefe et al., 2009). The reason of using S-band for these three frequencies will be explained in section 2.1.2.

In-phase (I) and quadrature (Q) channels are allocated in S1, one being used for ranging by the PRN code and the other offering, e.g., 12 kb/s inter-satellite communication for measurement exchange as well as for command and control purposes. The signals in these two channels do not overlap since the in-phase and quadrature carriers have 90° phase shift and thus can be orthogonally multiplexed by the QPSK modulation. On S2 and S3 frequencies, only low rate communication data are mod-

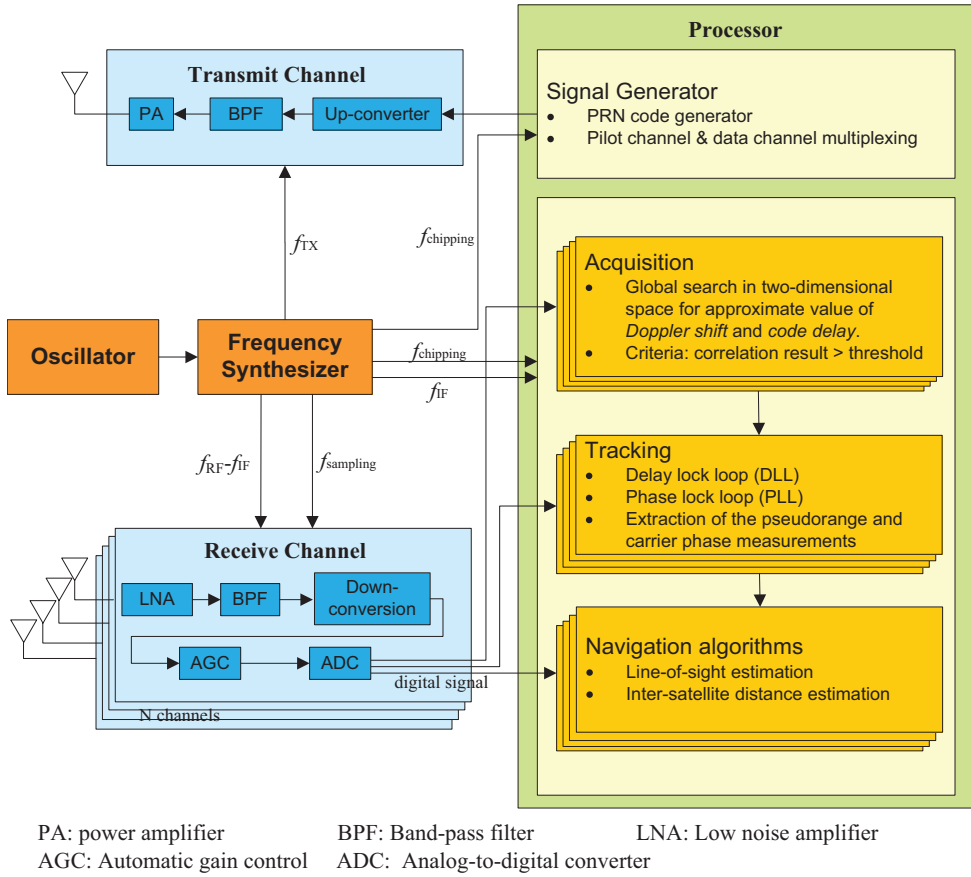


Figure 2.2: Architecture of an inter-satellite relative navigation transceiver

ulated onto the carrier and the high rate PRN ranging code does not exist. The purpose is to maximally avoid the PRN code despreading process in the receiver but maintain extra carriers to facilitate IAR. Variable communication data rates can be allocated to different carriers in order to efficiently transmit different types of data.

Figure 2.2 illustrates the block diagram of the transceiver architecture, which inherits GNSS technologies. Ranging is performed using the PRN code.

The signal generation begins from the PRN code generation at the upper right in the baseband processor in Figure 2.2. The signal will be up-converted (modulated), passed to a passband filter (BPF) and a power amplifier chain before being transmitted by the antenna.

In the receiver, multiple channels, connecting to separate front-ends and separate intermediate-frequency (IF) signal processing, are allocated for signals received by different antennas. In each channel, the signal goes through a low-noise amplifier (LNA) and a bandpass filter chain before it is down-converted to the IF band and digitized by the A/D converter. The digitized signals are then output to the IF band processor, being processed through acquisition and tracking. The acquisition is a global search in a two dimensional search space for approximate values of Doppler shift and code delay. After acquisition, control is handed over to the delay lock loop (DLL) and phase lock loop (PLL), where the fine estimates of the code and carrier

Table 2.2: Frequency candidates for the inter-satellite link (Delpech et al., 2011; Tien et al., 2004; Bertiger et al., 2002)

Band	Frequency Range	Examples
S	2.025-2.110 GHz 2.200-2.290 GHz	PRISMA TPF, PROBA-3
Ku	13.75-14.3 GHz 14.5-15.35 GHz	Star ranger for TechSat 21
Ka	22.55-23.55 GHz 25.25-27.50 GHz 32.30-33.40 GHz	Iridium GRACE (K/Ka band) StarLight
W	59 - 64 GHz 65 - 71 GHz	

phases will be obtained and continuously updated. Variations due to the dynamics between the spacecraft will also be adaptively tracked. After the tracking loops, pseudorange and carrier phase observables will be extracted accordingly, which are then fed into the navigation algorithms at certain intervals for the inter-satellite distance and LOS estimation.

It can also be seen in Figure 2.2 that the core to the whole architecture is a frequency standard synthesizer that provides coherent frequency references from the oscillator to various parts of the transceiver, e.g., for generating and modulating the PRN code components onto the carrier, for demodulating the carrier to the baseband or IF band, and also for controlling signal sampling and keeping time synchronization.

2.1.2 Frequency allocation

The appropriate frequency band (or bands) is an important part of any recommendation for the inter-satellite link. The choice of frequency bands depends upon the spectrum regulations specified by the ITU, technical characteristics and constraints (including availability of hardware), as well as mission requirements.

International and national spectrum regulations are an important consideration when identifying the proper spectrum or when making frequency allocations for inter-satellite links. Based on the service designation by ITU, several frequency allocation options may be available for inter-satellite communications (Edwards, 2002), as listed in Table 2.2. Examples of several distributed spacecraft missions with different frequency allocations are also listed in Table 2.2.

Besides these spectrum regulations, the system designer needs to consider the availability of hardware and the technical characteristics of the frequency bands. Several technical parameters influence the selection of frequency bands.

Link budget

This associates with required transmitter power, propagation, and antenna characteristics, which can vary greatly depending upon frequencies.

It is well understood that the Friis transmission equation is used to formulate the power received by one antenna under idealized conditions given another antenna some distance away transmitting a known amount of power. The Friis transmission

equation is expressed as (Wikipedia, 2014a)

$$P_r = P_t G_t G_r \left(\frac{4\pi d}{\lambda} \right)^2 = P_t G_t G_r \left(\frac{4\pi d f}{c} \right)^2 \quad (2.1)$$

where P_r are P_t are the input power of the receiving antenna and the output power of the transmitting antennas, G_t and G_r are the antenna gains of the transmitting and receiving antennas respectively, λ is the carrier wavelength, d is the distance between the antennas and c is the speed of light. The inverse of the factor in parentheses is the so-called free-space path loss, which is proportional to the square of the carrier frequency f .

The directivity of an antenna, its ability to direct radio waves in one direction or receive from a single direction, is measured by the antenna gain G . The antenna gain is the ratio of the power received by the antenna to the power that would be received by a hypothetical isotropic antenna, which receives power equally well from all directions. The antenna gain is a function of the antenna aperture or effective area A , which measures how effective an antenna is at transmitting/receiving the power of radio waves. The antenna gain is also frequency dependent

$$G = \frac{4\pi A}{\lambda^2} = \frac{4\pi A f^2}{c^2}. \quad (2.2)$$

With the increase of the frequency the requirement of keep the antenna gain intact will cause an antenna aperture to be decreased, which will result in less energy being captured with the smaller antennas. If keeping the antenna aperture intact, a higher frequency narrows the beam widths of the antenna and thus requires higher pointing accuracy.

Ionospheric effects, multipath, integer ambiguity resolution and Doppler shifts effects

ISLs are normally used both for the distribution of data among spacecraft and for navigation purposes. Performing pseudorange and carrier phase measurements is a common way to realize relative navigation. The frequency allocation has an effect on the navigation error budget, including ionospheric effects, carrier multipath, signal acquisition error and integer ambiguity resolution. The ionospheric delay is inversely proportional to the square of the carrier frequency. The carrier multipath and thermal noise are both inversely proportional to the carrier frequency. Therefore, a higher frequency allocation helps to reduce the errors caused by ionospheric path delay, phase tracking loop and phase multipath. However, due to higher maximum Doppler shifts at the higher frequency, the Doppler search region increases, which negatively influences the signal acquisition. Assuming identical code length, signal acquisition takes a longer time at the high frequency band in order to cover all possible Doppler regions. Moreover, the integer ambiguity resolution in terms of achieving a high success rate becomes more difficult in the high frequency band. This is due to the fact that a given range will comprise more integer cycles and thus the probability of successfully fixing these cycles becomes lower. More formally speaking, the ambiguity dilution of precision is higher at high frequencies.

Isolation with other onboard communication systems like the TT&C

If the inter-satellite system and the TT&C subsystem work in the same frequency band, i.e. S band, sufficient frequency separation between the inter-satellite link

Table 2.3: Frequency allocation for PRISMA mission (Lestarquit et al., 2006)

Carrier frequency (S band)	FFRF	
	S1: 2275 MHz	S2: 2105 MHz
	TT&C	
	TM: 2214 MHz	TC: 2035 MHz

and the TT&C uplink and downlink is necessary to reduce the risk of disturbance. The PRISMA mission is a good example when assigning carrier frequencies to its Formation Flying RF (FFRF) sensor (Lestarquit et al., 2006). Table 2.3 shows the frequency allocation on PRISMA. Apart from the consideration of separating the ISL and TT&C, the FFRF selected two frequencies S1 and S2 and a reasonable isolation between S1 and S2 in order to optimize the integer ambiguity resolution functionality by building a so-called widelane measurement (Lestarquit et al., 2006).

Considering both ITU regulations and technical constraints, frequencies allocated to the inter-satellite transceiver are recommended to use S-band frequencies, specifically at S1 (222×10.23 MHz), S2 (206×10.23 MHz) and S3 (201×10.23 MHz) in this research.

2.1.3 PRN code structure

The PRN code, as the ranging code in the inter-satellite system, consists of a sequence of +1's and -1's, which has limited length and is periodic forward in time. A PRN code only matches up, or strongly correlates to another PRN code, when they are exactly aligned. This property is called *correlation*. Auto-correlation measures the similarity between any PRN sequence and time shifts of itself, while cross-correlation compares a given sequence with all time shifts of a second sequence. The auto-correlation function for the PRN code can be expressed as (Misra and Enge, 2001)

$$R(\tau) = \frac{1}{LT_c} \int_0^{LT_c} c(t)c(t-\tau)dt \quad (2.3)$$

where $c(t)$ is the PRN sequence, T_c is called a chip period when +1 or -1 in a PRN sequence is called a chip. The reciprocal of the chip period is known as the chipping rate f_c and L is the number of chips in each repeat of the PRN sequence. As shown in this equation, auto-correlation multiplies $c(t)$ by a time-shifted replica of itself $c(t-\tau)$ and integrates the product. If $c(t-\tau)$ resembles $c(t)$, $R(\tau)$ will be large. It is assumed that the code $c(t)$ and $c(t-\tau)$ repeat indefinitely. In this case, Eq.(2.3) is called the periodic or circular auto-correlation function. A single period of the periodic code, $c_1(t)$, can be written as

$$c_1(t) = \sum_{l=0}^{L-1} c_l p\left(\frac{t-lT_c}{T_c}\right) \quad (2.4)$$

where c_l is the l th element in the sequence that is equal to +1 or -1, and $p(t)$ is the elemental chip waveform with unit width, unit length and centered at the origin. In this equation, $p(t)$ is modified to have a duration T_c and delay lT_c .

The chip waveform $p(t)$ in the GPS system has a rectangular shape. In principle, any shape could be used and different shapes can be used for different chips. Henceforth, the sequence generated using binary phase shift keying (BPSK) signalling with

rectangular chips is denoted as BPSK-R(n) signal, which has the chipping rate of $n \times 1.023$ MHz. The C/A code is a BPSK-R(1) signal while the P code is a BPSK-R(10) signal. Several variations of the basic waveform that employ non-rectangular symbols have also been investigated in recent years, such as the binary offset carrier (BOC) signals for the Galileo satellite navigation system (Betz, 2001; Hein et al., 2004, 2006; Avila-Rodriguez et al., 2008). A BOC signal may be viewed as the product/modulation of a BPSK-R signal with a square wave subcarrier (Kaplan and Hegarty, 2006). The notation BOC(m,n) is shorthand for a BOC modulation generated using a square wave subcarrier frequency $f_{sc} = m \times 1.023$ MHz and a chipping rate $f_c = n \times 1.023$ MHz. For example, BOC(10,5) means having a 10.23 MHz subcarrier frequency and a 5.115 MHz chipping rate. The chip waveform for the BPSK-R(n) and BOC(m,n) are

$$p_{\text{BPSK-R}}(t) = \begin{cases} 1, & 0 \leq t \leq T_c \\ 0, & \text{elsewhere} \end{cases} \quad (2.5)$$

$$p_{\text{BOC}}(t) = p_{\text{BPSK-R}}(t) \text{sgn} \left[\sin \left(\frac{\pi t}{T_{sc}} + \psi \right) \right] \quad (2.6)$$

where $\text{sgn}(t)$ is the signum function (1 if the argument is positive, -1 if the argument is negative) and ψ is a selectable phase angle. Two common values of ψ are 0° or 90° , for which the resultant BOC signals are referred to as *sine phased* or *cosine phased*, respectively. T_{sc} is the half-period of a square wave generated with subcarrier frequency $T_{sc} = 1/(2f_{sc})$. The number of square wave half-periods in a chip is typically selected to be an integer k

$$k = \frac{T_c}{T_{sc}} = \frac{2f_{sc}}{f_c} = \frac{2m}{n} . \quad (2.7)$$

Figure 2.3 illustrates several examples of the auto-correlation functions for either BPSK-R or BOC modulated signals. As can be seen, the auto-correlation function of the BPSK-R modulation has a sharp and distinct triangle peak. The closed form auto-correlation function for the BPSK-R modulation is

$$R_{\text{BPSK-R}}(\tau) = \begin{cases} 1 - |\tau|/T_c, & |\tau| \leq T_c \\ 0, & \text{elsewhere} \end{cases} . \quad (2.8)$$

The maximum auto-correlation occurs when the relative shift is zero, and the auto-correlation drops to zero for all other shifts larger than T_c .

The BOC modulation in Figure 2.3 exhibits sharper auto-correlation as compared to the BPSK-R modulation given the same chipping rate. This property assures better ranging accuracy (Betz, 2001). However, the auto-correlation function of the BOC modulation consists of multiple segments of connected lines with multiple zero-crossings and multiple peaks, leading to the difficulty of maintaining track of the main peak in the code tracking process. The number of positive and negative peaks is $2k - 1$, and the peaks are separated in delay by T_{sc} . Other characteristics of the BOC auto-correlation is summarized in Table 2.4.

Two characteristics are of great importance for PRN signals: one is the aforementioned auto-correlation function and the other is power spectral density (PSD). The PSD is defined to be the Fourier transform of the auto-correlation function, which describes the distribution of power with respect to frequency. After the modulation with the PRN code, the signal spectrum is spread as wider bandwidth will be occupied by the high-rate PRN waveform. In general, the bandwidth is proportional to

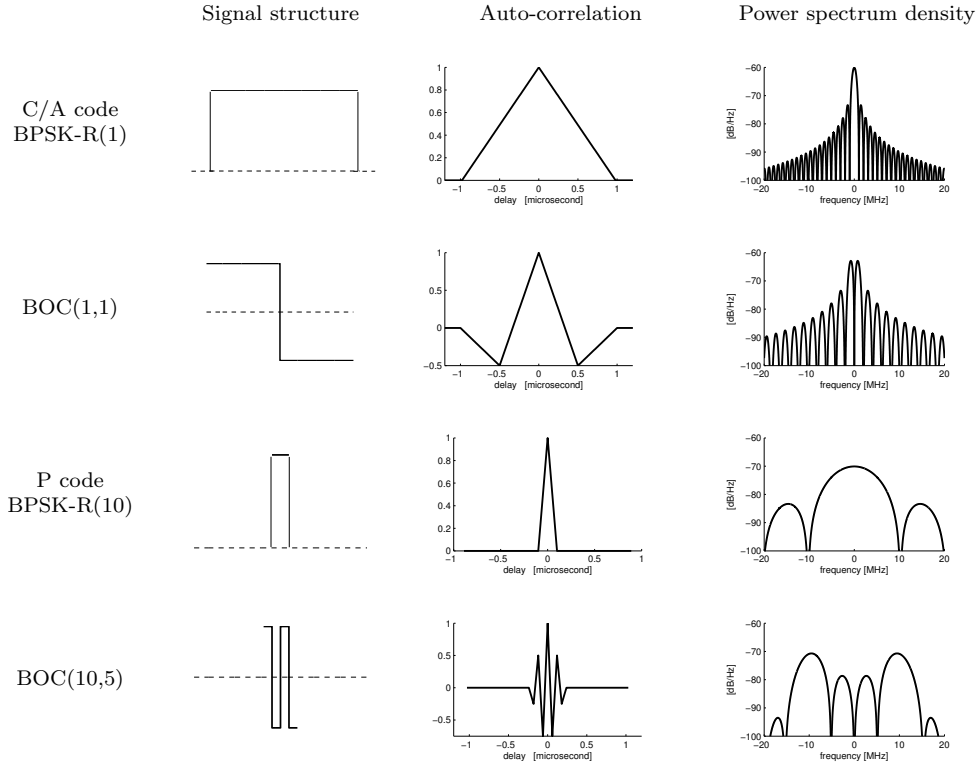


Figure 2.3: Examples of the auto-correlation and power spectrum density for either BPSK-R or BOC signals

Table 2.4: Auto-correlation function characteristics for BOC(m,n) modulation (Betz, 2001). Here, j denotes the index of the auto-correlation peak with $j = 0$ indicating the main peak and $j = 1$ being the first peak to the right of the main peak, etc.

Number of positive and negative peaks	Delay values of peaks	Auto-correlation values for peaks at $\tau = jT_{sc}$	Zero crossings nearest the main peak
$2k - 1$	$\tau = jT_{sc},$ $j = -k, \dots, k$	$(-1)^{ j } \frac{k - j }{k}$	$\pm \frac{k}{2k - 1} T_{sc}$

the chipping rate. Figure 2.3 also exhibits some examples of the PSD for BPSK-R or BOC modulated signals. The BPSK-R modulated signals, e.g., C/A code and P code, place most of the signal power in the middle frequencies of their bands, while the BOC signals split most of the power into two symmetrical mainlobes, which are shifted from the central frequency by the amount equal to the subcarrier frequency. Different locations of mainlobes show that their power is located in different portions of the band. The sum of the number of mainlobes and sidelobes between two mainlobes is equal to k . The closed-form expressions of the normalized (unit area over infinite bandwidth) PSD, denoted by $G_{\text{BPSK-R}}(f)$ and $G_{\text{BOC}}(f)$, are given as (Betz, 2001)

$$G_{\text{BPSK-R}}(f) = T_c \text{sinc}^2(\pi f T_c) \quad (2.9)$$

$$G_{\text{BOC}}(f) = \begin{cases} T_c \text{sinc}^2(\pi f T_c) \tan^2\left(\frac{\pi f}{2f_{sc}}\right), & k \text{ even} \\ T_c \frac{\cos^2(\pi f T_c)}{(\pi f T_c)^2} \tan^2\left(\frac{\pi f}{2f_{sc}}\right), & k \text{ odd} \end{cases} \quad (2.10)$$

where $\text{sinc}(x) = (\sin x)/x$. Note that the BPSK-R(n) can be treated as a special case of BOC(m,n) when $k = 2m/n = 1$. The C/A code equals BOC(0.5,1) and P code equals BOC(5,10).

2.2 Transmitter architecture

The specific signal generation in the transmitter is described by the block diagram in Figure 2.4. At the far left, the main clock signal is supplied to the remaining blocks. The clock signal has a standard frequency of 10.23 MHz. When multiplied by 222, 206 and 201, respectively, it generates the S-band carrier signals at three frequencies. At the bottom left, the clock signal is supplied to the BPSK-R(n) code or BOC(m,n) code generators. The resultant signal spectrum at three frequencies is shown in Figure 2.5. It can be seen that only the central spectrum is spread by the PRN code with high chipping rate, whereas two narrow tones on S2 and S3 are only modulated slowly by the low rate data. In this way, two extra phase measurements are obtained without the need of extra despreading, and they are separated away from the central carrier for the acceleration of the integer ambiguity resolution. This signal structure was originally proposed by Tien et al. (2004) and named as *ultra-BOC* signal structure. It is a variant structure of using multiple frequencies, showing a similar capability of facilitating the integer ambiguity resolution. The frequency spectrum designed in Figure 2.5 is representative of using three frequencies in S-band.

The BPSK-R(n) or BOC(m,n) ranging code is supplied to the in-phase BPSK modulator only on the S1 frequency, while different types of the communication data, including e.g., the navigation data, scientific observation data and the command and control, are separately supplied to the quadrature BPSK modulator on the S1 frequency and to the BPSK modulators on S2 and S3 frequencies. To this end, the ranging code and the communication data do not overlap as the signals are transmitted by using either orthogonally multiplexed channels or different carriers.

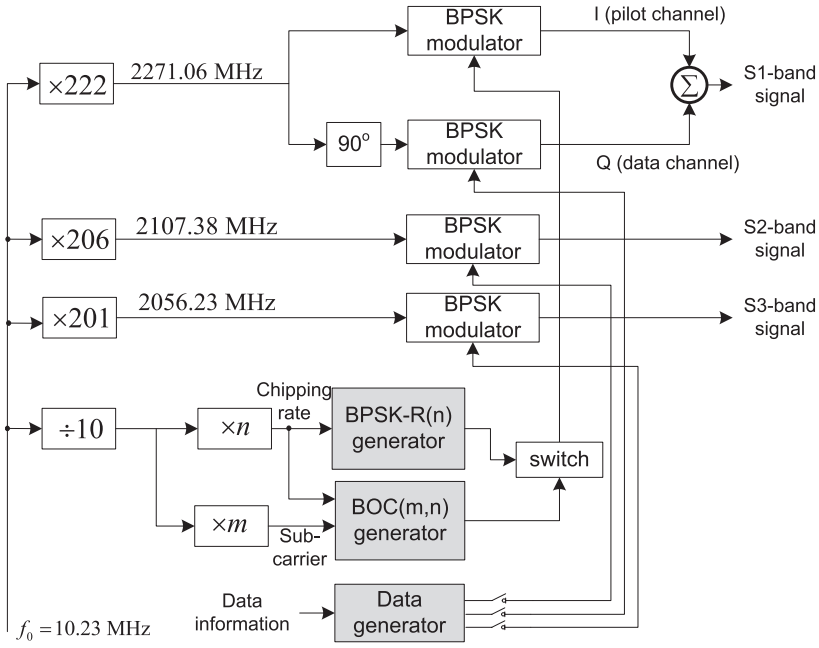
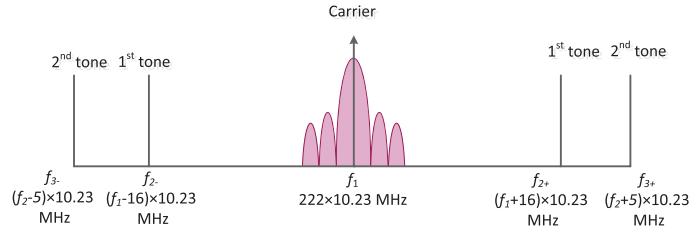


Figure 2.4: Generation of RF-based ranging signals in the transmitter

Figure 2.5: Ultra-BOC signal structure, representative of using three carrier frequencies at 222×10.23 MHz, $206 (222-16) \times 10.23$ MHz and $201 (206-5) \times 10.23$ MHz

2.3 Receiver architecture and analysis

2.3.1 Signal conditioning in the front-end

The front-end of a receiver establishes the starting point in the design of any receiver. As illustrated in the system architecture in Figure 2.2, it utilizes a combination of amplifier(s), mixer(s), filter(s) and its local oscillator to condition the received weak analog signal before converting it to digital samples.

An amplifier is a component that increases the signal magnitude, while it also adds noise. Specifically, the low noise amplifier (LNA) in the front-end is the component that amplifies the signal and adds minimal noise. The fundamental parameters used to describe an amplifier include (1) *gain*, usually expressed in dB, and often assumed constant over a specified frequency range; and (2) *noise figure*, also expressed in dB, and indicative of the amount of noise that will be added to the signal. A

typical amplifier in GNSS receivers requires a gain larger than 30 dB (Borio, 2012).

The mixer and local oscillator combination is used to translate the input carrier to a lower intermediate frequency (IF). The mixer operates through the trigonometric identity expressed as

$$\cos(\omega_1 t) \cos(\omega_2 t) = \frac{1}{2} \cos((\omega_1 - \omega_2)t) + \frac{1}{2} \cos((\omega_1 + \omega_2)t) \quad (2.11)$$

where ω_1 represents the angular frequency of the incoming signal while ω_2 denotes the locally generated angular frequency in the receiver. It is obvious that the outputs of the mixer are the sum $\omega_1 + \omega_2$ and difference frequencies $\omega_1 - \omega_2$. Of interest here is the difference frequency, which is the desired IF. The sum frequency is filtered out by the subsequent band-pass filter (BPF).

A filter is a frequency selective device that allows only certain frequencies to pass and attenuates others. An ideal filter (rectangular filter) would pass a range of frequencies and completely removes other frequencies outside that range. Unfortunately, such an ideal filter does not exist. The transition between those frequencies that are passed and removed is a gradual transition. An important parameter to characterize a filter is its 3 dB bandwidth. It is a frequency range within which the spectral density is above half of its maximum value, that is, above -3 dB relative to the peak.

The PRN signals and outer tones, shown in Figure 2.5, can be filtered separately by a technology called multi-band filtering (Lunot et al., 2008). Tones are away from the central spectrum and narrow band, whereas the PRN signals have spread spectrum and are wide band in nature. Therefore, the PRN signal has to be filtered in a larger bandwidth, which at least needs to include the mainlobe of its power spectrum. Figure 2.6 illustrates different bandwidth to BPSK-R(1) and BOC(10,5) signals. For the BPSK-R(1), the PSD mainlobe locates within 2 MHz, while for the BOC(10,5), two separate mainlobes locate within 30 MHz. The wider is the bandwidth for the PRN code, the less is the power loss after bandlimiting. This power loss can lead to a rounded auto-correlation peak as shown in Figure 2.7. As a consequence, it reduces the ranging accuracy as compared to the infinite-bandwidth case.

The final component in the front-end path is the analog-to-digital converter (ADC). The function of an ADC is two-fold: *sampling* and *quantization* which transform the continuous time-various and valued signal to discrete time-various and valued signal. By Shannon's sampling theorem, sampling can be achieved with no loss of information under the condition that the sampling frequency f_s has to be at least twice the signal bandwidth. Quantization, on the other hand, always leads to some losses. This is related to the number of bits used in quantization. It has been reported in Bastide et al. (2003) that if single bit quantization is used, the degradation in the resulting processing is less than 2 dB. If a two or more bit quantization is utilized, the degradation is less than 1 dB. It is important to recognize that if multibit quantization is employed, an automatic gain control (AGC) needs to be implemented to properly scale the input signal in order to fully use the ADC dynamics. Strong signals are clipped if the signal amplitude is beyond the peak-to-peak range of the ADC (Kaplan and Hegarty, 2006).

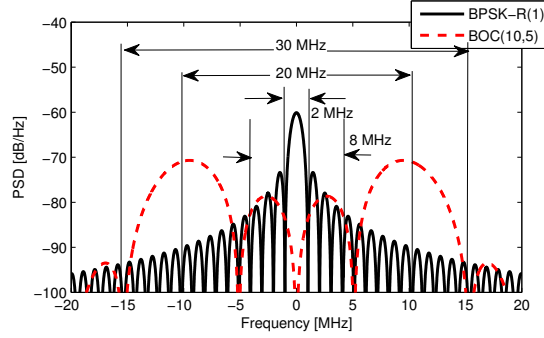


Figure 2.6: BPSK-R(1) and BOC(10,5) bandlimited PSD

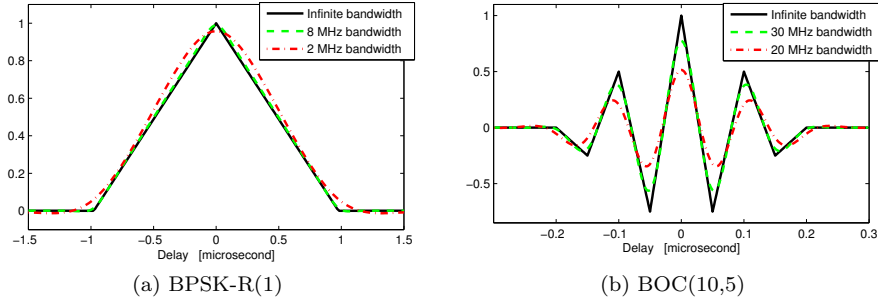


Figure 2.7: Bandlimiting effects in auto-correlation function

2.3.2 Acquisition

After amplification, filtering, down-conversion and quantization, the digitalized received signal $r[n]$ on the S1 frequency can be expressed as:

$$r[n] = \sqrt{2C}c(nT_s - \tau) \cos(2\pi(f_{IF_1} + f_{d_1})nT_s + \varphi) + \eta[n] \quad (2.12)$$

where τ and φ are the code delay and carrier phase, f_{IF_1} is the IF frequency down-converted from the S1 frequency, f_{d_1} is the Doppler frequency on S1, C is the received signal power, $T_s = 1/f_s$ is the sampling period, and $\eta[n]$ is the noise, which can be modelled as Gaussian distribution $\eta[n] \sim \mathcal{N}(0, \sigma_{IF_1}^2)$ with the variance $\sigma_{IF_1}^2 = N_0\beta_r$ (Borio, 2008), N_0 is the noise floor and β_r is the front-end bandwidth for the PRN code filtering.

Once the digital signals are obtained, the first task of the IF-band signal processing in the receiver is to determine whether a signal is present in the collected data. More specifically, the processor has at its disposal N signal samples to choose between two hypotheses:

$$\begin{aligned} H_1 : r[n] &= y[n] + \eta[n] && \text{signal present} \\ H_0 : r[n] &= \eta[n] && \text{signal absent} \end{aligned}, \quad 0 \leq n \leq N \quad (2.13)$$

where $y[n] = \sqrt{2C}c(nT_s - \tau) \cos(2\pi(f_{IF_1} + f_{d_1})nT_s + \varphi)$.

The decision statistics can be obtained by computing the likelihood ratio Λ between two maximum likelihood estimators under H_1 and H_0 hypotheses. The deriva-

tions can be found in Borio (2008)

$$\begin{aligned}\Lambda &= \frac{\max_{\tau, f_{d_1}, \varphi} \mathcal{L}(\tau, f_{d_1}, \varphi | H_1)}{\mathcal{L}(H_0)} \\ &= \exp\left\{-\frac{NC}{2\sigma_{IF_1}^2}\right\} \max_{\tau, f_{d_1}, \varphi} \exp\left\{\frac{1}{\sigma_{IF_1}^2} \sum_{n=0}^{N-1} r[n]y[n]\right\}\end{aligned}\quad (2.14)$$

The equivalent decision statistics can be obtained by removing the constant terms and taking the log of Λ

$$\begin{aligned}\Lambda_0 &= \max_{\tau, f_{d_1}, \varphi} \sum_{n=0}^{N-1} r[n]y[n] \\ &= \max_{\tau, f_{d_1}, \varphi} \frac{1}{N} \sum_{n=0}^{N-1} r[n]c(nT_s - \tau) \cos(2\pi(f_{IF_1} + f_{d_1})nT_s + \varphi) \\ &= \max_{\tau, f_{d_1}, \varphi} \Re\left\{\frac{1}{N} \sum_{n=0}^{N-1} r[n]c(nT_s - \tau) \exp\{-j2\pi(f_{IF_1} + f_{d_1})nT_s - j\varphi\}\right\}.\end{aligned}\quad (2.15)$$

This equation shows a fundamental function for signal processing: the arguments that maximize this function are the maximum likelihood estimators for the code delay, Doppler frequency and carrier phase. We can rewrite the function Λ_0 as

$$\Lambda_0 = \max_{\tau, f_{d_1}, \varphi} \Re\{R_c(\tau, F_{d_1}) \exp(-j\varphi)\}\quad (2.16)$$

where $R_c(\tau, F_d)$ is defined as a generalization of the standard auto-correlation function which indicates the similarities between two signals as a function of time and frequency shifts

$$R_c(\tau, F_{d_1}) = \frac{1}{N} \sum_{n=0}^{N-1} r[n]c(nT_s - \tau) \exp\{-j2\pi F_{d_1} nT_s\}, \quad F_d = f_{IF_1} + f_{d_1}.\quad (2.17)$$

For $F_d = 0$, the standard auto-correlation function can be found.

The maximization with respect to the phase can be solved in closed-form

$$\Lambda_0 = \max_{\tau, f_{d_1}} |R_c(\tau, F_d)|\quad (2.18)$$

$$\hat{\varphi} = \arctan\left(\frac{\Im\{R_c(\tau, F_d)\}}{\Re\{R_c(\tau, F_d)\}}\right).\quad (2.19)$$

Then, the final test for determining the signal presence becomes

$$\max_{\tau, f_d} |R_c(\tau, F_d)|^2 \begin{cases} > T_h, & \text{signal present} \\ < T_h, & \text{signal absent} \end{cases}\quad (2.20)$$

where T_h is the decision threshold. This equation indicates that the test should be implemented in two steps: (1) computation of $R_c(\tau, F_d)$ over a finite discrete bi-dimensional grid of frequencies and delay values; (2) search of the maximum and decision. However, the phase of the incoming signal, the noise and other impairments

can degrade the reliability of $R_c(\tau, F_d)$. To remove the dependence of the input signal phase, the absolute value of $R_c(\tau, F_d)$ is squared in Eq.(2.20) before it can be compared with the threshold. Due to noise, the value of $|R_c(\tau, F_d)|^2$ is a random variable, namely the decision variable, which can be characterized by two probability density functions (pdf) referring to the presence or absence of the desired signal. The probability that the decision variable passes a threshold is called *detection probability* $P_d(T_h)$ if the desired signal is present, and *false alarm probability* $P_{fa}(T_h)$ if it is absent. Their closed-form expressions have been derived in Borio (2008)

$$P_{fa}(T_h) = \exp\left(-\frac{T_h}{2\sigma_n^2}\right) \quad (2.21)$$

$$P_d(T_h) = Q_1\left(\sqrt{\frac{C}{2\sigma_n^2}}, \sqrt{\frac{T_h}{\sigma_n^2}}\right) \quad (2.22)$$

where $\sigma_n^2 = N_0/4T$ is the post-integration noise variance with T as the integration time, $C/(2\sigma_n^2)$ is treated as post-integration SNR, and $Q_K(a, b)$ is the generalized Marcum Q-function, defined as (Cantrell and Ojha, 1987; Borio, 2008)

$$Q_K(a, b) = \frac{1}{a^{K-1}} \int_b^{+\infty} x^K \exp\left(-\frac{a^2 + x^2}{2}\right) I_{K-1}(ax) dx \quad (2.23)$$

with the modified Bessel function I_{K-1} of order $K - 1$. The plot of $P_d(T_h)$ versus $P_{fa}(T_h)$ will qualify the performance of the detector. The decision threshold, T_h , is usually chosen in order to provide a fixed false alarm probability P_{fa}^{target}

$$T_h = -2\sigma_n^2 \ln P_{fa}^{target} \quad (2.24)$$

where the noise variance σ_n^2 , in general, is unknown and can be estimated by correlating the input signal with an unused/fictitious PRN code. This method of obtaining σ_n^2 can guarantee the correlator output is a zero mean Gaussian random noise.

The acquisition process is essentially implemented based on detection and estimation theory. The acquisition block diagram is illustrated in Figure 2.8, where code delays are swept from zero to the code length and Doppler frequencies are swept in between the potential minimum and maximum Doppler, which are determined by the user velocity. An example of acquisition results in the presence or absence of the desired signal is given in Figure 2.9, where $|R_c(\tau, F_d)|^2$ has been calculated over the bi-dimensional grid of code delays and frequencies. By comparing $|R_c(\tau, F_d)|^2$ with the threshold T_h , the decision of whether signal is present can be made.

The acquisition process by sweeping over all possible code delays and Doppler frequencies is time and computation consuming. A parallel code/frequency one-dimensional search strategy can be used by taking advantage of Fast Fourier Transform (FFT). Exhaustive analysis about variable parallel search strategies can be found in Borre et al. (2007) and Borio (2008). In Borio (2008), more advanced detection theories are also described, which do not only depend on the statistical properties of the acquisition but also on different strategies (e.g., the parallel search strategy) adopted for the signal detection. This section only introduces the functionality of the acquisition process and an implementation example using the software-defined receiver for the proof of the concept.

The acquisition of the outer tones on S2 and S3 frequencies is skipped since they are not modulated by any PRN code and therefore the correlation peak will not

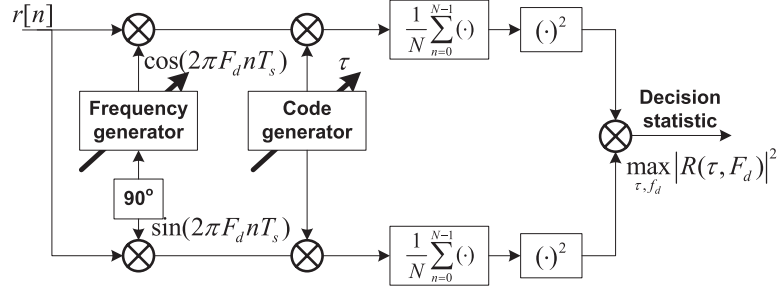
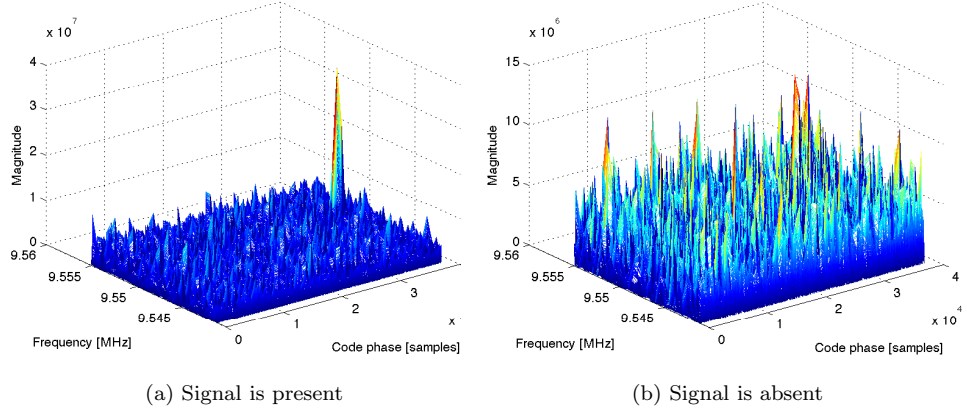


Figure 2.8: Acquisition block diagram

Figure 2.9: Acquisition result: BPSK-R(1) code, $f_{IF} = 9.548$ MHz, $f_s = 38.192$ MHz. Frequency is swept in 9.548 MHz \pm 10 kHz and code delay is swept from 0 to 38192 samples.

show up for acquisition. A simple way to obtain the coarse Doppler frequencies on S2 and S3 is to multiply the acquired S1 Doppler by the frequency ratios of f_2/f_1 and f_3/f_1 . This is based on the following equations about the Doppler effect

$$\frac{f_{d1}}{f_1} = \frac{f_{d2}}{f_2} = \frac{f_{d3}}{f_3} = \frac{\Delta v}{c} \quad (2.25)$$

where f_{d1} , f_{d2} and f_{d3} are the Doppler frequencies on S1, S2 and S3, respectively, and f_1 , f_2 and f_3 are three carrier frequencies. The relative velocity between satellites is denoted as Δv , and the speed of light is denoted as c .

2.3.3 Tracking

The code delay and Doppler frequency estimated by the acquisition block are too rough for being used for positioning and navigation. In order to obtain the fine estimates of code delay, phase, frequency and also track their changes, control is handed over to the signal tracking blocks.

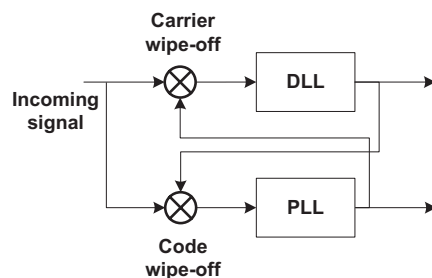


Figure 2.10: The coupled DLL and PLL

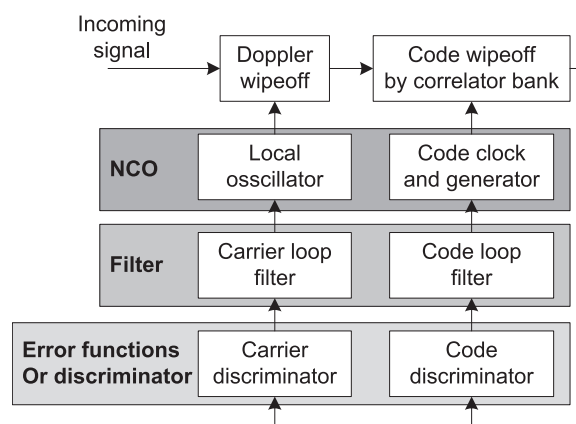


Figure 2.11: Basic DLL and PLL structures (Misra and Enge, 2001)

Tracking loops

Two separated tracking loops, namely the delay lock loop (DLL) and the phase lock loop (PLL), are coupled in the sense that the DLL (PLL) requires precise carrier (code) wipe-off for operating correctly. This is illustrated in Figure 2.10. The DLL and PLL can be modelled as control systems like the one shown in Figure 2.11. Both take measured correlations as input. Both have discriminators to strip out the desired error signals. Both use feedback to control the behavior of the local numerically controlled oscillator (NCO) so that the generated replica code/carrier will be aligned in time with the incoming signal.

The comprehensive DLL and PLL block diagrams for tracking the code delays and carrier phases on S1 are illustrated in Figure 2.12 and 2.13. If it is assumed that the code, frequency and phase replicas have misalignment $\delta\tau$, δf_{d_1} and $\delta\phi$, respectively, with respect to the incoming signal, the correlator output P , after the code and carrier wipe-off and after the integration over N samples, can be expressed as (Borio, 2012)

$$P = \sqrt{C/2R}(\delta\tau) \frac{\sin(\pi\delta f_{d_1}NT_s)}{N \sin(\pi\delta f_{d_1}T_s)} \exp\{j\delta\phi\} + \eta \quad (2.26)$$

where η is the residual noise term, which is zero mean, complex valued and Gaussian distributed. The real and imaginary parts of η comply with $\eta \sim \mathcal{N}(0, \sigma_n^2 \mathbf{I}_2)$, where

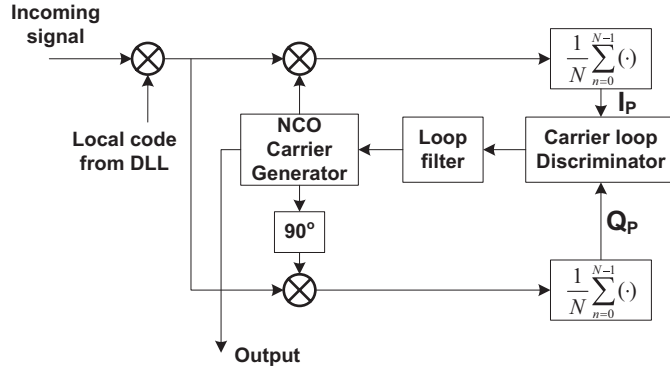


Figure 2.12: PLL block diagram

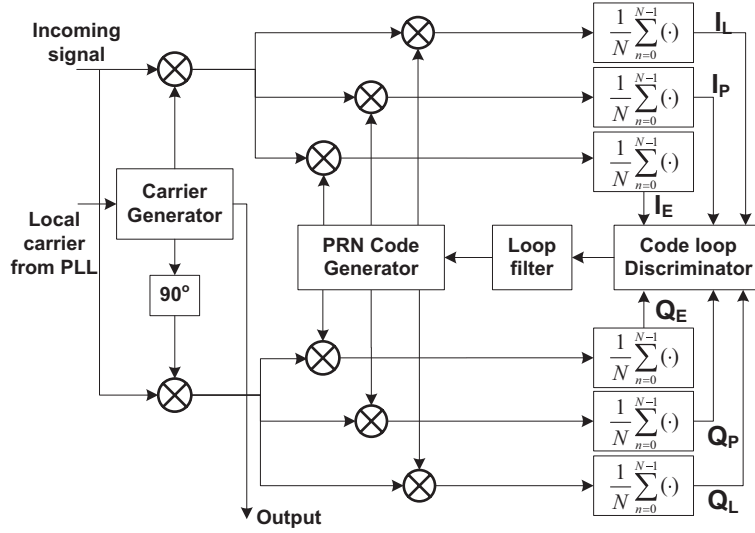


Figure 2.13: DLL block diagram

σ_n^2 is the noise variance.

Assuming a small frequency mismatch, the correlator output becomes

$$P = \sqrt{C/2R}(\delta\tau) \exp \{j\delta\phi\} + \eta . \quad (2.27)$$

This complex correlator is usually implemented as a pair of real correlators in the in-phase (I) arm and quadrature (Q) phase arms expressed as

$$\begin{aligned} I_P &= \sqrt{C/2R}(\delta\tau) \cos(\delta\phi) + \eta_I \\ Q_P &= \sqrt{C/2R}(\delta\tau) \sin(\delta\phi) + \eta_Q . \end{aligned} \quad (2.28)$$

Here, the subscript P means the prompt correlator as opposed to the early (E) or

late (L) correlators in the I and Q arms

$$\begin{aligned}
 I_E &= \sqrt{C/2}R(\delta\tau - dT_c/2) \cos(\delta\phi) + \eta_I \\
 Q_E &= \sqrt{C/2}R(\delta\tau - dT_c/2) \sin(\delta\phi) + \eta_Q \\
 I_L &= \sqrt{C/2}R(\delta\tau + dT_c/2) \cos(\delta\phi) + \eta_I \\
 Q_L &= \sqrt{C/2}R(\delta\tau + dT_c/2) \sin(\delta\phi) + \eta_Q
 \end{aligned} \tag{2.29}$$

where d is the early-late spacing in chips.

The phase misalignment can be determined from the arctangent of Q_P/I_P , which is treated as the discriminator function $D(\delta\phi)$ in the PLL

$$D(\delta\phi) = \arctan\left(\frac{Q_P}{I_P}\right) = \arctan\left(\frac{\sin(\delta\phi)}{\cos(\delta\phi)}\right). \tag{2.30}$$

The discriminator is a non-linear function of the error that the tracking loop is trying to minimize. Any misalignment in the replica carrier phase with respect to the incoming signal carrier phase produces a nonzero phase angle, so that the amount and direction of the phase change can be detected and corrected by the PLL. When the PLL is phase locked, the I component is maximum (signal plus noise) while the Q component is minimum (containing only noise).

Similarly, the misalignment of the code delay in the DLL is detected by the DLL discriminator $D(\delta\tau)$, which is designed as approximations of the derivative of the correlation function.

$$D(\delta\tau) \propto \frac{\partial R(\delta\tau)}{\partial \delta\tau} \approx \frac{R(\delta\tau - dT_c/2) - R(\delta\tau + dT_c/2)}{dT_c} \tag{2.31}$$

where \propto is an abbreviation for “proportional to”. This can be implemented by the subtraction between the early and late correlators, $D(\delta\tau) = I_E - I_L = (R(\delta\tau - dT_c/2) - R(\delta\tau + dT_c/2)) \cos(\delta\phi)$. However, this unfortunately shows dependency on phase errors $\delta\phi$. To obtain a phase independent discriminator, several non-linear discriminators can be applied, as indicated in Table 2.6.

Various types of PLL and DLL discriminators are summarized in Table 2.5 and 2.6. The choice of discriminator is dependent on the type of applications and the noise in the signal.

An example of the tracking loops is illustrated in Figure 2.14. From 2.14 (a), it can be seen that once the tracking loop is locked, the energy is kept in the I arm, while the Q arm only contains noise. The prompt correlator in the I arm keeps tracking the maximum point of the auto-correlation. The early and late correlators yield same outputs (except for the embedded noise) since any difference between them will be captured by the DLL discriminator and provided to the code NCO to speed-up or slow-down the local code replica generator. Similarly, the Q arm contains only noise because any error in the Q arm will be captured by the PLL discriminator and reported to the carrier NCO. The task of the PLL and DLL is to maintain the code and phase discriminator outputs zero, as illustrated in bottom figures of (b) and (c). The assumed Doppler frequency in the carrier is -537 Hz, which has been correctly tracked in the PLL. Similar to the carrier Doppler, the code Doppler is also present which represents the code chipping rate offset due to the relative dynamics between the satellite and user, see bottom figures in (b) and (c).

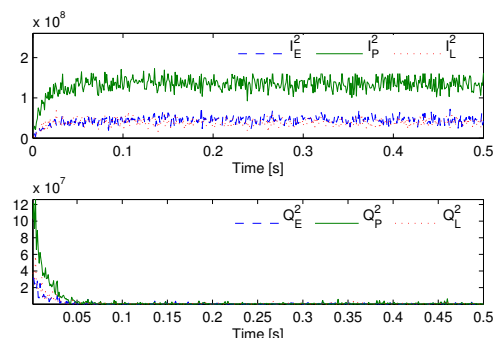
Table 2.5: PLL discriminator (Kaplan and Hegarty, 2006)

Discriminator function	Output phase error	Characteristics
$\arctan\left(\frac{Q_P}{I_P}\right)$	$\delta\phi$	Two-quadrant arctangent. Optimal ¹ at high and low SNR. Amplitude independent. Highest computational burden.
$\arctan 2(Q_P, I_P)$	$\delta\phi$	Four-quadrant arctangent. Optimal ¹ at high and low SNR. Amplitude independent. Sensitive to data transition. High computational burden.
$\frac{Q_P}{I_P}$	$\tan(\delta\phi)$	Suboptimal but good at high and low SNR. Amplitude independent. High computational burden.
$I_P \cdot Q_P$	$\sin(2\delta\phi)$	Standard Costas discriminator. Near optimal at high SNR. Proportional to signal amplitude squared. Moderate computational burden.
$\text{Sign}(I_P) \cdot Q_P$	$\sin(\delta\phi)$	Decision directed Costas discriminator. Near optimal at high SNR. Proportional to signal amplitude. Least computational burden.

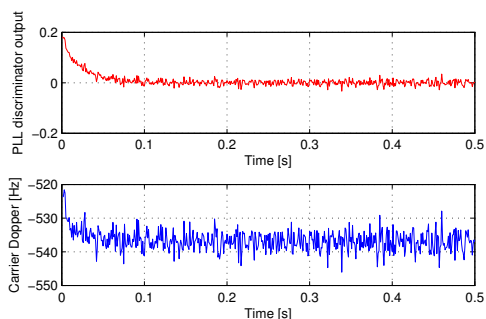
¹ Optimal: indicates maximum likelihood estimator.

Table 2.6: DLL discriminator (Kaplan and Hegarty, 2006)

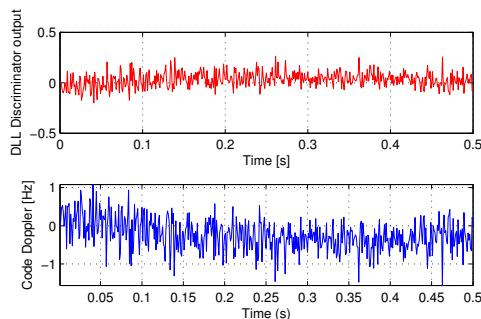
Discriminator function	Characteristics
$I_E - I_L$	Coherent early-minus-late (CELP). Requires phase lock conditions. Proportional to signal amplitude. Least computational load.
$(I_E - I_L)I_P + (Q_E - Q_L)Q_P$	Quasi-coherent dot product. Phase independent. Proportional to signal amplitude squared. Low computational load.
$(I_E^2 + Q_E^2) - (I_L^2 + Q_L^2)$	Non-coherent early-minus-late power (NELP). Phase independent. Proportional to signal amplitude squared. Moderate computational load.
$\frac{\sqrt{I_E^2 + Q_E^2} - \sqrt{I_L^2 + Q_L^2}}{\sqrt{I_E^2 + Q_E^2} + \sqrt{I_L^2 + Q_L^2}}$	Normalized non-coherent early-minus-late envelope. Phase independent. Signal amplitude independent. High computational load.



(a) Six correlator outputs in the in-phase and quadrature arms of the tracking loop.



(b) PLL tracking outputs



(c) DLL tracking outputs

Figure 2.14: The coupled PLL and DLL tracking results in the software defined receiver. The *arctan* PLL discriminator and *normalized NELP* DLL discriminator are used. Early-late spacing is 0.5 chips and the PLL and DLL loop noise bandwidth are 25 Hz and 2 Hz, respectively.

The tracking process on S2 and S3 frequencies consists of stand-alone PLLs since there is no PRN code to be wiped-off and therefore no coupling with DLLs. Only carrier phase measurements will be yielded on S2 and S3 frequencies.

Tracking accuracy

The tracking accuracy is determined by the statistic noise (e.g. thermal noise) and the systematic noise (e.g. multipath). The thermal noise on the carrier phase measurement in an arctangent PLL is computed as (Kaplan and Hegarty, 2006)

$$\sigma_{\text{PLL}}^2 = \frac{B_L}{C/N_0} \left(1 + \frac{1}{2TC/N_0} \right) [\text{rad}^2] \quad (2.32)$$

where T is the integration time, B_L is the equivalent tracking noise bandwidth in Hz, and C/N_0 is the carrier to noise ratio. Note that the carrier noise is independent of the code type.

Approximations for the thermal noise on code measurement using the BPSK-R modulation are provided in Betz and Kolodziejcki (2000), where three cases of particular interest are analysed in Eq.(2.33) under the constraint that $\beta_r T_c > 1$, i.e, the front-end filter has a bandwidth β_r that will at least pass the center half of the

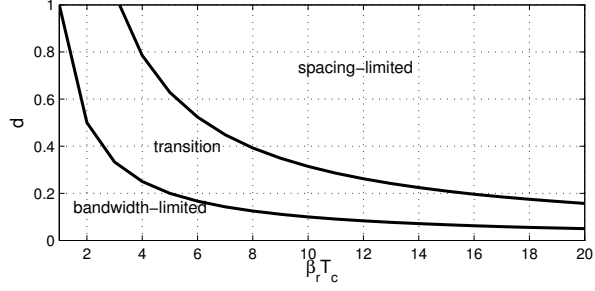


Figure 2.15: Three regions where different approximations of the code tracking error apply (Betz and Kolodziejcki, 2000)

mainlobe of the signal spectrum

$$\sigma_{\text{DLL}}^2 = \begin{cases} \frac{B_L(1 - 0.5B_L T)}{2C/N_0} d, & dT_c\beta_r \geq \pi \\ \frac{B_L(1 - 0.5B_L T)}{2C/N_0} \left[\frac{1}{\beta_r T_c} + \frac{\beta_r T_c}{\pi - 1} \left(d - \frac{1}{\beta_r T_c} \right)^2 \right], & 1 < dT_c\beta_r < \pi \\ \frac{B_L(1 - 0.5B_L T)}{2C/N_0} \left(\frac{1}{\beta_r T_c} \right), & dT_c\beta_r \leq 1 \end{cases} \quad (2.33)$$

where β_r is the front-end bandwidth, σ_{DLL} is in unit of chips. These approximations apply in a coherent early-minus-late DLL, while larger thermal noise can be expected in a non-coherent DLL (Betz and Kolodziejcki, 2000).

Three different regions are illustrated in Figure 2.15 where different approximations of the code thermal noise apply. The condition $dT_c\beta_r \geq \pi$ is referred to as *spacing-limited*, since the bandwidth satisfies $\beta_r T_c > 3$ for all values in $0 < d \leq 1$, indicating that the complete mainlobe of the signal spectrum (or the majority of the power) has been captured and the error thus primarily depends on the early-late spacing, not the bandwidth. This also indicates that narrow spacing should be accompanied by large front-end bandwidth to avoid rounding of the correlation peak in the region where the narrow correlators are being operated. In fact, there is no benefit to further reducing d to less than the reciprocal of the $T_c\beta_r$, which is the condition referred to as *bandwidth-limited*. The error in this condition depends primarily on the front-end bandwidth on the contrary. In this case, the early-late processing provides a near optimal spacing-independent code tracking accuracy that a coherent discriminator can achieve. The condition $1 < dT_c\beta_r < \pi$ indicates the *transition* region between the two distinct extreme cases.

BOC tracking techniques

Since the BOC-modulated signal has multiple closely spaced peaks in its auto-correlation function, it can be difficult to distinguish the main peak from side peaks due to noise. The time delay of the first side peak, combined with its magnitude ratio to the main peak indicates the degree to which the DLL may have difficulty maintaining track of the true peak. Using the BOC-modulated signal means to cope with multiple peak ambiguities in the DLL. Several techniques have been proposed to solve the problem.

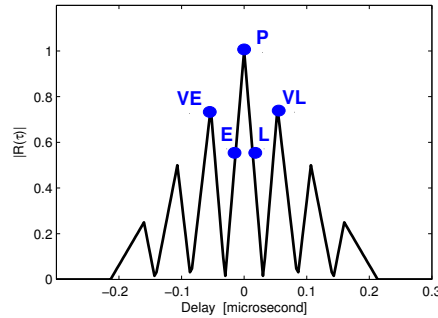


Figure 2.16: Bump-jump method for the BOC(10,5) signal using extra correlators for detecting adjacent peaks in the squared correlation of a NELS DLL discriminator

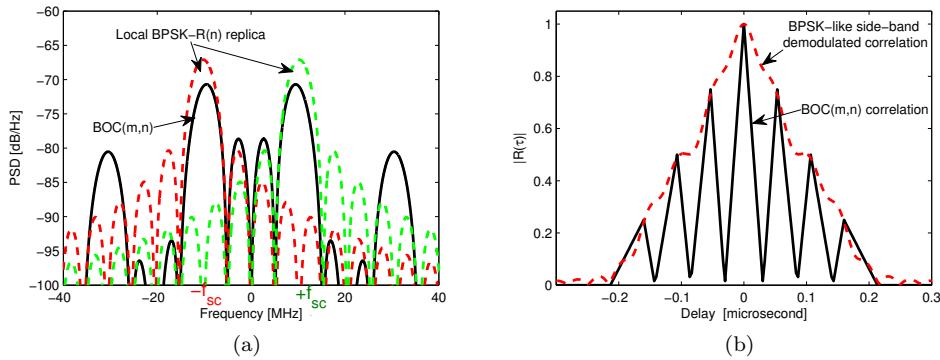


Figure 2.17: BPSK-like technique for the BOC(10,5) signal. (a) Treating the BOC(m,n) signal as the sum of two BPSK-R(n) signals located symmetrically on the BOC subcarriers; (b) Unambiguous correlation function after the BPSK-like side-band demodulation.

One technique suggested by Fine and Wilson (1999), called “Bump-jump” method, is used to detect and correct false locks by introducing a couple of extra correlators, designated as very-early (VE), very-late (VL) correlators as depicted in Figure 2.16. Under correct lock conditions, the VE and VL correlators detect the squared amplitude of adjacent peaks, which shall be consistently lower than the prompt peak. The Bump-jump method thus consists in measuring and comparing these adjacent peaks with respect to the currently tracked peak and jumping left or right depending on the comparison result, until a maximum is found.

Other techniques, e.g. side-band filtering (Fisman et al., 2000) and the BPSK-like technique (Martin et al., 2003), are faster and simpler to implement. These techniques focus on eliminating peak ambiguities in a so-called *transition process* before going to the classical tracking processing. In this transition process, the BOC(m,n) signal is treated as the sum of two BPSK-R(n) signals with carrier frequencies symmetrically positioned on the negative and positive BOC subcarriers $\pm f_{sc}$. Thus the local BPSK-R(n) replica shifted by $\pm f_{sc}$ in the transition process can demodulate a single lobe independently and provides unambiguous correlation function, as shown in Figure 2.17. This unambiguous correlation is continuously tracked in the transition process until an unbiased local code delay converges toward the received code.

The accuracy of the code convergence at this stage is lower than the case of the nominal BOC tracking, which will be implemented afterwards in order to provide the full BOC performance without biases. These techniques have a drawback that the transition process has to be repeatedly implemented in a highly noisy or dynamic environment where the signal easily loses the lock on the nominal mainpeak track and has to recover from the transition process.

Recognizing that the BOC signal is the combination of the periodic subcarrier and BPSK-R code, the unambiguous correlation can also be created if they have been treated independently and the code wipe-off is spilt into the subcarrier wipe-off and the BPSK-R code wipe-off. This technique is called double estimator, which introduces two separate code tracking loops, one being the DLL and the other being the subcarrier lock loop (SLL) (Hodgart and Blunt, 2007; Hodgart et al., 2008). The final code delay estimation is then the function of the delays from both the DLL and SLL. It is more robust than the other aforementioned techniques since the tracking loops are automatically adaptive to changes of the incoming signals.

2.3.4 Lower bound of code tracking accuracy

A lower bound code tracking accuracy for white noise has been obtained in Betz (2001) and Betz and Kolodziejcki (2009), which is based on the performance of a maximum-likelihood estimator of the time of arrival using T seconds of data, driving a tracking loop. This lower bound does not depend on the early-late spacing or the type of discriminator, but instead relates how well any discriminator can perform with a given front-end bandwidth β_r , a given carrier to noise ratio C/N_0 and a given equivalent DLL tracking noise bandwidth B_L

$$\sigma_{LB}^2 = \frac{B_L(1 - 0.5B_L T)}{(2\pi)^2 \frac{C}{N_0} \int_{-\beta_r/2}^{\beta_r/2} f^2 G(f) df} \quad (2.34)$$

where $G(f)$ is the power spectral density normalized to unit area over infinite bandwidth, $\int_{-\infty}^{\infty} G(f) df = 1$, $G(f)$ is expressed in closed-form in Eq.(2.9) for the BPSK-R modulation and in Eq.(2.10) for the BOC modulation.

Table 2.7: Lower bound of the code tracking accuracy. The integration time T is equal to one sequence length, $T = lT_c$

Modulation	C/N_0 [dB/Hz]	One-sided code tracking noise bandwidth [Hz]	Front-end bandwidth [MHz]	Out-of-band power loss	Code tracking accuracy [m]
BPSK-R(1)	45	1	24	0.0085	0.2391
	45	0.5	24	0.0085	0.1691
	45	0.5	8	0.0253	0.2917
BOC(1,1)	45	1	24	0.0253	0.1374
	45	0.5	24	0.0253	0.0972
	45	0.5	8	0.0749	0.1683
BPSK-R(10)	45	1	24	0.0946	0.0812
	45	0.5	24	0.0946	0.0574
	45	0.5	8	0.3344	0.1082
BOC(10,5)	45	1	24	0.2370	0.0339
	45	0.5	24	0.2370	0.0240
	45	0.5	8	0.9344	0.2812

Table 2.7 shows the lower bound of the code tracking accuracy for different modulations. It is clear that at the same noise level with the same tracking loop bandwidth and the same front-end bandwidth, a higher chipping rate leads to a higher ranging accuracy. For the codes with the same chipping rate, the BOC modulation assures higher accuracy than the BPSK-R modulation. These two conclusions stand true only if the front-end bandwidth is large enough to capture the power spectrum mainlobe, which is centred at zero frequency in baseband for the BPSK-R modulation but shifted to the sub-carrier frequencies $\pm f_{sc}$ for the BOC modulation. The out-of-band power loss in the penultimate column indicates the amount of power that is lost by bandlimiting. The larger is the out-of-band power loss, the less is the in-band power available to a receiver. For instance, the out-of-band power loss for the 8 MHz bandlimited BOC(10,5) is 93%, which is so large that the benefits of its high chipping rate and BOC pulse shape have vanished due to the improper bandlimiting.

2.3.5 Multipath effects

Multipath refers to the phenomenon of a signal reaching an antenna via two or more paths: a direct line-of-sight path and one or more of its reflections from structures in the vicinity. A reflected signal is delayed and usually weaker than the direct signal. Code and carrier phase measurements in the presence of multipath are the sum of the direct and the reflected signals. Unlike terrestrial applications where the multipath effect is caused by reflections from relatively far-away objects like buildings and trees, the RF-based ranging system on-board the spacecraft in space suffer from short-delay multipath that is reflected from the surfaces on the spacecraft itself or from other spacecraft during the operations of rendezvous and docking.

The measure of multipath immunity is built into the PRN signal structure. A reflected signal which is delayed by more than 1.5 chips (e.g., 450 m for BPSK-R(1) and 45 m for BPSK-R(10)) would be suppressed automatically in the correlation process in a receiver because the auto-correlation for the PRN code is nearly zero for delays longer than 1.5 chips. The benefit of using a higher chipping rate code is to assure a greater multipath immunity (Misra and Enge, 2001). A reflected signal delayed by less than 1.5 chips will degrade the code tracking accuracy to several meters and phase accuracy to several centimetres. The underlying reason is elaborated as follows.

In the absence of multipath, the outputs of the CELP DLL discriminator and arctan PLL discriminator are

$$\begin{aligned} D(\delta\tau) &= I_P - I_L = A_0 [R(\delta\tau + dT_c/2) - R(\delta\tau - dT_c/2)] \cos(\delta\phi) \\ D(\delta\phi) &= \arctan\left(\frac{Q_P}{I_P}\right) = \arctan\left(\frac{\sin(\delta\phi)}{\cos(\delta\phi)}\right) \end{aligned} \quad (2.35)$$

where A_0 is denoted as the signal amplitude. The task of the DLL and PLL is to maintain the code and phase discriminator outputs zero. Solutions of $D(\delta\tau) = 0$ and $D(\delta\phi) = 0$ are straightforward in the absence of multipath: $\delta\tau = 0$ and $\delta\phi = 0$.

In the presence of multipath, the receiver tracking process continues to follow the rules of $D(\delta\tau) = 0$ and $D(\delta\phi) = 0$, but the values of $\delta\tau$ and $\delta\phi$ that fulfil them are no longer zeros, indicating that the loops are no more tracking the direct signal, but a combination of the direct and the reflected ones. This yields errors on code and phase estimates.

More specifically, adding a multipath component, the received signal on the S1 frequency after ADC will be changed to

$$\begin{aligned} r[n] &= A_0 c(nT_s - \tau_0) \cos(2\pi(f_{IF_1} + f_{d_1})nT_s - \varphi_0) \\ &+ A_1 c(nT_s - (\tau_0 + \tau_1)) \cos(2\pi(f_{IF_1} + f_{d_1})nT_s - (\varphi_0 + \psi_1)) + \eta[n] \end{aligned} \quad (2.36)$$

where a multipath signal has been superimposed on the LOS signal with an amplitude of A_1 , an extra delay of τ_1 and an extra phase of ψ_1 . Note that τ_1 is always positive since the multipath always arrives later than the LOS signal, and ψ_1 is a function of τ_1 , $\psi_1 = 2\pi f\tau_1$. With this new received signal, the set of correlator and discriminator outputs becomes

$$\begin{aligned} I_P &= A_0 R(\delta\tau) \cos(\delta\phi) + A_1 R(\delta\tau - \tau_1) \cos(\delta\phi - \psi_1) \\ Q_P &= A_0 R(\delta\tau) \sin(\delta\phi) + A_1 R(\delta\tau - \tau_1) \sin(\delta\phi - \psi_1) \\ D(\delta\tau) &= A_0 [R(\delta\tau + dT_c/2) - R(\delta\tau - dT_c/2)] \cos(\delta\phi) \\ &+ A_1 [R(\delta\tau - \tau_1 + dT_c/2) - R(\delta\tau - \tau_1 - dT_c/2)] \cos(\delta\phi - \psi_1) \\ D(\delta\phi) &= \arctan \left(\frac{A_0 R(\delta\tau) \sin(\delta\phi) + A_1 R(\delta\tau - \tau_1) \sin(\delta\phi - \psi_1)}{A_0 R(\delta\tau) \cos(\delta\phi) + A_1 R(\delta\tau - \tau_1) \cos(\delta\phi - \psi_1)} \right) \\ &\approx \arctan \left(\frac{A_1 R(\tau_1) \sin \psi_1}{A_0 + A_1 R(\tau_1) \cos \psi_1} \right). \end{aligned} \quad (2.37)$$

Approximations in $D(\delta\phi)$ hold true as the code/phase errors due to multipath are negligible compared to the multipath delay that causes these errors.

Figure 2.18 shows the auto-correlation functions for the BPSK-R(1) modulation when $\tau_1 = 0.5$ chips, $\psi_1 = 0^\circ$ or 180° and multipath-to-signal amplitude ratio (MSR) $A_1/A_0 = 0.5$. The signal strength, represented by reading of the punctual correlator output I_P , is changed by multipath. Positive multipath ($0^\circ \leq \psi_1 \leq 90^\circ$) enlarges the signal strength while negative multipath ($90^\circ < \psi_1 \leq 180^\circ$) decreases the signal strength. The code discriminator outputs for the positive or the negative multipath are also illustrated in Figure 2.19. It is clear that the central stable tracking point of discriminator is no longer $(0, 0)$, leading to a remaining tracking bias due to multipath. Multipath error is then computed measuring the bias between the point at $(0, 0)$ and the point of the real zero-crossing of the discriminator function.

Figure 2.18 and 2.19 are examples for a specific multipath delay. A more common and recognized way to analyse multipath is to vary all the geometric multipath delays from 0 to 1.5 chips with a given amplitude ratio. Rules of keeping $D(\delta\tau) = 0$ and $D(\delta\phi) = 0$ with the definitions given by Eq.(2.37) are used to compute the code and phase multipath errors. As shown in Figure 2.20, multipath errors exhibit periodic oscillations due to the change of multipath phase $\psi_1 = 2\pi f\tau_1$. The code multipath envelope can be considered as the upper and lower bounds of the multipath errors when multipath phase is equal to 0° and 180° . The phase multipath also has an envelope, but it is the extreme when multipath phase is 90° or 270° . Code and phase multipath errors have an out-of-phase (quadrature) relationship (Sleewaegen, 1997).

Figure 2.21 compares the multipath error envelopes for the BPSK-R modulated codes with different chipping rates and early-late spacings. It is clear that for the multipath delayed more than 1.5 chips (e.g., 450 m for BPSK-R(1) and 45 m for BPSK-R(10)), the multipath error is zero. However, for the short-delay multipath, the error is independent of the type of modulation and correlator spacing. The

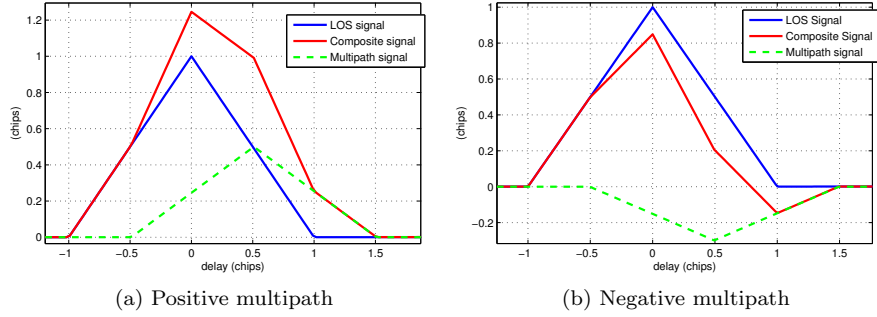


Figure 2.18: BPSK-R(1) correlator output with one reflected signal. Multipath-to-signal amplitude ratio $A_1/A_0 = 0.5$, multipath time delay $\tau_1 = 0.5$ chips and phase delay $\psi_1 = 0^\circ$ in (a) and 180° in (b). Early-late spacing d is 0.1 chips.

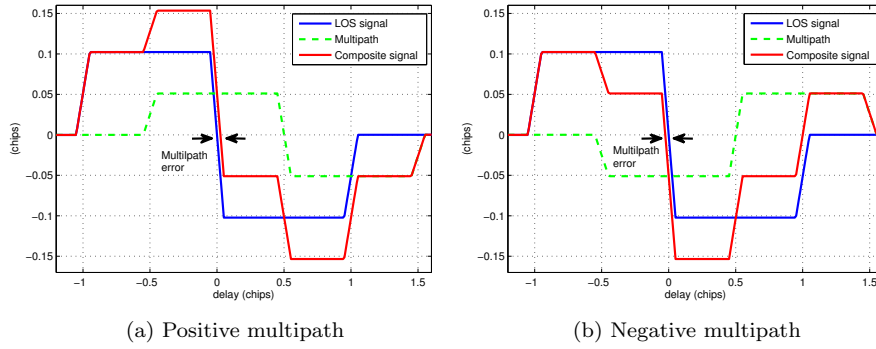


Figure 2.19: BPSK-R(1) code coherent discriminator output with one reflected signal. Multipath-to-signal amplitude ratio $A_1/A_0 = 0.5$, multipath time delay $\tau_1 = 0.5$ chips and phase delay $\psi_1 = 0^\circ$ in (a) and 180° in (b). Early-late spacing d is 0.1 chips.

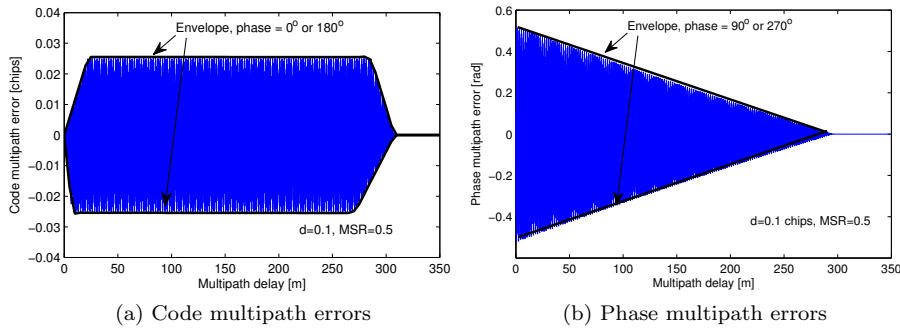


Figure 2.20: BPSK-R(1) multipath error exhibits periodic swings between the indicated upper and lower bounds in function of multipath delays. The code envelope shows the extremes of both the positive and negative errors when multipath phase is 0° and 180° , while the phase multipath extremes occur when multipath phase is 90° and 270° . They show out-of-phase relationship.

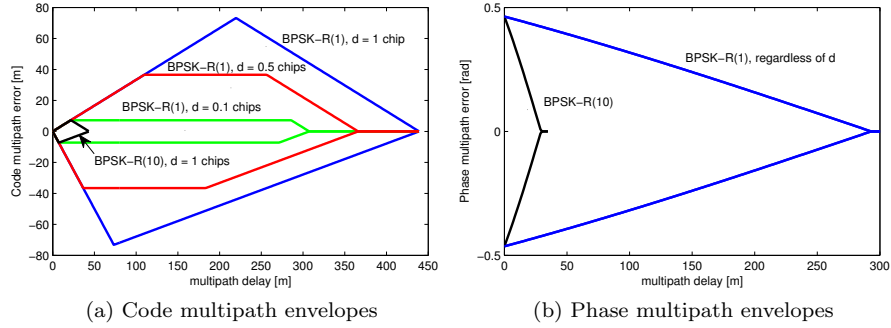


Figure 2.21: Code and phase multipath envelopes with respect to different chipping rates and early-late spacings. The actual error periodically swings between the upper and lower bounds as the relative phase changes. $A_1/A_0 = 0.5$

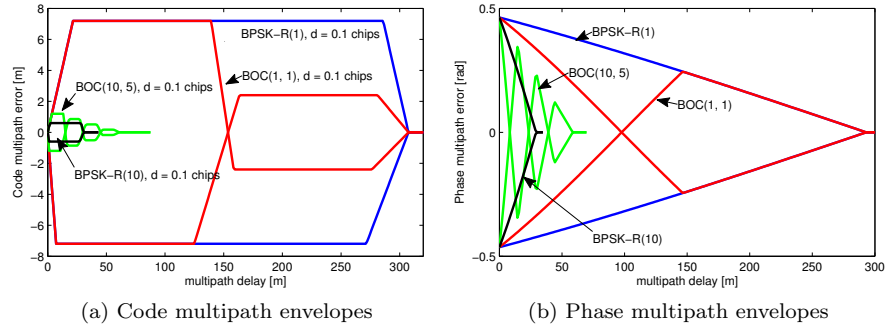


Figure 2.22: Code and phase multipath envelopes for various BPSK-R and BOC modulations. The actual error periodically swings between the upper and lower bounds as the relative phase changes. $A_1/A_0 = 0.5$

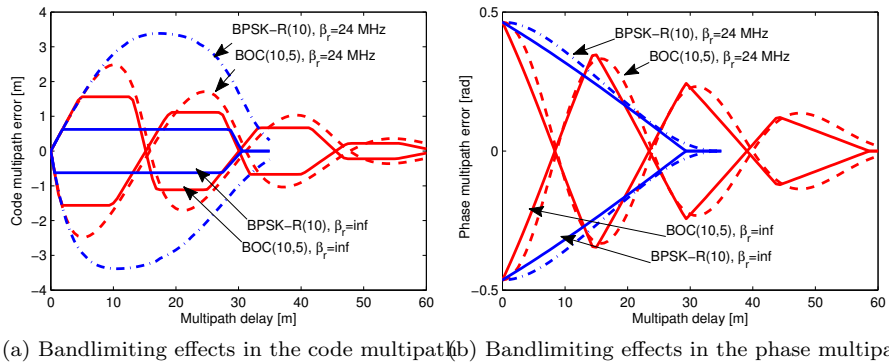


Figure 2.23: Bandlimited code and phase multipath envelopes for the BPSK-R(10) and BOC(10,5) modulated codes. The actual error periodically swings between the upper and lower bounds as the relative phase changes. $A_1/A_0 = 0.5$, $d = 0.1$ chips.

equation of the short-delay multipath error is (van Nee, 1995)

$$\delta\tau = \frac{A_1 \cos(\psi_1)\tau_1}{A_0 + A_1 \cos(\psi_1)}, \quad 0 \leq \tau_1 < a \quad (2.38)$$

where a is the transition point in the multipath envelope where the linearly increased error turns into the constant

$$a = \frac{A_0 \pm A_1}{2A_0} dT_c, \quad \begin{cases} + & \text{positive multipath} \\ - & \text{negative multipath} \end{cases} \quad (2.39)$$

Although the short-delay multipath error in Eq.(2.38) is independent of the type of modulation and correlator spacing. The transition point of (2.39), however, shows that reducing d or T_c shortens the error growing period and consequently guarantees better multipath rejection performance. However, smaller d or T_c require wider front-end bandwidth to prevent the correlation peak from being severely rounded.

In Figure 2.21, the carrier phase multipath envelope linearly decreases with the multipath delay. Positive and negative phase errors are symmetrical. The phase error is irrelevant to the correlator spacing. This can be explained by Eq.(2.37), which is calculated using the in-phase and quadrature prompt correlators and has no relation to early or late correlators.

Figure 2.22 compares the multipath envelopes for the BPSK-R and BOC modulated codes. The BOC(m,n) modulation only outperforms BPSK-R(n) modulation for the long or medium delayed multipath, while for the short-delay multipath, they show similar performance at the same chipping rate of $n \times 1.023$ MHz. The BOC(1,1) and BOC(10,5) have 2 and 4 ($2m/n$) square waves in a chip, respectively, resulting in twice or 4 times envelope fluctuations.

Figure 2.23 illustrates the bandlimited multipath envelopes. As depicted, bandlimiting has significant effects to the code multipath errors but only slightly influences the phase errors.

The multipath immunity capability of the given code is an crucial criterion for evaluating its eligibility to be used in the RF-based relative navigation system. The following section will compare multipath immunity performance for different codes. Multipath mitigation methods will also be proposed in chapter 4 and 5 in order to further reduce its impacts on the navigation accuracy as well as to minimize the difficulty of the carrier phase integer ambiguity resolution in the presence of multipath.

2.3.6 Evaluation of BPSK-R and BOC codes

In order to choose an eligible code among different candidates in code families for the RF-based inter-satellite ranging system, many aspects should be evaluated, mainly including the code tracking accuracy in thermal noise, multipath immunity capability and the complexity of the code acquisition and tracking process based on the consideration of the overall inter-satellite system design.

Table 2.8 provides the characteristics of various codes and also the criteria for evaluating them. The lower bound of code tracking accuracy is the first criterion as it represents the best achievable accuracy for a given code in thermal noise. Multipath error is also evaluated specifically for the very-short-delay multipath for applications in space. For this type of multipath, the code error envelope linearly grows with the multipath delay, and the negative multipath grows faster than the

Table 2.8: Characterizing the BPSK-R and BOC-modulated signals for the RF-based inter-satellite ranging system

Criteria	Characteristic	BPSK-R(1)	BOC(1,1)	BPSK-R(10)	BOC(10,5)
Thermal noise ¹	Lower bound of tracking accuracy [m]	0.169	0.097	0.057	0.024
Multipath ²	Code multipath delayed by less than 4 m [m]	[-3.79,1.33]	[-3.79,1.33]	[-2.67,1.29]	[-2.39,1.29]
	Phase multipath delayed by 4 m [cm]	[-0.97,0.97]	[-0.96,0.96]	[-0.93,0.93]	[-0.73,0.73]
	Maximum phase multipath [cm]	± 0.97	± 0.97	± 0.97	± 0.97
Receiver complexity	Time delay of the first sidepeak from the mainpeak [chips]	-	0.5	-	0.25
	Ratio of the first sidepeak to mainpeak normed amplitude	-	0.5	-	0.75
	Bandwidth needed to include mainlobes [MHz]	2	4	20	30
	90% power bandwidth [MHz]	1.8	6.3	17.4	64.5

¹ Computed with the front-end bandwidth of 24 MHz, the tracking loop noise bandwidth of 0.5 Hz and the signal to noise ratio of 45 dB/Hz.

² Computed with the front-end bandwidth of 24 MHz, the early-late spacing of 0.1 chips, the multipath-to-signal amplitude ratio of 0.5, and carrier frequency of 2271.06 MHz in S-band. Multipath could be much more severe if its relative amplitude ratio is larger than 0.5, or is processed in a receiver with either a smaller bandwidth or larger early-late spacing.

positive multipath. Therefore, the maximum code multipath error occurs when it is caused by a negative multipath. On the contrary, the carrier phase multipath envelope linearly decreases with the multipath delay, indicating that the largest phase multipath error (approximately 0.97 cm in given specifications in Table 2.8) is caused by the extremely short multipath, which is the case in space. The positive and negative phase multipath are symmetrical.

The BPSK-R(10) and BOC(10,5) codes have higher chipping rate than the BPSK-R(1) and BOC(1,1), providing higher ranging accuracy in thermal noise and better multipath immunity. The BOC(10,5) has $k=4$ ($k = 2m/n$) square waves in one chip, resulting in 4 times of multipath envelope fluctuations and thus smaller errors than the non-fluctuating BPSK-R(10). However, for the multipath delayed by less than 4 m, the BPSK-R(10) and BOC(10,5) codes have comparable immunity performance. This is due to the fact that the BOC(10,5) multipath with 4 m delay has not yet approached its first fluctuation point, which occurs at approximately 15 m delay in its code multipath envelope, see Figure 2.23 (a).

Another criterion to evaluate various codes is to compare the receiver complexity, which involves the degree of difficulty to track of the peak and the necessary bandwidth to maintain the code tracking capability. The time delay of the first side peak in the auto-correlation function, combined with its magnitude ratio with respect to the main peak, indicates the degree to which receivers may have difficulty maintaining track of the main peak. Side peaks that are close both in delay and in magnitude, e.g., the BOC(10,5), makes this challenging. In fact, the higher is $k = 2m/n$ in a BOC(m,n) code, the harder is the code tracking process, as extra

processing has to be implemented by either using extra correlators or extra loops, see section 2.3.3. In this aspect, the BPSK-R modulation is superior to the BOC modulation since tracking and maintaining on a single peak do not involve any extra processing.

In addition, wider bandwidth is required for the BOC(m,n) modulation as its mainlobes are split and shifted to the location of subcarrier frequencies at $\pm f_{sc}$. The necessary bandwidth to include mainlobes are wider for higher subcarrier. A 90% power bandwidth is also characterized, which indicates a bandwidth to pass 90% of the signal power. As shown in the Table, the BPSK-R(1) and BPSK-R(10) requires 1.8 MHz and 17.4 MHz, respectively, to pass the 90% power, which are smaller than their mainlobe bandwidth of 2 MHz and 20 MHz. On the contrary, for the BOC(1,1) and BOC(10,5), the 90% power bandwidths are much wider than their mainlobe bandwidths, indicating that both mainlobes and several sidelopes are needed to reach 90% power. Not only the front-end bandwidth increases for the BOC code, the associated ADC sampling rate needs also to be increased.

Given the aforementioned criteria, the BPSK-R(10) is finally chosen as the ranging code for the inter-satellite relative navigation system. On one hand, it provides higher ranging accuracy and better multipath immunity capability than the codes of lower chipping rates, e.g., BPSK-R(1) or BOC(1,1). On the other hand, it does not require a complicated code tracking loop design as in the case of the BOC(10,5) since the single peak auto-correlation function avoids the process of distinguishing the main peak from side peaks. The required bandwidth and the associated ADC sampling rate are also smaller than the BOC(10,5) modulation so that the front-end design becomes easier and the power consumption becomes less.

2.4 Code and phase observations

2.4.1 Undifferenced observation model

Two primary observables are available after the acquisition and tracking process in the receiver: the code observable (also called pseudorange) and the carrier phase observable. Both the observables are affected by a number of errors, including the errors generated at the transmitter (transmitter clock errors, instrumental delays), at the receiver (receiver clock errors, instrumental delays and the aforementioned multipath and thermal noise) and the errors caused by the transmitting media (ionospheric and tropospheric delays). Taking into account the full set of errors, the code and phase observations can be described as (Teunissen and Kleusberg, 1998; Giorgi, 2011)

$$\begin{aligned}
 \rho_{r,f}^s(t) &= r_{\rho,r}^s(t, t - \tau_r^s) + I_{r,f}^s + T_r^s + c[dt_r(t) - dt^s(t - \tau_r^s)] \\
 &\quad + lb_{\rho,r,f}(t) + lb_{\rho,f}^s(t - \tau_r^s) + \delta\tau_{mp_{r,f}} + \varepsilon_{\rho,r,f} \\
 \phi_{r,f}^s(t) &= r_{\phi,r}^s(t, t - \tau_r^s) - I_{r,f}^s + T_r^s + c[dt_r(t) - dt^s(t - \tau_r^s)] \\
 &\quad + lb_{\phi,r,f}(t) + lb_{\phi,f}^s(t - \tau_r^s) + \lambda_f[\theta_{r,f}(t_0) - \theta_f^s(t_0)] \\
 &\quad + \lambda_f N_{r,f}^s + \delta\phi_{mp_{r,f}} + \varepsilon_{\phi,r,f}
 \end{aligned} \tag{2.40}$$

where the superscript s indicates the transmitting spacecraft and the subscripts r and f indicate the receiver and the carrier frequency, respectively. Other terms

denote:

ρ, ϕ	code and phase observations [m]
r_ρ, r_ϕ	true range between receiver and transmitter [m]
τ_r^s	signal travel time from transmitter to receiver [s]
I, T	ionospheric and tropospheric delays [m]
dt_r	receiver clock error [s]
dt^s	transmitter clock error [s]
c	speed of light [m/s]
lb_r	receiver instrumental delays, mainly including the line bias [m]
lb^s	transmitter instrumental delays [m]
$\theta_r(t_0)$	initial phase of the generated replica carrier signal [cycle]
$\theta^s(t_0)$	initial phase of the original transmitter carrier signal [cycle]
t_0	time of reference for phase synchronization [s]
λ_f	signal wavelength at frequency f [m]
N	number of complete carrier phase cycles (integer ambiguity) [cycle]
$\delta\tau_{mp}, \delta\phi_{mp}$	code and phase multipath errors [m]
$\varepsilon_\rho, \varepsilon_\phi$	remaining unmodeled code and phase thermal noise [m].

The effects of the atmosphere on the signal (e.g., delay, bending and reflections) are reflected in the ionospheric (I) and tropospheric (T) delays. The troposphere stretches to about 16 km at the equator and 9 km above the poles (Misra and Enge, 2001). Since the spacecraft's orbit is normally not close to the Earth's surface, the tropospheric delay is normally not a considerable error source unless the navigation processor requires argumentations from the ground station. Whereas the ionosphere from about 85 km to 1000 km altitude is a dominant error source, especially for the spacecraft in LEO. Note also that the sign of the ionospheric term in the carrier phase is negative whereas in the pseudorange it is positive. This is because the ionosphere, as a dispersive medium, slows down the speed of propagation of signal modulations (the PRN codes and the communication data) to below the vacuum speed of light whereas the speed of propagation of the carrier is actually increased beyond the speed of light in vacuum (Teunissen and Kleusberg, 1998).

The effects of the non-perfect time synchronization between receiver and transmitter is captured by the term dt_r and dt^s , which indicate the clock errors. Instrumental delays in transmitter and receiver are expressed by the term lb_r and lb^s . The line bias is embedded in lb_r which indicates the physical length of the cable from the antenna to the receiver. Both the code and carrier phase observations are also affected by multipath errors that are expressed by the term $\delta\tau_{mp}$ and $\delta\phi_{mp}$, respectively.

The mathematical expression of the carrier phase observation is very similar to the pseudorange observation - the major difference being the presence of the integer cycle ambiguity N and the initial phases of the original and replica carriers $\theta^s(t_0)$ and $\theta_r(t_0)$.

Some of the error sources are independent of the carrier frequency. They include the receiver and transmitter satellite clock errors as well as the tropospheric delay effect. All other terms will in general be different for different carrier frequencies.

2.4.2 Single-differenced model between receivers/antennas

Single-differenced (SD) models may be formed by differencing observations from two different receivers/antennas, transmitters, frequencies, epochs or observations. For

the inter-satellite LOS estimation, the SD between two receivers/antennas on-board the spacecraft is required and will thus be presented in this section. An exhaustive analysis of other SD models used in GNSS applications can be found in Teunissen and Kleusberg (1998).

The SD pseudorange and carrier phase model between two receivers/antennas r_1 and r_2 for the same transceiver spacecraft s can be written as

$$\begin{aligned}
\rho_{r_1,f}^s(t) - \rho_{r_2,f}^s(t) &= r_{\rho,r_1}^s(t, t - \tau_{r_1}^s) - r_{\rho,r_2}^s(t, t - \tau_{r_2}^s) \\
&\quad + I_{r_1,f}^s - I_{r_2,f}^s + T_{r_1}^s - T_{r_2}^s \\
&\quad + c[dt_{r_1}(t) - dt^s(t - \tau_{r_1}^s)] - c[dt_{r_2}(t) - dt^s(t - \tau_{r_2}^s)] \\
&\quad + [lb_{\rho,r_1,f}(t) + lb_{\rho,f}^s(t - \tau_{r_1}^s)] - [lb_{\rho,r_2,f}(t) + lb_{\rho,f}^s(t - \tau_{r_2}^s)] \\
&\quad + \delta\tau_{mp_{r_1,f}} - \delta\tau_{mp_{r_2,f}} + \varepsilon_{\rho,r_1,f} - \varepsilon_{\rho,r_2,f} \\
\phi_{r_1,f}^s(t) - \phi_{r_2,f}^s(t) &= r_{\phi,r_1}^s(t, t - \tau_{r_1}^s) - r_{\phi,r_2}^s(t, t - \tau_{r_2}^s) \\
&\quad - I_{r_1,f}^s + I_{r_2,f}^s + T_{r_1}^s - T_{r_2}^s \\
&\quad + c[dt_{r_1}(t) - dt^s(t - \tau_{r_1}^s)] - c[dt_{r_2}(t) - dt^s(t - \tau_{r_2}^s)] \\
&\quad + [lb_{\phi,r_1,f}(t) + lb_{\phi,f}^s(t - \tau_{r_1}^s)] - [lb_{\phi,r_2,f}(t) + lb_{\phi,f}^s(t - \tau_{r_2}^s)] \\
&\quad + \lambda_f[\theta_{r_1,f}(t_0) - \theta_f^s(t_0)] - \lambda_f[\theta_{r_2,f}(t_0) - \theta_f^s(t_0)] \\
&\quad + \lambda_f[N_{r_1,f}^s - N_{r_2,f}^s] \\
&\quad + \delta\phi_{mp_{r_1,f}} - \delta\phi_{mp_{r_2,f}} + \varepsilon_{\phi,r_1,f} - \varepsilon_{\phi,r_2,f}. \tag{2.41}
\end{aligned}$$

The initial phase of the carrier $\theta_f^s(t_0)$ from the common transmitter, can be completely eliminated. Since antennas on-board the spacecraft have a short baseline, the travel time difference with respect to two antennas $\tau_{r_1}^s$ and $\tau_{r_2}^s$ will be very small. The transmitter instrumental delays $lb_f^s(t - \tau_{r_1}^s)$ and $lb_f^s(t - \tau_{r_2}^s)$, and transmitter clock errors $dt^s(t - \tau_{r_1}^s)$ and $dt^s(t - \tau_{r_2}^s)$ for two antennas can thus be considered equal over short time span $\tau_{r_1}^s - \tau_{r_2}^s$. In addition, by applying the SD between antennas with a short baseline, the spatially correlated ionospheric delays $I_{r_1,f}^s$ and $I_{r_2,f}^s$, and tropospheric delays $T_{r_1}^s$ and $T_{r_2}^s$ can also be significantly cancelled out. Eliminating common errors, Eq.(2.41) reads

$$\begin{aligned}
\Delta\rho_{r_{12},f}^s &= \Delta r_{\rho,r_{12}}^s + c\Delta dt_{r_{12}} + \Delta lb_{\rho,r_{12},f} \\
&\quad + \Delta\delta\tau_{mp_{r_{12},f}} + \Delta\varepsilon_{\rho,r_{12},f} \\
\Delta\phi_{r_{12},f}^s &= \Delta r_{\phi,r_{12}}^s + c\Delta dt_{r_{12}} + \Delta lb_{\phi,r_{12},f} + \lambda_f\Delta\theta_{r_{12},f} + \lambda_f\Delta N_{r_{12},f}^s \\
&\quad + \Delta\delta\phi_{mp_{r_{12},f}} + \Delta\varepsilon_{\phi,r_{12},f} \tag{2.42}
\end{aligned}$$

where Δ is the SD operator and subscript r_{12} indicates the difference between two antennas. As shown, the remaining error sources include the relative receiver clock error $c\Delta dt_{r_{12}}$, the relative receiver instrumental delays $\Delta lb_{\rho,r_{12},f}$ and $\Delta lb_{\phi,r_{12},f}$, multipath $\Delta\delta\tau_{mp_{r_{12},f}}$ and $\Delta\delta\phi_{mp_{r_{12},f}}$, and thermal noise $\Delta\varepsilon_{\rho,r_{12},f}$ and $\Delta\varepsilon_{\phi,r_{12},f}$ in both the pseudorange and carrier phase observations, while the relative initial carrier phase $\Delta\theta_{r_{12},f}$ and integer ambiguity $\Delta N_{r_{12},f}^s$ exist only in the phase observation.

The combined effects of the clock error, instrumental delay and the initial phase act as biases, which will be further discussed in the following section.

2.4.3 Bias analysis

The instrumental delay $\Delta lb_{r_{12},f}$ in the receiver mainly includes the line bias from the antenna to the receiver and the electronic circuit delay inside the receiver. The

relative electronic circuit delay of two receivers is normally very small and negligible, whereas the relative line bias is due to the different cable lengths from two antennas to receivers, which can be seen as a constant term over time and can be pre-calibrated.

To further eliminate the relative receiver clock errors $c\Delta dt_{\rho,r_{12}}$ and $c\Delta dt_{\phi,r_{12}}$, a perfect synchronization of the receiver clocks should be satisfied. This requirement initiates two possible arrangements:

- (1) a single receiver with connections to multiple antennas using a single internal clock; or
- (2) multiple receivers driven by a common external clock.

These two arrangements will both eliminate the relative clock errors $c\Delta dt_{\rho,r_{12}}$ and $c\Delta dt_{\phi,r_{12}}$. However, they have different impacts to the relative initial phase of the carrier replica $\Delta\theta_{r_{12},f}$, which should also be eliminated by the SD or pre-calibrated as the presence of $\Delta\theta_{r_{12},f}$ implies that the ambiguities cannot be treated as integers.

In the first arrangement, as the carrier replicas for yielding phase observations of two antennas come from two different channels in a single receiver, they do not only share the same internal clock, but also have the synchronized initial carrier phases between channels. This is the case for most of the receivers that are designed in the context of carrier phase-based applications (Giorgi, 2011). Therefore, this arrangement allows for eliminating both the clock errors and the initial phases by SD between antennas, leaving only the random noise and unmodelled multipath. However, it requires a receiver with connections to multiple antennas. Not many of the current receivers are available in the market with multi-antenna collections. Fortunately, most of the future space qualified GNSS receivers support multiple antenna connections, e.g. the SSTL SRG-20 receiver with 4 antennas was already demonstrated on TopSat mission (Duncan et al., 2008), the SSTL SGR-ReSI receiver has up to 8 single- or 4 dual-frequency antennas (Unwin et al., 2012), and the ESA AGGA-4 receiver has 4 antennas (Rosello et al., 2012) that is developed as next generation receiver. It is thus possible to modify these GNSS receivers such that they can operate as transceivers for the future RF-based inter-satellite relative navigation.

In the second arrangement, several receivers are driven by an external common clock. This arrangement only assures the clock drift over time is identical, while the initial phases of the carrier replicas for variable receivers are likely to be different (Keong, 1999). The reason is that the phase measurements are accumulations of the phase rate. A common clock can only guarantee the integrals are implemented in the same time slice, but the initial phases are different in different receivers. Therefore, the SD with a common external clock only eliminate the term $c\Delta dt_{\rho,r_{12}}$ and $c\Delta dt_{\phi,r_{12}}$, but leaves a constant non-zero initial relative phase bias $\Delta\theta_{r_{12},f}$ in the phase measurement.

Given the fact that both the line bias and the remaining initial phase bias are constant in the second arrangement, their combined effect can be treated as a constant bias that will either be pre-calibrated or estimated in a filter (Keong, 1999). The estimation methods will be discussed in Section 3.2.

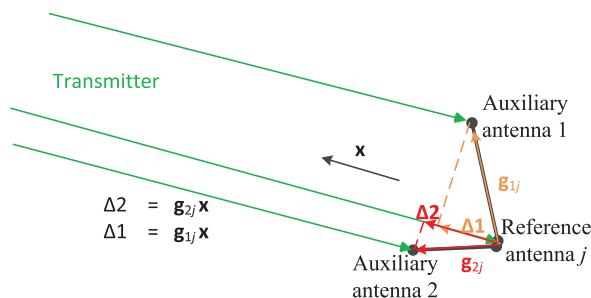


Figure 2.24: Geometry of a single transmitter and several receiving antennas, where the single differenced range is equal to the antenna baseline projection onto the LOS direction

2.5 Relative navigation model

2.5.1 Line-of-sight estimation model

The inter-satellite LOS can be estimated from the phase differences of carriers arriving at different antennas.

Assuming several antennas are fixed to the rigid spacecraft platform, their relative geometry with respect to the transmitting spacecraft is shown in Figure 2.24. The antenna baseline length is negligible compared to the inter-satellite distance. It is thus reasonable to assume the signals that arrive at different antennas are parallel. The relative orientations of these two spacecraft can then be expressed by a single LOS unit vector $\mathbf{x}^s = (x^s, y^s, z^s)^T$, subject to $\|\mathbf{x}^s\| = 1$ in the Cartesian body fixed frame. The superscript s represents a certain transmitting spacecraft s .

As shown in Figure 2.24, the single differenced range between a pair of antennas is equal to the baseline projection onto the LOS direction, which enables the establishment of the LOS observation model

$$\begin{aligned} \Delta \rho_{ij}^s &= \mathbf{g}_{ij}^T \mathbf{x}^s + \Delta \epsilon_{\rho_{ij}}^s \\ \Delta \phi_{ij}^s &= \mathbf{g}_{ij}^T \mathbf{x}^s + \lambda \Delta N_{ij}^s + \Delta \epsilon_{\phi_{ij}}^s \end{aligned}, \text{ subject to } \|\mathbf{x}^s\| = 1 \quad (2.43)$$

where $\Delta \rho_{ij}^s$ and $\Delta \phi_{ij}^s$ denote the SD pseudorange and carrier phase observations between reference antennas j and auxiliary antenna i , $\mathbf{g}_{ij} = (g_{x_{ij}}, g_{y_{ij}}, g_{z_{ij}})^T$ is the antenna baseline vector, and ΔN_{ij}^s is the unknown initial integer ambiguity for that antenna baseline. The SD errors $\Delta \epsilon_{\rho_{ij}}^s$, $\Delta \epsilon_{\phi_{ij}}^s$ include the receiver random noise and the unmodelled errors, e.g., multipath. Biases are assumed to be pre-calibrated and removed from the model.

The underlying assumption of having several paralleled arriving signals implicates the fact that the system operability is range-limited. The minimum range is a function of the range difference of two quasi-paralleled signals. Assuming the inter-satellite distance is around 160 m, the angle between two quasi-paralleled signals to a one-meter baseline will be approximately equal to $\arctan(1/160) = 0.0063$ rad. The range difference will then be $1(m)/\sin(0.0063) - 160 \text{ m} = 0.0010 \text{ m}$, which can be neglected in the observation model as it is on the same order as the random noise. In this case, the minimum operating range is around 160 m for one-meter baseline. Baselines less than one meter allow for operating in a shorter distance. The minimum range is thus extended. Note that this parallel signal assumption does not limit the maximum operating range. Signals are more parallel for longer inter-satellite distances.

Extending Eq.(2.43) to n baselines ($n+1$ antennas) yields:

$$\begin{aligned}\Delta\mathbf{P}^s &= \mathbf{G}\mathbf{x}^s + \Delta\epsilon_\rho^s \\ \Delta\Phi^s &= \mathbf{G}\mathbf{x}^s + \lambda\Delta\mathbf{N}^s + \Delta\epsilon_\phi^s, \text{ subject to } \|\mathbf{x}^s\| = 1\end{aligned}\quad (2.44)$$

where $\mathbf{G} = [\mathbf{g}_{1j}^T; \mathbf{g}_{2j}^T; \dots; \mathbf{g}_{nj}^T]$ is the baseline coordinate matrix for n baselines, SD pseudorange and carrier phase observations between the auxiliary antennas and reference antenna are grouped into $\Delta\mathbf{P}^s = [\Delta\rho_{1j}^s, \Delta\rho_{2j}^s, \dots, \Delta\rho_{nj}^s]^T$ and $\Delta\Phi^s = [\Delta\phi_{1j}^s, \Delta\phi_{2j}^s, \dots, \Delta\phi_{nj}^s]^T$, and $\Delta\mathbf{N}^s$ includes n integers $\Delta\mathbf{N}^s = [\Delta N_{1j}^s, \Delta N_{2j}^s, \dots, \Delta N_{nj}^s]^T$.

The model in Eq. (2.44) for the RF-based inter-satellite LOS estimation differs to the standard GPS-based attitude determination model, which is written below as comparison

$$\begin{aligned}\Delta\mathbf{P}_{ij} &= \mathbf{g}_{ij}^T\mathbf{X} + \Delta\epsilon_{\rho_{ij}} \\ \Delta\Phi_{ij} &= \mathbf{g}_{ij}^T\mathbf{X} + \lambda\Delta\mathbf{N}_{ij} + \Delta\epsilon_{\phi_{ij}}, \text{ subject to } \|\mathbf{g}_{ij}\| = L\end{aligned}\quad (2.45)$$

where M LOS vectors with respect to M GPS satellites are embedded in the matrix $\mathbf{X} = [\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^M]^T$ with the element \mathbf{x}^m as the LOS vector to the m th GPS satellite, $\Delta\mathbf{N}_{ij}$ now contains M ambiguities for M satellites, and \mathbf{g}_{ij}^T represents a single antenna baseline to be estimated. Here, both the LOS and baseline vectors are expressed in the local coordinate frame such as the north-east-up frame, in which the baseline vector \mathbf{g}_{ij}^T becomes unknown, and LOS vectors in \mathbf{X} with respect to all visible GPS satellites comprises the design matrix.

For the attitude determination model in Eq.(2.45), the LOS vector between GPS satellite and the user can be coarsely calculated using the GPS satellite ephemeris and pseudoranges. However, in a RF-based relative navigation system of model (2.44), precise ephemeris (absolute positions) of the transmitting spacecraft is likely to be unknown to the other spacecraft in the formation. Therefore, the LOS vector is unknown, and antenna baseline matrix becomes known and will be used as design matrix for the LOS estimation.

In the computational aspect, the model (2.44) shows an advantage over the model (2.45). The design matrix in Eq.(2.44) is the antenna baseline matrix \mathbf{G} , which has constant elements in the body fixed frame as long as they are fixed on the rigid platform, while the design matrix in Eq.(2.45) is the LOS matrix \mathbf{X} , which changes over time as GPS satellites and/or platform move in the local coordinate frame. The design matrix needs to be decomposed for the integer ambiguity resolution. The decomposition of \mathbf{G} is only performed once while the decomposition of \mathbf{X} has to be repeated every epoch until the resolved ambiguities are validated (Sutton, 2002). Therefore, the LOS estimation model of Eq.(2.44) implies less computational load.

However, Eq.(2.44) also presents a drawback compared to Eq.(2.45). The baseline matrix \mathbf{G} has a weak geometry diversity since antennas are all fixed on the spacecraft with limited dimensions, while the geometry diversity in Eq.(2.45) is determined by more diversely sparse positions of GPS satellites. Properly arranging antennas on different locations on the spacecraft can compensate this drawback to some extent.

By using an ultra-BOC signal structure, additional carrier phase measurements at different frequencies will be available. Thus, the inter-satellite LOS estimation model for n baselines, m frequencies at a single-epoch can be written by extending Eq.(2.44). The superscript s is omitted to generalize the LOS vector between any

transmitting spacecraft and receiving spacecraft in the formation

$$\begin{bmatrix} \Delta \mathbf{P} \\ \Delta \Phi_{f_1} \\ \Delta \Phi_{f_2} \\ \vdots \\ \Delta \Phi_{f_m} \end{bmatrix} = \begin{bmatrix} \mathbf{G} & & & & \\ \mathbf{G} & \lambda_1 \mathbf{I}_n & & & \\ \mathbf{G} & & \lambda_2 \mathbf{I}_n & & \\ \vdots & & & \ddots & \\ \mathbf{G} & & & & \lambda_m \mathbf{I}_n \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \Delta \mathbf{N}_{f_1} \\ \Delta \mathbf{N}_{f_2} \\ \vdots \\ \Delta \mathbf{N}_{f_m} \end{bmatrix} + \Delta \boldsymbol{\varepsilon} \quad (2.46)$$

subject to $\|\mathbf{x}\| = 1$

where $\Delta \mathbf{N}_{f^*}$ indicates the SD ambiguities on frequency f^* , which has the size of $n \times 1$. The number of unknowns, including the LOS vector and ambiguities, is $mn + 3$, while the number of measurements is $(m + 1)n$. To solve the equation, at least 3 baselines are needed. This means there must be at least 4 antennas mounted on the receiving spacecraft. Moreover, the \mathbf{G} matrix shall have a full column rank, requiring antennas to be arranged in a non-planar geometry.

Note that the LOS vector is represented in Cartesian coordinates $\mathbf{x} = (x, y, z)^T$. It can also be expressed by bearing angles of elevation (el) and azimuth (az) in polar coordinates

$$\begin{aligned} el &= \arctan \left(\frac{z}{\sqrt{x^2 + y^2}} \right) \\ az &= \arctan \left(\frac{y}{x} \right). \end{aligned} \quad (2.47)$$

Both coordinate frames are equivalent in providing the information of the inter-satellite relative orientations. The LOS vector has one more unknown than bearing angles. However, a constraint that \mathbf{x} is subject to $\|\mathbf{x}\| = 1$ can be exploited as a-priori available information to improve the estimation performance.

2.5.2 Inter-satellite distance estimation model

The inter-satellite distance needs to be estimated using the undifferenced pseudorange and carrier phase observables. This estimation will then suffer from clock errors or the oscillator instability that are embedded in the undifferenced observation model.

Recall the undifferenced observation model in section 2.4.1, the errors from the oscillator instability are written as:

$$e_r^s = c[dt_r(t) - dt^s(t - \tau_r^s)] \quad (2.48)$$

where e_r^s is the relative clock offset observed by r and transmitted by s , $dt_r(t)$ is caused by the receiver oscillator instability and recorded at the receive time tag of t , while $dt^s(t - \tau_r^s)$ is caused by the transmitter oscillator instability and recorded at the transmit time tag of $t - \tau_r^s$ with τ_r^s as the signal travel time.

A dual one-way ranging method was proposed by Kim and Tapley (2002) to minimize the oscillator noise effect by combining two one-way ranging measurements. The dual one-way method has been used on GRACE (Kim and Tapley, 2002; Kim and Lee, 2009) and PRISMA mission (Thevenet and Grelier, 2012). The concept of this method is illustrated in Figure 2.25. With identical transmission and reception subsystems, each satellite transmits a RF-based signal to the other satellite. The received signal at each spacecraft is on-board processed and the associated pseudorange

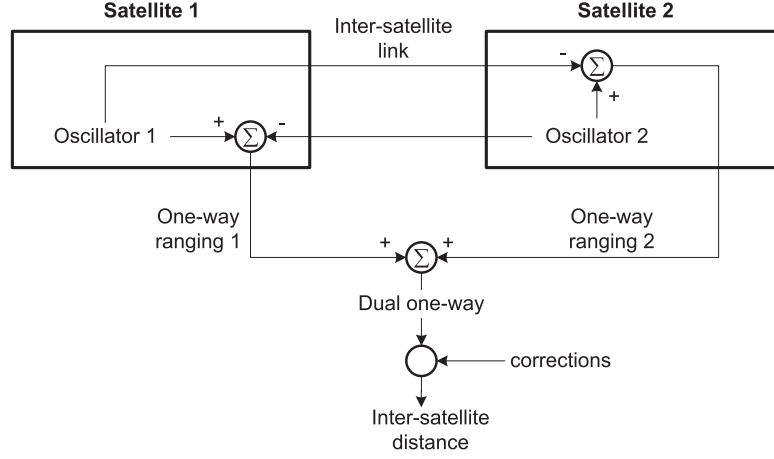


Figure 2.25: Architecture of the dual one-way ranging method (Kim and Tapley, 2002)

and carrier phase measurements are extracted and recorded. These measurements are called the one-way ranging measurements, and there are two groups of one-way measurements from the pair of spacecraft.

Then, a similar equation as Eq.(2.48) can be written to express the relative clock offset observed by s and transmitted by r

$$e_s^r = c[dt_s(t) - dt^r(t - \tau_s^r)]. \quad (2.49)$$

According to Kim and Tapley (2002, 2003), the relative clock offset due to the oscillator instability and recorded on both satellites have nearly equal and opposite effects on each one-way measurements. Summation of these two measurements cancels most of the oscillator noise that have long and medium period parts. Only the high-frequency noise, whose period is shorter than the signal travel time, remains after the the summation process

$$e_r^s + e_s^r \approx 0. \quad (2.50)$$

To this end, summation of two one-way undifferenced pseudorange and carrier phase measurements will remove the relative clock offset and result in the following observation model

$$\begin{aligned} \rho_{r,f_s}^s(t) + \rho_{s,f_r}^r(t) &= r_{\rho,r}^s(t, t - \tau_r^s) + r_{\rho,s}^r(t, t - \tau_s^r) \\ &\quad + I + lb_\rho + \delta\tau_{mp} + \varepsilon_\rho \\ \phi_{r,f_s}^s(t) + \phi_{s,f_r}^r(t) &= r_{\phi,r}^s(t, t - \tau_r^s) + r_{\phi,s}^r(t, t - \tau_s^r) \\ &\quad - I + lb_\phi + \lambda_{f_s} \theta_{r,f_s}^s(t_0) + \lambda_{f_r} \theta_{s,f_r}^r(t_0) \\ &\quad + \lambda_{f_s} N_{r,f_s}^s + \lambda_{f_r} N_{s,f_r}^r + \delta\phi_{mp} + \varepsilon_\phi \end{aligned} \quad (2.51)$$

where the superscript s (or r) and subscript r (or s) donate the signal transmitted from s (or r) and received by r (or s), and the associated transmission frequencies are f_s and f_r , respectively. It is worth mentioning that measurements on each spacecraft may not be recorded at a common time because of the clock de-synchronization. For the GRACE mission, this effect is called time-tag error and is removed by interpolating the raw one-way measurements into the corrected time-tag (Kim and

Tapley, 2002). The solution for the PRISMA mission is to store measurements in the onboard navigation process unit, perform samples selection and propagate measurements using the Doppler information so as to re-synchronize the raw one-way measurements to a common time (Thevenet and Grelier, 2012).

Both the summed pseudorange and carrier phase measurements of (2.51) have several remaining errors, including the ionospheric delays I , the instrumental delay lb , multipath errors $\delta\tau_{mp}$, $\delta\phi_{mp}$ and thermal noise ε . Phase measurements also consist of the initial phase bias $\theta(t_0)$ and integer ambiguity N . The instrumental delays together with the initial phase bias require pre-calibration to be eliminated. The ionospheric effect may be eliminated using the conventional ionospheric-free dual-frequency combinations. The remaining errors will then include multipath and thermal noise. Then, the geometrical distance between two satellites can be computed as the half of the dual one-way summation

$$\begin{aligned} r_{\rho,r}^s(t, t - \tau_r^s) &= r_{\rho,s}^r(t, t - \tau_s^r) = \frac{\rho_{r,f_s}^s(t) + \rho_{s,f_r}^r(t)}{2} + \tilde{\varepsilon}_\rho \\ r_{\phi,r}^s(t, t - \tau_r^s) &= r_{\phi,s}^r(t, t - \tau_s^r) = \frac{\phi_{r,f_s}^s(t) + \phi_{s,f_r}^r(t)}{2} + \tilde{N} + \tilde{\varepsilon}_\phi \end{aligned} \quad (2.52)$$

where $\tilde{\varepsilon}_\rho$, $\tilde{\varepsilon}_\phi$ include all the remaining errors in the pseudorange and carrier phase measurements, and \tilde{N} is the combined integer ambiguity term from the dual one-way measurements.

The half-sum in this distance computation enables the removal of the relative clock offset between satellites. Similarly, the half-difference of the dual one-way measurements will eliminate the contribution of distance and yield the relative clock offset estimation

$$\begin{aligned} e_{\rho,s}^r &= -e_{\rho,r}^s = \frac{\rho_{r,f_s}^s(t) - \rho_{s,f_r}^r(t)}{2} + \tilde{\tilde{\varepsilon}}_\rho \\ e_{\phi,s}^r &= -e_{\phi,r}^s = \frac{\phi_{r,f_s}^s(t) - \phi_{s,f_r}^r(t)}{2} + \tilde{N} + \tilde{\tilde{\varepsilon}}_\phi \end{aligned} \quad (2.53)$$

where $\tilde{\tilde{\varepsilon}}_\rho$, $\tilde{\tilde{\varepsilon}}_\phi$ are used to denote errors of the half-differenced clock estimation model, and \tilde{N} is the integer ambiguity in the half-differenced phase measurement.

A reliable integer ambiguity resolution in the dual one-way ranging method is a non-trivial task. Several significant error sources, e.g., the ionospheric effect, shall be primarily removed. The inter-satellite distance can thus be estimated in the first instance by using only the pseudorange measurements. The accuracy is in the meter level. The LOS ambiguity resolution results can assist in the distance ambiguity resolution, as reported by the PRISMA mission (Grelier et al., 2011).

2.6 Chapter summary

This chapter presented RF-based relative navigation transceiver concept and architecture for future formation flying missions. The transceiver design inherits basic GNSS technologies but by utilizing a locally generated PRN ranging code.

The transceiver architecture, functionality and performance were elaborately discussed in this chapter. Two different signal structures, BPSK-R and BOC, were comprehensively analysed and evaluated in terms of the lower bound accuracy, multipath performance and acquisition and tracking strategy. The BPSK-R(10) was suggested

as appropriate inter-satellite ranging code as it does not only provide high ranging accuracy and good multipath performance, but also requires a low computational tracking strategy without the need of distinguishing multiple peak ambiguities.

Three carrier frequencies in the S-band, S1, S2 and S3, were proposed to facilitate the carrier phase integer ambiguity resolution (IAR). However, the PRN ranging code only needs to be modulated onto the S1 frequency, while other frequencies are modulated by low rate communication data or unmodulated so as to maximally avoid the code despreading process in the receiver and also to maintain the capability of extra frequency-aided fast IAR.

This chapter also introduced basic models for the inter-satellite LOS and distance estimation. The following chapters focus more on the LOS estimation and the associated IAR and multipath mitigation. Chapter 3 will elaborate the LOS model and propose unaided and instantaneous (single-epoch) IAR methods. Innovative multipath mitigation methods will be introduced in chapter 4 and 5 for reducing multipath errors on pseudorange and carrier phase measurements, respectively. The multipath effects on IAR will also be discussed in chapter 5.

Chapter 3

Line-of-sight Estimation

Determining the line-of-sight (LOS) between two spacecraft in a formation plays a crucial role in formation acquisition and maintenance. One of the ways to achieve this is the use of a RF-based inter-satellite system. With multiple antennas mounted on the rigid spacecraft body, the LOS can be precisely estimated using the carrier phase differences between antennas. The basic LOS model has been derived in chapter 2, as well as the analysis of dominating error sources. This chapter focuses on elaborating the LOS model, resolving the associated integer ambiguities efficiently and reliably, evaluating the antenna geometry impacts, and characterizing the estimation performance by both numerical simulations and field tests.

3.1 Problem statement and existing methods

The RF based LOS estimation relies on highly precise carrier phase observations. However, it is well known that the phase measurements are ambiguous by an unknown integer number of cycles. A process called *integer ambiguity resolution* (IAR) is required which resolves these unknown ambiguities as integers, and it is the key to be able to exploit the very high precision of the carrier phase data. The chapter aims at proposing an instantaneous (single-epoch) IAR for the LOS estimation.

3.1.1 Integer ambiguity resolution

Various integer ambiguity resolution (IAR) methods have been developed, differing in the way of how the performance can be achieved and how the computational efficiency can be improved. The performance means the capability to discriminate a correct ambiguity set from all candidate sets. This capability has intrinsic relations to the strength of the underlying model, e.g., the noise on observations and the available frequencies. A good resolution method is thus difficult to be distinguished from the others based on their performances as the underlying model they apply may differ from each other. It is easier to use the computational efficiency rather than the performance to compare different methods (Kim and Langley, 2000).

Most the ambiguity resolution methods are carried out based on the theory of the integer least squares (ILS). Three steps would be involved - the float solution, the integer ambiguity mapping/search process, and the fixed solution. The most computationally intensive part is the ambiguity search process. Two approaches in terms

Table 3.1: Search space reduction approaches for various ambiguity resolution techniques (Kim and Langley, 2000). Here, \checkmark (or \times) denotes the specific technique *is* (or *is not*) used the certain method.

Ambiguity resolution methods	Search space transformation	Conditional search	References
LAMBDA	\checkmark	\checkmark	Teunissen (1995)
FARA	\times	\checkmark	Frei and Beutler (1990)
Null space	\checkmark	\times	Martin-Neira et al. (1995)
FASF	\times	\checkmark	Chen and Lachapelle (1995)
OMEGA	\checkmark	\checkmark	Kim and Langley (1999)

of reducing the search space can be used to classify different ambiguity resolution methods. They are the search domain transformation and the conditional search (Kim and Langley, 2000). Table 3.1 lists various ambiguity resolution methods that utilize these approaches.

In the search domain transformation, the original ambiguity sets are transformed through a “many-to-one” relationship and/or through redefining a more efficient search space (Kim and Langley, 2000). For instance, the LAMBDA (Least-squares AMBIGUITY Decorrelation Adjustment) method introduces a decorrelation transformation procedure, which maximumly decorrelates the covariances between ambiguities and returns new ambiguities that show a dramatic improvement in correlation and precision (Teunissen, 1995). The search for the transformed integers is more efficient in the LAMBDA method. Defining conditional search in multi-level searches is also an efficient approach to reduce the search space, e.g. the FARA (Fast Ambiguity Resolution Approach) and FASF (Fast Ambiguity Search Filter) methods. These methods were proposed based on the fact that the ambiguity parameters of lower search levels can be conditioned on those of upper search, in which way the search space is reduced. The methods that simultaneously utilize both approaches include, e.g., the LAMBDA and OMEGA (Optimal Method for Estimation GPS Ambiguities), as indicated in Table 3.1.

Although the aforementioned different methods were proposed to improve the IAR computational efficiency, a good IAR performance (high success rate) is the ultimate goal. As mentioned, the performance is highly dependent on the strength of the underlying model. In existing GNSS-based carrier phase positioning algorithms, the model becomes stronger by taking advantage of at least one of the following redundancies: (1) additional observables over time in changing satellite-user geometry (Cohen, 1992; Park, 2001); (2) additional data sources such as inertial sensors (Scherzinger, 2000, 2002; Petovello, 2003) or other positioning systems (Grellier et al., 2011); (3) multiple frequencies (Teunissen et al., 2002; O’Keefe et al., 2009); (4) more antennas (Sutton, 2002; Giorgi, 2011; Teunissen, 2011); and (or) (5) constraints (Sutton, 2002; Park and Teunissen, 2003; Monikes et al., 2005).

The first source of redundancy triggered the development of the earliest strategies to facilitate IAR, which were called motion-based methods. The platform motion is required to create geometrical variations that can be exploited by the methods. The main disadvantage is that it does not provide an instantaneous solution. The time required for the ambiguity initialization is in the order of few seconds when the platform is properly moved, but a few minutes if the changes in the satellite-user geometry are only given by the satellite motion (Giorgi, 2011). In space applications, this concept was also widely used. In Park (2001), a local pseudolite (pseudo

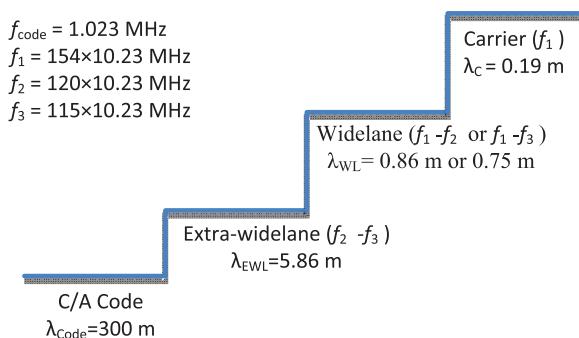


Figure 3.1: Cascading integer ambiguity resolution structure in the GNSS triple frequency system

GPS-like transmitter) onboard the spacecraft in the formation was used to augment the differential GPS. The local pseudolite range measurements allow a faster relative motion that has been used to efficiently solve ambiguities in near real-time. In the PRISMA mission, attitude manoeuvring was performed by rotating one spacecraft around the axis of its three-antenna triplet base with respect to another spacecraft in order to accelerate the IAR for the inter-satellite LOS estimation (Grelier et al., 2011). The magnitude of attitude rotation was set to 50° and the increased magnitudes of 60° , 70° , 80° and 100° were also tested in order to improve the IAR robustness. Although these motion-based methods allow for a shorter time span IAR, they are inherently non-instantaneous methods. Furthermore, the required platform motion limits the scope of applications.

The second type of redundancy is to take advantage of the *a-priori* position and/or velocity knowledge from other sensors, e.g. the inertial measurement units (IMUs) (Scherzinger, 2000, 2002; Petovello, 2003). IMUs are completely autonomous instruments that sense accelerations and rotation rates and integrate them to attitude, velocity and position increments. The position and velocity measurements from these inertial navigation system (INS) can normally be used to aid GPS during complete or partial GPS data outages, i.e. the absence of updates. In INS aided solutions, ambiguities can be re-initialized in a shorter time. As reported by Scherzinger (2000), it is possible to recover L1 integer ambiguities within seconds after a short-duration GPS outage and also maintain decimeter-level accuracy throughout the outage. Possible improvements and limitations of the INS aided IAR during variable GPS data outage durations were discussed in Petovello (2003). In the PRISMA mission, apart from the aforementioned motion-aided IAR, GNC-aided IAR was also performed for the LOS estimation when the signal loses lock for some reason (e.g., antenna handover). Instead of using IMUs, differential GPS solutions obtained in the GNC filter are converted into distance and LOS measurements which then serve as references to facilitate the LOS IAR on the RF-based relative navigation system (Grelier et al., 2011).

The method of involving additional frequencies as redundancy was initiated by using the so-called widelane observation. Since the non-ambiguous pseudorange has the random noise in the meter level, it is far too noisy to be directly used for estimating integer cycles of carriers with centimeter-level wavelengths. The widelane observation is then formed by two carrier phases on two frequencies so that a much bigger wavelength quasi-carrier can be introduced. Taking GPS L1, L2 frequencies

as example, the widelane wavelength λ_{WL} is established as

$$\begin{aligned}\lambda_{L1} &= c/f_{L1} = 0.19 \text{ m} \\ \lambda_{L2} &= c/f_{L2} = 0.24 \text{ m} \\ \lambda_{WL} &= c/(f_{L1} - f_{L2}) = 1/(1/\lambda_{L1} - 1/\lambda_{L2}) = 0.86 \text{ m}\end{aligned}\quad (3.1)$$

and the widelane observation ϕ_{WL} is built as

$$\begin{aligned}\phi_{WL} &= \frac{\phi_{L1}/\lambda_{L1} - \phi_{L2}/\lambda_{L2}}{1/\lambda_{L1} - 1/\lambda_{L2}} \\ &= r_r^s + \lambda_{WL}(N_1 - N_2) + \text{other error sources}\end{aligned}\quad (3.2)$$

where r_r^s is the true range between satellite and receiver, the widelane ambiguity $N_1 - N_2$ is easier and more reliable to resolve as λ_{WL} is the much larger than carrier wavelengths of λ_{L1} or λ_{L2} . With the correct integer, the next step is to treat the fixed widelane as a new pseudorange to fix the ambiguity on carriers. Since the fixed widelane is much more precise than the pseudorange, it becomes possible to estimate the carrier ambiguity with more sufficient confidence in a shorter time. The sequence of steps between pseudorange, widelane, and carrier has led to this method being called cascading. Triple-frequency variations of this method have also been proposed for the modernized GPS and Galileo where an additional step is added to make use of the third frequency to form an extra-widelane observable with an even longer wavelength. The ambiguity on the extra-widelane will be resolved in the first place before cascading down to the widelane observable. This can be depicted by Figure 3.1, taking the L1, L2 and L5 triple-frequency modernized GPS system for example. These multi-frequency based methods are generally referred to as either three carrier ambiguity resolution (TCAR), multiple carrier ambiguity resolution (MCAR), or simply cascading integer resolution (CIR) methods (Jung, 1999; Zhang et al., 2003; O'Keefe et al., 2009).

The ultra-BOC signal, proposed in chapter 2, is a variant structure of using multiple frequencies. It combines several carriers into one signal structure by introducing separate carrier tones apart from the central carrier spectrum, see section 2.2. This ultra-BOC signal shows a similar capability as of using multiple frequencies but can be processed more easily without the need of code wipe-off in the signal acquisition and tracking processes. The ultra-BOC structure can be preceded in a cascading way to facilitate integer ambiguity resolution. Nevertheless, the cascading integer resolution can be treated as a special usage of multiple frequencies. It does single differencing between frequencies but cannot guarantee it is the optimal combination of frequencies. Unlike the cascading methods that rely on specific linear combinations of the carrier phase measurements, the LAMBDA method uses the decorrelation to the ambiguity covariance matrix so that it can intrinsically assure an optimal combination between frequencies and other influencing factors such that the correlations between ambiguities are minimized (Teunissen et al., 2002; Verhagen and Joosten, 2004). It has been shown in Teunissen et al. (2002) that the various cascading schemes are theoretically suboptimal compared to the LAMBDA-derived linear combination. In O'Keefe et al. (2009), this conclusion was also demonstrated by simulations. Therefore, the observation model in following sections will utilize the ultra-BOC structure in the LAMBDA method to assure the optimal usage of multiple frequencies.

Having redundancies of more antennas and constraints are usually coupled, e.g., employing the antenna baseline length and/or the baseline geometry on the rigid

platform as constraints. The constraints can then be treated as a-priori information to strengthen the underlying model and to augment the reliability of the ambiguity estimation. Employing more antennas not only provides redundant observations, but also introduces more associated geometrical constraints to allow for improved ambiguity resolution. This especially works well for the GNSS-based attitude determination when multiple antennas are rigidly mounted on the platform (Sutton, 2002; Park and Teunissen, 2003; Kuylen et al., 2005; Teunissen, 2006, 2007; Giorgi, 2011). It can also benefit the relative navigation between two platforms if each platform carries a number of antennas (Buist et al., 2009, 2011).

For the RF-based relative LOS estimation in this chapter, it is crucially important to have a fast, unaided and non-motion-based ambiguity resolution since autonomous spacecraft formation flying missions usually operate in a tightly controlled time-critical mode. The precise LOS estimation not only needs to be provided timely to the subsequent relative orbit or attitude propagation, but shall also avoid any a-priori information from other sensors, so that the RF-based system can be foreseen as the first-stage methodology in autonomous navigation before its incorporation with other systems for more accurate and robust navigation. The potential redundancies that can be used for improving the IAR performance in this RF-based system can then include frequencies, antennas and constraints. The LAMBDA method is chosen to assure computational efficiency. In fact, the LAMBDA method is currently the benchmarking technique to solve integer ambiguities, as it is known to be optimal not only in the sense that it works in a highly efficient way, but also in the sense that it can provide the highest possible success rates (Teunissen, 1999; Verhagen and Teunissen, 2006).

3.1.2 Benchmarking solution: LAMBDA

A generalized carrier-phase based model can be cast in a linear(ized) system of observation equations as

$$E(\mathbf{y}) = \mathbf{B}\mathbf{x} + \mathbf{A}\mathbf{a} = \begin{bmatrix} \mathbf{B} & \mathbf{A} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{a} \end{bmatrix}, \mathbf{x} \in \mathbb{R}^q, \mathbf{a} \in \mathbb{Z}^p \quad (3.3)$$

$$D(\mathbf{y}) = \mathbf{Q}_{\mathbf{y}\mathbf{y}} \quad (3.4)$$

where $E(\cdot)$ and $D(\cdot)$ are the mathematical expectation and dispersion operators, \mathbf{y} is the vector of observables, which consists of the code and carrier phase measurements, \mathbf{x} is the real-valued vector of unknowns (order q), \mathbf{a} is the integer-valued vector of unknowns (order p), \mathbf{B} and \mathbf{A} are the design matrices that link the vector of observations to the vector of unknowns \mathbf{x} and \mathbf{a} , respectively. The dispersion of \mathbf{y} , denoted with $D(\mathbf{y})$, is characterized by the covariance matrix $\mathbf{Q}_{\mathbf{y}\mathbf{y}}$.

One usually applies the least squares principle to solve the unknowns in Eq.(3.3) in the form of a minimization problem

$$\min_{\mathbf{a} \in \mathbb{Z}^p, \mathbf{x} \in \mathbb{R}^q} \|\mathbf{y} - \mathbf{A}\mathbf{a} - \mathbf{B}\mathbf{x}\|_{\mathbf{Q}_{\mathbf{y}\mathbf{y}}}^2 \quad (3.5)$$

where $\|\cdot\|_{\mathbf{Q}_{\mathbf{y}\mathbf{y}}}^2 = (\cdot)^T \mathbf{Q}_{\mathbf{y}\mathbf{y}}^{-1} (\cdot)$. Note that the minimization takes place in $\mathbb{R}^q \times \mathbb{Z}^p$. The integer nature of the subset unknowns makes it impossible to obtain an analytical closed form expression to Eq.(3.5). The well-known LAMBDA method provides a good solution to this problem. In the following, the key process of the LAMBDA method is described.

There are basically three steps in LAMBDA. The first step is to obtain the so-called float solution using a standard least squares adjustment. The integer nature of the ambiguities $\mathbf{a} \in \mathbb{Z}^p$ is disregarded and real-valued estimates of \mathbf{x} and \mathbf{a} can be obtained, together with their associated covariance matrix

$$\begin{bmatrix} \hat{\mathbf{x}} \\ \hat{\mathbf{a}} \end{bmatrix}, \quad \begin{bmatrix} \mathbf{Q}_{\hat{\mathbf{x}}\hat{\mathbf{x}}} & \mathbf{Q}_{\hat{\mathbf{x}}\hat{\mathbf{a}}} \\ \mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{x}}} & \mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{a}}} \end{bmatrix}. \quad (3.6)$$

This solution is referred to as the float solution. Since the fact that the ambiguities are integers has not yet been exploited, this float solution is not as precise as possible. Therefore, in the second step, the float ambiguity estimate $\hat{\mathbf{a}}$ is used to compute the corresponding integer ambiguity estimate $\check{\mathbf{a}}$. This implies that a mapping (or search) procedure $\mathbb{R}^p \rightarrow \mathbb{Z}^p$, from the p -dimensional space of real values to the p -dimensional space of integers. Search is implemented to minimize a standard ambiguity objective function $\mathbf{J}(\mathbf{a})$ based on the Integer Least Squares (ILS) adjustment

$$\check{\mathbf{a}} = \min_{\mathbf{a} \in \mathbb{Z}^p} \mathbf{J}(\mathbf{a}) = \min_{\mathbf{a} \in \mathbb{Z}^p} \|\hat{\mathbf{a}} - \mathbf{a}\|_{\mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{a}}}}^2 \quad (3.7)$$

where \mathbf{a} represents the integer candidate.

The search space is governed by $\mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{a}}}$ which represents an elongated ellipse due to the correlations between individual ambiguities. The method of measuring the nearest integer vector to $\hat{\mathbf{a}}$ is to perform the search in a sequential conditional adjustment in volume χ^2

$$(\hat{\mathbf{a}} - \mathbf{a})^T \mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{a}}}^{-1} (\hat{\mathbf{a}} - \mathbf{a}) \leq \chi^2. \quad (3.8)$$

where the size χ^2 is obtained by an initial rounding or bootstrapping method (Teunissen, 1998).

Using the LDL^T -decomposition of matrix $\mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{a}}}^{-1}$ with lower triangular matrix \mathbf{L} and diagonal matrix \mathbf{D} , the quadratic inequality in Eq.(3.8) can be written as (de Jonge and Tiberius, 1996)

$$\sum_{i=1}^p d_i \left[(a_i - \hat{a}_i) + \sum_{j=1}^{i-1} l_{ij} (a_j - \hat{a}_j) \right]^2 \leq \chi^2 \quad (3.9)$$

where d_i and l_{ij} are the diagonal elements in \mathbf{D} and the lower triangular elements in \mathbf{L} , respectively.

By defining a conditional estimate for \hat{a}_i as $\hat{a}_{i|1,2,\dots,i-1}$ which is conditioned on candidate integers a_1, a_2, \dots, a_{i-1} , the sequential conditional adjustment can be performed by rewriting Eq.(3.9) in p sequential intervals for searching each of the ambiguities

$$\begin{aligned} (a_1 - \hat{a}_1)^2 &\leq \chi^2/d_1 \\ (a_2 - \hat{a}_{2|1})^2 &\leq (\chi^2 - d_1(a_1 - \hat{a}_1)^2)/d_2 \\ (a_3 - \hat{a}_{3|1,2})^2 &\leq (\chi^2 - (d_1(a_1 - \hat{a}_1)^2 + d_2(a_2 - \hat{a}_{2|1})^2))/d_3 \\ &\dots \\ (a_p - \hat{a}_{p|1,2,\dots,p-1})^2 &\leq \left(\chi^2 - \sum_{i=1}^{p-1} d_i (a_i - \hat{a}_{i|1,2,\dots,i-1})^2 \right) / d_p \end{aligned} \quad (3.10)$$

with

$$\hat{a}_{i|1,2,\dots,i-1} = \hat{a}_i - \sum_{j=1}^{i-1} l_{ji}(a_j - \hat{a}_j). \quad (3.11)$$

This conditional search makes full use of the correlations between ambiguities expressed in \mathbf{L} and is thus working efficiently. The search terminates when all valid candidates inside the ellipsoid have been treated. An ambiguity vector that yields the minimum for Eq.(3.7) is thus obtained, and regarded as the fixed ambiguity $\hat{\mathbf{a}}$.

In order to analyze how the ambiguities and the fixed estimators are distributed, the concept of pull-in region is introduced by Teunissen (2002). These regions are used to identify the set of real-valued (float) ambiguities which are “pulled” to the same integer ambiguity matrix following a given integer estimation process. The pull-in regions in Figure 3.2 are based on integer least squares (ILS). They are centered at integer grid points. If the float ambiguity solution resides in a specific pull-in region, the corresponding integer grid point will be the ILS solution.

The probability that the float ambiguity vector $\hat{\mathbf{a}}$ is mapped to an integer vector \mathbf{a} is (Teunissen, 2002):

$$P(\Psi(\hat{\mathbf{a}}) = \mathbf{a}) = \int_{S_{\mathbf{a}}} f_{\hat{\mathbf{a}}}(x) dx \quad (3.12)$$

where $S_{\mathbf{a}}$ is the pull-in region, $f_{\hat{\mathbf{a}}}(x)$ is the probability density function of $\hat{\mathbf{a}}$, $\Psi(\hat{\mathbf{a}})$ is the mapping function $\Psi(\hat{\mathbf{a}}) : \mathbb{R}^p \rightarrow \mathbb{Z}^p$. Due to the integer nature of \mathbb{Z}^p , the map $\Psi(\hat{\mathbf{a}})$ will not be an one-to-one, but instead a many-to-one map.

It is common practise to use the *success rate* to decide on acceptance or rejection of the integer ambiguity resolution. The success rate is defined as the probability that the float ambiguity vector is mapped to the correct integer vector. The success rate is completely determined by the distribution of float ambiguities. In Teunissen (1998), a bootstrapped-based lower bound success rate is given as

$$P_{LB} = \prod_{i=1}^p \left(2\Phi \left(\frac{1}{2\sigma_{\hat{a}_{i|1,2,\dots,i-1}}} \right) - 1 \right) \quad (3.13)$$

with $\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp\{-\frac{1}{2}v^2\} dv$ denoted as the normal distribution function and $\sigma_{\hat{a}_{i|1,2,\dots,i-1}}^2$ as the variance of the i^{th} ambiguity estimate conditioned on the previous integers. It should be mentioned that $\sigma_{\hat{a}_{i|1,2,\dots,i-1}}^2$ is equal to the i^{th} diagonal element d_i in the LDL^T -decomposed matrix \mathbf{D} (de Jonge and Tiberius, 1996).

The P_{LB} is an easy-to-compute lower bound of the success rate. It differs to the empirical success rate P_E , which is defined as the percentage of occurrences that the computed integer solution in the experiment is equal to the true integer vector.

Once the integer ambiguity solution is accepted, the third and last step of LAMBDA consists of correcting the float solution of all other real-valued parameters of interest by virtue of their correlation with the ambiguities. As a result, the fixed solution $\check{\mathbf{x}}$ will be obtained. Provided of correctly fixed ambiguities, the fixed solution will have a precision that is in accordance with the high precision of the phase measurement

$$\check{\mathbf{x}} = \hat{\mathbf{x}} - \mathbf{Q}_{\hat{\mathbf{x}}\hat{\mathbf{a}}} \mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{a}}}^{-1} (\hat{\mathbf{a}} - \check{\mathbf{a}}) \quad (3.14)$$

$$\mathbf{Q}_{\check{\mathbf{x}}\check{\mathbf{x}}} = \mathbf{Q}_{\hat{\mathbf{x}}\hat{\mathbf{x}}} - \mathbf{Q}_{\hat{\mathbf{x}}\hat{\mathbf{a}}} \mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{a}}}^{-1} \mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{x}}}. \quad (3.15)$$

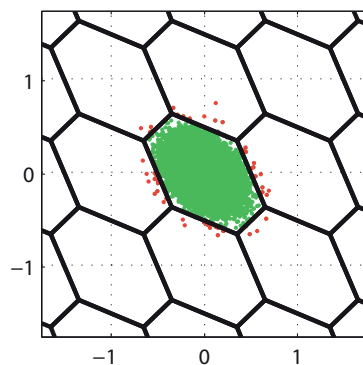


Figure 3.2: Two-dimensional pull-in-regions (black) and the float ambiguity solutions, which are in green if ambiguities are correctly fixed, and in red if they are wrongly fixed (Verhagen, 2012)

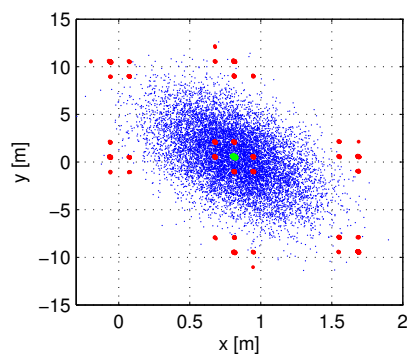


Figure 3.3: Scatterplot of the two-dimensional distribution of float solutions $\hat{\mathbf{x}}$ (blue) and the corresponding fixed solutions $\check{\mathbf{x}}$ (green or red). In this case, 69.2% of the float ambiguity solutions are correctly fixed (green), and 30.8% is wrongly fixed (red), which results in $\check{\mathbf{x}}$ distributed in non-physical locations.

In general, the elements in $\mathbf{Q}_{\check{\mathbf{x}}\check{\mathbf{x}}}$ shall be approximately two orders smaller than the elements in $\mathbf{Q}_{\hat{\mathbf{x}}\hat{\mathbf{x}}}$. However, incorrect integer ambiguity estimation may result in the opposite effect in terms of positioning accuracy: rather than a dramatic precision improvement, wrong ambiguity solutions can cause very large position errors, which may exceed those of the float solution. This is demonstrated by a specific example in Figure 3.3, which shows a scatterplot of the float and fixed LOS vectors $\hat{\mathbf{x}}$ and $\check{\mathbf{x}}$ based on 10^4 single-epoch solutions. The ambiguities are fixed correctly in only 69.2% of the cases. The errors of $\hat{\mathbf{x}}$ are shown in blue, whereas the corresponding errors of $\check{\mathbf{x}}$ are shown as either red or green dots: red if the ambiguities are fixed incorrectly, green if they are fixed correctly. It can be seen that in case of incorrect integer ambiguity estimation, $\check{\mathbf{x}}$ tends to be of the same size or even much larger than $\hat{\mathbf{x}}$. The underlying reason can be explained by Figure 3.2, where a part of float ambiguities reside in the wrong ISL pull-in regions, resulting in the corresponding fixed solution of the real-valued parameters $\check{\mathbf{x}}$ located in non-physical positions. Therefore, improving the IAR success rate is of crucial importance.

Apart from the float and fixed solutions, another type of solution, called conditional solution $\hat{\mathbf{x}}(\mathbf{a})$, needs to be introduced. It is conditioned on the known integer

Table 3.2: Prior work of applying nonlinear constraints

	Methods	References
Rigorous methods	C-LAMBDA	Teunissen (2006); Park and Teunissen (2009)
	MC-LAMBDA	Teunissen (2007); Giorgi et al. (2012, 2010)
	WC-LAMBDA	Teunissen (2010)
	AC-LAMBDA	Teunissen (2011)
Approximate methods	LC-LAMBDA	Giorgi and Teunissen (2012)
	LWC-LAMBDA	Teunissen (2010)
	Validation	Kuylen et al. (2005); Fan et al. (2005)
	Subset Ambiguity Bounding	Sutton (2002); Monikes et al. (2005)
	ARCE	Park et al. (1996)

candidate \mathbf{a} and has the same precision as $\check{\mathbf{x}}$

$$\hat{\mathbf{x}}(\mathbf{a}) = \hat{\mathbf{x}} - \mathbf{Q}_{\check{\mathbf{x}}\check{\mathbf{a}}} \mathbf{Q}_{\check{\mathbf{a}}\check{\mathbf{a}}}^{-1}(\hat{\mathbf{a}} - \mathbf{a}). \quad (3.16)$$

Provided a correctly fixed integer vector $\check{\mathbf{a}}$, $\hat{\mathbf{x}}(\check{\mathbf{a}})$ is equal to $\check{\mathbf{x}}$.

3.1.3 Constrained LAMBDA

In case of the LOS estimation, the LOS vector is subject to the *a priori* available constraint $\|\mathbf{x}\| = 1$. This information can be exploited to improve the estimation performance. Now, two types of constraints should be clarified: the integer constraints on the ambiguities, and the length constraint (quadratic nonlinear constraint) on the LOS vector. They play a distinct role in the estimation process. The presence of the integer ambiguities enables a *precise* estimation, whereas the presence of the length constraint will enable to achieve a high ambiguity resolution success rate and therefore a *reliable* estimation.

Most of prior work deals with different ways of applying the nonlinear constraint into LAMBDA for attitude determination when the antenna baseline length and/or the baseline geometry are employed as nonlinear constraints. Table 3.2 categorizes two classes of methods: rigorous methods and approximate methods.

Rigorous methods

The first class of methods rigorously incorporates nonlinear constraints and solves the corresponding model rigorously, e.g., the C-LAMBDA (constrained-LAMBDA) method in case of a single constraint $\|\mathbf{x}\| = l$ (Teunissen, 2006; Park and Teunissen, 2009), the WC-LAMBDA (weighted constrained-LAMBDA) method in case of a single constraint with uncertainties $\|\mathbf{x}\| = E(l), \sigma_l^2 = D(l)$ (Teunissen, 2010), and the MC-LAMBDA (multivariate constrained-LAMBDA) method for multivariate orthogonal constraints $\mathbf{R}^T \mathbf{R} = \mathbf{I}_q$ where \mathbf{R} belongs to the class of orthogonal matrices $\mathbf{R} \in \mathbb{O}^{3 \times q}$ (Teunissen, 2007; Giorgi et al., 2010, 2012). The C-LAMBDA and MC-LAMBDA methods in this class assure the highest possible success rates as they have fully and rigorously explored hard constraints, while the WC-LAMBDA method rigorously takes into account a soft constraint with uncertainties. In an extreme case of $\sigma_l^2 \rightarrow 0$, the WC-LAMBDA method reduces to C-LAMBDA. In case of the other extreme, $\sigma_l^2 \rightarrow \infty$, it then has no constraint and becomes the original LAMBDA method.

Table 3.3: Objective functions of the C-LAMBDA, MC-LAMBDA and WC-LAMBDA methods

Methods	Objective function
C-LAMBDA $E(y) = \mathbf{B}\mathbf{x} + \mathbf{A}\mathbf{a}$ $\ \mathbf{x}\ = l$	$\check{\mathbf{a}} = \min_{\mathbf{a} \in \mathbb{Z}^p} J_C(\mathbf{a}) = \min_{\mathbf{a} \in \mathbb{Z}^p} \left(\ \hat{\mathbf{a}} - \mathbf{a}\ _{\mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{a}}}}^2 + H_C(\mathbf{a}, \check{\mathbf{x}}(\mathbf{a})) \right)$ <p>where $\check{\mathbf{x}}(\mathbf{a}) = \min_{\mathbf{x} \in \mathbb{R}^3, \ \mathbf{x}\ =l} H_C(\mathbf{a}, \mathbf{x})$</p> $= \min_{\mathbf{x} \in \mathbb{R}^3, \ \mathbf{x}\ =l} \ \hat{\mathbf{x}}(\mathbf{a}) - \mathbf{x}\ _{\mathbf{Q}_{\hat{\mathbf{x}}(\mathbf{a})\hat{\mathbf{x}}(\mathbf{a})}}^2$
MC-LAMBDA ¹ $E(\mathbf{Y}) = \mathbf{B}\mathbf{X} + \mathbf{A}\mathbf{a}$ $\mathbf{X} = \mathbf{R}\mathbf{X}_{uvw}$ $\mathbf{R}^T \mathbf{R} = \mathbf{I}_q$	$\check{\mathbf{a}} = \min_{\mathbf{a} \in \mathbb{Z}^{p \times r}} J_{MC}(\mathbf{a}) = \min_{\mathbf{a} \in \mathbb{Z}^{p \times r}} \left(\ \text{vec}(\hat{\mathbf{a}} - \mathbf{a})\ _{\mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{a}}}}^2 + H_{MC}(\mathbf{a}, \check{\mathbf{R}}(\mathbf{a})) \right)$ <p>where $\check{\mathbf{X}}(\mathbf{a}) = \min_{\mathbf{R} \in \mathbb{O}^{3 \times q}, \mathbf{R}^T \mathbf{R} = \mathbf{I}_q} H_{MC}(\mathbf{a}, \mathbf{R})$</p> $= \min_{\mathbf{R} \in \mathbb{O}^{3 \times q}, \mathbf{R}^T \mathbf{R} = \mathbf{I}_q} \left\ \text{vec}(\hat{\mathbf{R}}(\mathbf{a}) - \mathbf{R}) \right\ _{\mathbf{Q}_{\hat{\mathbf{R}}(\mathbf{a})\hat{\mathbf{R}}(\mathbf{a})}}^2$
WC-LAMBDA $E(y) = \mathbf{B}\mathbf{x} + \mathbf{A}\mathbf{a}$ $\ \mathbf{x}\ = E(l)$ $\sigma_l^2 = D(l)$	$\check{\mathbf{a}} = \min_{\mathbf{a} \in \mathbb{Z}^p} J_{WC}(\mathbf{a}) = \min_{\mathbf{a} \in \mathbb{Z}^p} \left(\ \hat{\mathbf{a}} - \mathbf{a}\ _{\mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{a}}}}^2 + H_{WC}(\mathbf{a}, \check{\mathbf{x}}(\mathbf{a})) \right)$ <p>where $\check{\mathbf{x}}(\mathbf{a}) = \min_{\mathbf{x} \in \mathbb{R}^3} H_{WC}(\mathbf{a}, \mathbf{x})$</p> $= \min_{\mathbf{x} \in \mathbb{R}^3} \left(\ \hat{\mathbf{x}}(\mathbf{a}) - \mathbf{x}\ _{\mathbf{Q}_{\hat{\mathbf{x}}(\mathbf{a})\hat{\mathbf{x}}(\mathbf{a})}}^2 + \sigma_l^{-2}(l - \ \mathbf{x}\)^2 \right)$

¹ The MC-LAMBDA method extends the C-LAMBDA method from a single constraint to r baseline constraints. A rotation matrix \mathbf{R} is applied to convert \mathbf{X} from the unknown frame xyz into the known frame uvw : $\mathbf{X} = \mathbf{R}\mathbf{X}_{uvw}$, where \mathbf{R} satisfies $\mathbf{R}^T \mathbf{R} = \mathbf{I}_q$ with q equal to 1, 2 or 3, respectively, when r is equal to 1, 2 or greater than 2. The vec in the MC-LAMBDA objective function denotes the vec -operator, which stacks the columns of a $p \times q$ matrix into a column vector of order pq .

Instead of integer minimizing the standard ambiguity objective function $\mathbf{J}(\mathbf{a}) = \|\hat{\mathbf{a}} - \mathbf{a}\|_{\mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{a}}}}^2$, the so-called constrained integer least squares objective functions $\mathbf{J}_C(\mathbf{a})$, $\mathbf{J}_{MC}(\mathbf{a})$ and $\mathbf{J}_{WC}(\mathbf{a})$ are employed, which minimize the sum of the standard ambiguity objective function $\mathbf{J}(\mathbf{a})$ and the baseline objective function $H_C(\mathbf{a}, \mathbf{x})$, $H_{MC}(\mathbf{a}, \mathbf{X})$ or $H_{WC}(\mathbf{a}, \mathbf{x})$ for the C-LAMBDA, MC-LAMBDA and WC-LAMBDA methods, respectively. Table 3.3 gives the associated equations.

These additional baseline objective functions $H_C(\mathbf{a}, \mathbf{x})$, $H_{MC}(\mathbf{a}, \mathbf{X})$ and $H_{WC}(\mathbf{a}, \mathbf{x})$, as the second term in $\mathbf{J}_C(\mathbf{a})$, $\mathbf{J}_{MC}(\mathbf{a})$ and $\mathbf{J}_{WC}(\mathbf{a})$, are conditioned on ambiguity candidates and are usually some orders of magnitude larger than the first term $\|\hat{\mathbf{a}} - \mathbf{a}\|_{\mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{a}}}}$ (Teunissen, 2010; Giorgi, 2011). As a consequence, the search space becomes non-ellipsoidal and its size is so large that many candidates will be unnecessarily examined. Moreover, minimizing the baseline objective function within a constraint, e.g., $\min_{\mathbf{x} \in \mathbb{R}^3, \|\mathbf{x}\|=l} H_C(\mathbf{a}, \mathbf{x})$ for the C-LAMBDA method, needs to be repeatedly calculated for all candidates in the search space. This constrained minimization $\min_{\mathbf{x} \in \mathbb{R}^3, \|\mathbf{x}\|=l} H_C(\mathbf{a}, \mathbf{x})$ can be rewritten in Teunissen (2006, 2007) as the minimization of an ellipsoid $\min_{\mathbf{x} \in \mathbb{R}^3} \|\hat{\mathbf{x}}(\mathbf{a}) - \mathbf{x}\|_{\mathbf{Q}_{\hat{\mathbf{x}}(\mathbf{a})\hat{\mathbf{x}}(\mathbf{a})}}^2$ centered at $\hat{\mathbf{x}}(\mathbf{a})$, subject to a sphere $\|\mathbf{x}\|_{\mathbf{I}_3}^2 = l^2$ centered at origin. The problem then becomes to find the smallest ellipsoid that just touches the sphere. Although singular value decomposition (SVD) or iterative orthogonal projections can be used to rigorously solve the problem, they are still computationally intensive, especially because they have to be repeatedly calculated for a large amount of candidates in the search space.

In Teunissen (2006) and Giorgi and Teunissen (2012), the lower and upper bounding objective functions $\mathbf{J}_{C1}(\mathbf{a})$ and $\mathbf{J}_{C2}(\mathbf{a})$ have been introduced for the C-LAMBDA method in order to shorten the search time. In their methods, the search time can be largely reduced by iteratively and adaptively expanding or shrinking the size of

the search space from the lower or upper bounding functions $\mathbf{J}_{C_1}(\mathbf{a})$ and $\mathbf{J}_{C_2}(\mathbf{a})$

$$\begin{aligned}\mathbf{J}_{C_1}(\mathbf{a}) &= \|\hat{\mathbf{a}} - \mathbf{a}\|_{\mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{a}}}}^2 + \nu_m (\|\hat{\mathbf{x}}(\mathbf{a})\|_{\mathbf{I}_3}^2 - l)^2 \\ \mathbf{J}_{C_2}(\mathbf{a}) &= \|\hat{\mathbf{a}} - \mathbf{a}\|_{\mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{a}}}}^2 + \nu_M (\|\hat{\mathbf{x}}(\mathbf{a})\|_{\mathbf{I}_3}^2 - l)^2\end{aligned}\quad (3.17)$$

where ν_m and ν_M are the smallest and largest eigenvalues of the matrix $\mathbf{Q}_{\hat{\mathbf{x}}(\mathbf{a})\hat{\mathbf{x}}(\mathbf{a})}^{-1}$. This assures $\mathbf{J}_{C_1}(\mathbf{a}) \leq \mathbf{J}_C(\mathbf{a}) \leq \mathbf{J}_{C_2}(\mathbf{a})$. The intensive computation of $\mathbf{H}_C(\mathbf{a}, \hat{\mathbf{x}}(\mathbf{a}))$ in $\mathbf{J}_C(\mathbf{a})$ can be largely avoided by substitutionally calculating only two squared norms $\nu_m (\|\hat{\mathbf{x}}(\mathbf{a})\|_{\mathbf{I}_3}^2 - l)^2$ and $\nu_M (\|\hat{\mathbf{x}}(\mathbf{a})\|_{\mathbf{I}_3}^2 - l)^2$ in these two bounding functions. However, as reported in Giorgi (2011), the results of minimizing the bounding functions may differ from the results of minimizing the constrained least squares. This means $\mathbf{J}_C(\mathbf{a})$ still needs to be evaluated, but in a reduced space. Although the final number of integer vectors in the reduced search space is small, it becomes larger as the strength of the underlying model decreases, e.g., for higher noise (Giorgi, 2011). A similar search strategy is also applied to the MC-LAMBDA method (Teunissen, 2007; Giorgi, 2011; Giorgi et al., 2012, 2010) and the WC-LAMBDA method (Teunissen, 2010).

An alternative rigorous method for the multivariate case, called AC-LAMBDA (Affine constrained LAMBDA), was proposed by Teunissen (2011). This method discards the nonlinear constraints but rigorously includes the remaining linear affine constraints to the search space. Intensive computations can be avoided as the search space remains ellipsoidal and the standard ambiguity objective function can be used. However, it only works in the multivariate case when the constraints can be split into affine and nonlinear constraints (Teunissen, 2011).

Approximate methods

Apart from the rigorous methods, the second class of methods solves the nonlinear constrained integer least squares problem in approximate ways, e.g., the LC-LAMBDA method (the linearized version of C-LAMBDA) (Giorgi and Teunissen, 2012) and LWC-LAMBDA method (the linearized version of WC-LAMBDA) (Teunissen, 2010). In these methods, the non-standard ambiguity objective functions $\mathbf{J}_C(\mathbf{a})$ and $\mathbf{J}_{WC}(\mathbf{a})$ are linearized to quadratic approximations so that quadratic constraints can be treated as linear constraints

$$\begin{aligned}\mathbf{J}_C(\mathbf{a}) &\approx \mathbf{J}_C(\bar{\mathbf{a}}) + (\bar{\mathbf{a}} - \mathbf{a})^T \left(\frac{1}{2}\partial_{aa}^2 \mathbf{J}_C(\bar{\mathbf{a}})\right) (\bar{\mathbf{a}} - \mathbf{a}) \\ &= \mathbf{J}_C(\bar{\mathbf{a}}) + \|\bar{\mathbf{a}} - \mathbf{a}\|_{\left(\frac{1}{2}\partial_{aa}^2 \mathbf{J}_C(\bar{\mathbf{a}})\right)^{-1}}^2\end{aligned}\quad (3.18)$$

$$\begin{aligned}\mathbf{J}_{WC}(\mathbf{a}) &\approx \mathbf{J}_{WC}(\bar{\mathbf{a}}) + (\bar{\mathbf{a}} - \mathbf{a})^T \left(\frac{1}{2}\partial_{aa}^2 \mathbf{J}_{WC}(\bar{\mathbf{a}})\right) (\bar{\mathbf{a}} - \mathbf{a}) \\ &= \mathbf{J}_{WC}(\bar{\mathbf{a}}) + \|\bar{\mathbf{a}} - \mathbf{a}\|_{\left(\frac{1}{2}\partial_{aa}^2 \mathbf{J}_{WC}(\bar{\mathbf{a}})\right)^{-1}}^2\end{aligned}\quad (3.19)$$

where $\partial_{aa}^2 \mathbf{J}_C(\bar{\mathbf{a}})$ and $\partial_{aa}^2 \mathbf{J}_{WC}(\bar{\mathbf{a}})$ are the Hessian of the baseline objective functions evaluated at $\bar{\mathbf{a}}$, $\bar{\mathbf{a}}$ is the constrained float ambiguity solution, which is the best float solution as the constraint has been treated to enable $\bar{\mathbf{a}}$ to be closer to the correct integer. It is thus reasonable to have $\bar{\mathbf{a}}$ as the point of approximation. Table 3.4 gives the objective functions for the LC-LAMBDA and LWC-LAMBDA.

As shown in Table 3.4, the benefit of using the linearized version is that the objective function is a quadratic integer minimization, and the standard LAMBDA search can thus be applied. However, these linearized methods have problems in finding

Table 3.4: Objective functions of the LC-LAMBDA and LWC-LAMBDA methods

Methods	Objective function
LC-LAMBDA	$\check{\mathbf{a}} = \min_{\mathbf{a} \in \mathbb{Z}^p} \ \check{\mathbf{a}} - \mathbf{a}\ _{\left(\frac{1}{2} \partial_{\mathbf{a}\mathbf{a}}^2 J_C(\check{\mathbf{a}})\right)^{-1}}^2$
	where $\check{\mathbf{a}} = \hat{\mathbf{a}} - \mathbf{Q}_{\hat{\mathbf{a}}\check{\mathbf{x}}} \mathbf{Q}_{\check{\mathbf{x}}\check{\mathbf{x}}}^{-1} (\check{\mathbf{x}} - \bar{\mathbf{x}})$
	$\bar{\mathbf{x}} = \min_{\mathbf{x} \in \mathbb{R}^3, \ \mathbf{x}\ =l} \ \check{\mathbf{x}} - \mathbf{x}\ _{\mathbf{Q}_{\check{\mathbf{x}}\check{\mathbf{x}}}}^2$
LWC-LAMBDA	$\check{\mathbf{a}} = \min_{\mathbf{a} \in \mathbb{Z}^p} \ \check{\mathbf{a}} - \mathbf{a}\ _{\left(\frac{1}{2} \partial_{\mathbf{a}\mathbf{a}}^2 J_{WC}(\check{\mathbf{a}})\right)^{-1}}^2$
	where $\check{\mathbf{a}} = \hat{\mathbf{a}} - \mathbf{Q}_{\hat{\mathbf{a}}\check{\mathbf{x}}} \mathbf{Q}_{\check{\mathbf{x}}\check{\mathbf{x}}}^{-1} (\check{\mathbf{x}} - \bar{\mathbf{x}})$
	$\bar{\mathbf{x}} = \min_{\mathbf{x} \in \mathbb{R}^3} \left(\ \check{\mathbf{x}} - \mathbf{x}\ _{\mathbf{Q}_{\check{\mathbf{x}}\check{\mathbf{x}}}}^2 + \sigma_l^{-2} (\ \mathbf{x}\ - l)^2 \right)$

the correct optimum if the quadratic constraint has a short length. As reported in Giorgi and Teunissen (2012), using the LC-LAMBDA method, the ambiguity resolution success rates were lower than the unconstrained original LAMBDA method when l is shorter than 50 m. This is due to the fact that the nonlinearity or statistical curvature becomes more severe for shorter l (Giorgi and Teunissen, 2012).

Several other approximate methods, including, e.g., the validation method (Kuylen et al., 2005; Fan et al., 2005) and the subset ambiguity bounding method (Sutton, 2002; Monikes et al., 2005; Wang et al., 2009) avoid linearization by replacing the hard quadratic equality constraint with soft quadratic inequality boundaries. These methods still aim for integer minimizing the standard ambiguity objective function, but now in a reduced search space (Teunissen, 2010). The idea of these methods is as follows. Since $\hat{\mathbf{x}}(\mathbf{a})$ is a very precise estimator, it can be expected that the length of $\hat{\mathbf{x}}(\mathbf{a})$ is very close to l , provided \mathbf{a} is the correct integer vector. Thus, the hard equality constraint can be represented by soft inequality boundaries $(l - \delta l)^2 \leq \|\hat{\mathbf{x}}(\mathbf{a})\|^2 \leq (l + \delta l)^2$ with a certain threshold δl . The correct integer vector should lie within a set \mathbb{C}

$$\mathbb{C} = \{\mathbf{a} \in \mathbb{Z}^p \mid (l - \delta l)^2 \leq \|\hat{\mathbf{x}}(\mathbf{a})\|^2 \leq (l + \delta l)^2\}. \quad (3.20)$$

The validation method (Kuylen et al., 2005; Fan et al., 2005) is the simplest way to include the length constraint into LAMBDA using soft inequality boundaries. It checks all ambiguity candidates in the standard LAMBDA ambiguity search space. Only the ones that yield $\hat{\mathbf{x}}(\mathbf{a})$ within the boundaries will be accepted for further search of integer least squares minimizer. In Sutton (2002), Monikes et al. (2005), Wang et al. (2009) and Park et al. (1996), ambiguities are divided into a so-called primary and secondary subset. Only the primary subset is used for building the quadratic inequality boundaries of Eq.(3.20), so that the constraint can be implemented in the early stage of search. This method of utilizing only the subset ambiguity is named as subset ambiguity bounding method in this dissertation.

In either the validation method or the subset ambiguity bounding method, the integer least squares minimization problem will be fulfilled in a reduced search space $\mathbb{C} \cap \mathbb{Z}^p$

$$\check{\mathbf{a}} = \min_{\mathbf{a} \in \mathbb{C} \cap \mathbb{Z}^p} \|\hat{\mathbf{a}} - \mathbf{a}\|_{\mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{a}}}}^2. \quad (3.21)$$

The objective function of Eq.(3.21) is in the form of a squared norm, same to

the standard ambiguity objective function, indicating that the validation method and the subset ambiguity bounding method can be more easily and efficiently implemented than the rigorous C-LAMBDA or WC-LAMBDA method. As opposed to the approximate LC-LAMBDA and LWC-LAMBDA methods, the validation and subset ambiguity bounding methods utilize a loose constraint inequality boundary instead of linearizing objective functions, thus being able to work regardless of the length of l .

The following sections focus on elaborating and improving the validation method and the subset ambiguity bounding methods for the LOS estimation. This includes the extended evaluation and comparison of these two methods, the innovative derivations in closed-form for the threshold δl and the insightful proposal of an easy-to-use measure for the constrained ambiguity resolution capability. The impact of antenna geometries will also be thoroughly discussed.

Both the validation method and the subset ambiguity bounding method are approximate methods that treat the constraint in the integer mapping/searching process. As comparison, another approximate method by means of linearization will also be elaborated in the following, where the constraint is treated on the float solution.

3.2 Theory of LOS estimation and associated constrained LAMBDA

Before elaborating specific constrained IAR methods for the LOS estimation, the LOS model, derived in chapter 2, will be rewritten into a generalized carrier phase based observation model. Error sources embedded in the model will be further analysed in order to assure that the remaining error can be regarded as Gaussian random noise.

3.2.1 Single-epoch LOS estimation model

With multiple antennas mounted on the rigid spacecraft platform, inter-satellite ranging signals received by those antennas are processed simultaneously in multiple channels of a single receiver, producing pseudorange and carrier phase observables for each antenna. The single-differenced (SD) observation between a pair of antennas is equal to the antenna baseline projection onto the unit LOS vector

$$\begin{aligned}\Delta\rho_{ij} &= \mathbf{g}_{ij}^T \mathbf{x} + c\Delta t_{ij} + \Delta lb + \Delta\varepsilon_\rho \\ \Delta\phi_{ij} &= \mathbf{g}_{ij}^T \mathbf{x} + c\Delta t_{ij} + \Delta lb + \lambda\Delta N_{ij} + \Delta\varepsilon_\phi\end{aligned}\quad (3.22)$$

subject to $\|\mathbf{x}\| = 1$

where $\mathbf{x} = [x, y, z]^T$ is the LOS vector, which is subject to a unit length constraint, $\Delta\rho_{ij}$ and $\Delta\phi_{ij}$ denote the SD pseudorange and carrier phase measurements between the reference antenna j and the auxiliary antenna i , $\mathbf{g}_{ij} = (g_{x_{ij}}, g_{y_{ij}}, g_{z_{ij}})^T$ is the antenna baseline vector, λ is the carrier wavelength, Δlb is the line bias, $c\Delta t_{ij}$ is the receiver clock error, and initial phase bias, and ΔN_{ij} is the unknown integer ambiguity. Both \mathbf{g}_{ij} and \mathbf{x} are expressed in the body fixed frame.

From the error analysis in chapter 2, it is known that after SD between antennas over a short baseline, the spatially correlated orbital, tropospheric and ionospheric

errors will significantly cancel, and the errors from the same source, like the transmitter clock error, can also be eliminated. However, SD does not eliminate the line bias Δlb due to different cable lengths from two antennas to a single receiver. The line bias can be treated as a constant over time. If antennas are connected to different receivers, i.e., the field demonstration in this chapter, the receiver clock error $c\Delta t_{ij}$ also remains after SD due to different receiver time tags. Driving multiple receivers by an external common clock aids in eliminating the clock drift over time but leaves constant non-zero initial phase bias. Given the fact that the combined effect of the line bias and the initial phase bias is constant over time, they can be lumped together to be pre-calibrated and removed from the SD model. After the bias removal, the remaining error $\Delta\varepsilon_\rho$ and $\Delta\varepsilon_\phi$ can then be regarded as Gaussian distributed random noise and the integer least squares ambiguity resolution can then be applied.

Assuming the bias has been removed, Eq.(3.22) can be extended to n baselines ($n+1$ antennas), m frequencies

$$\begin{bmatrix} \Delta\mathbf{P} \\ \Delta\Phi_{f_1} \\ \Delta\Phi_{f_2} \\ \vdots \\ \Delta\Phi_{f_m} \end{bmatrix} = \begin{bmatrix} \mathbf{G} & & & & \\ \mathbf{G} & \lambda_1\mathbf{I}_n & & & \\ \mathbf{G} & & \lambda_2\mathbf{I}_n & & \\ \vdots & & & \ddots & \\ \mathbf{G} & & & & \lambda_m\mathbf{I}_n \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \Delta\mathbf{N}_{f_1} \\ \Delta\mathbf{N}_{f_2} \\ \vdots \\ \Delta\mathbf{N}_{f_m} \end{bmatrix} + \Delta\varepsilon$$

subject to $\|\mathbf{x}\| = 1$ (3.23)

where $\mathbf{G} = [\mathbf{g}_{1j}^T; \mathbf{g}_{2j}^T; \dots; \mathbf{g}_{nj}^T]$ is the antenna baseline geometry of size $n \times 3$ in the body fixed frame, $\Delta\mathbf{P} = [\Delta\rho_{1j}, \Delta\rho_{2j}, \dots, \Delta\rho_{nj}]^T$ is the vector of the SD pseudorange observation, $\Delta\Phi_{f_*} = [\Delta\phi_{1j,f_*}, \Delta\phi_{2j,f_*}, \dots, \Delta\phi_{nj,f_*}]^T$ is the vector of the SD carrier phase observation of frequency f_* , $\Delta\mathbf{N}_{f_*} = [\Delta N_{1j,f_*}, \Delta N_{2j,f_*}, \dots, \Delta N_{nj,f_*}]^T$ is the corresponding SD ambiguity vector of frequency f_* and λ_* is the associated carrier wavelength.

The set of observations in Eq.(3.23) can now be grouped into a generalized carrier phase based observation model

$$\begin{aligned} E(\mathbf{y}) &= \mathbf{B}\mathbf{x} + \mathbf{A}\mathbf{a} = \begin{bmatrix} \mathbf{B} & \mathbf{A} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{a} \end{bmatrix}, \quad \mathbf{x} \in \mathbb{R}^3, \mathbf{a} \in \mathbb{Z}^p \\ &\text{subject to } \|\mathbf{x}\| = 1 \\ D(\mathbf{y}) &= \mathbf{Q}_{\mathbf{yy}} \end{aligned} \quad (3.24)$$

where \mathbf{y} includes n SD pseudorange and mn carrier phase observations, \mathbf{a} contains p (equal to mn) unknown integer-valued ambiguities $[\Delta\mathbf{N}_{f_1}; \dots; \Delta\mathbf{N}_{f_m}]$, and \mathbf{x} is the real-valued LOS unit vector. Here, \mathbf{A} is the $(m+1)n \times mn$ matrix which contains carrier wavelengths, while \mathbf{B} is the $(m+1)n \times 3$ antenna baseline matrix

$$\mathbf{A} = \begin{bmatrix} \mathbf{0} \\ \text{diag}(\lambda_1, \dots, \lambda_m) \end{bmatrix} \otimes \mathbf{I}_n, \quad \mathbf{B} = \mathbf{e}_{m+1}^T \otimes \mathbf{G} \quad (3.25)$$

where $\text{diag}()$ denotes the diagonal matrix, \mathbf{e}_{m+1} represents a unit column matrix with $m+1$ elements, and \otimes is the Kronecker product. The precision of observations is described by the covariance matrix $\mathbf{Q}_{\mathbf{yy}}$. Although undifferenced observations are independent, a common reference antenna is shared after SD, leading to the non-zero covariance elements in $\mathbf{Q}_{\mathbf{yy}}$.

3.2.2 Bias calibration

Recall that the line bias and initial phase bias are constant over time, they shall be pre-calibrated and removed from the SD model before solving integer ambiguities. A low pass filter, proposed in Keong (1999), is used in this chapter for the constant bias calibration. This low pass filter is formulated as

$$est_k = \frac{k-1}{k}est_{k-1} + \frac{1}{k}res_k \quad (3.26)$$

where est is the estimated bias, which will be fed back to the system at the subsequent epoch, res is defined as a residual bias and k is the epoch counter.

For the pseudorange observation, the residual bias is

$$res_{k,\rho} = \Delta\rho_k - \Delta r_{k-1} \quad (3.27)$$

where Δr_{k-1} is the estimated SD range in the last epoch after the feedback of the estimated bias.

For the phase observation, the residual bias may contain several wavelengths, which will be absorbed by the term of ΔN , leaving a fractional part with magnitude typically at the centimeter level. This fractional part of residual bias is

$$res_{k,\phi} = \Delta\phi_k - \Delta r_{k-1} - \lambda\Delta N_k \quad (3.28)$$

$$\Delta N_k = \left\lceil \frac{\Delta\phi_k - \Delta r_{k-1}}{\lambda} \right\rceil. \quad (3.29)$$

The bias estimation does not need to be as accurate as the LOS estimation, since the integer least squares ambiguity resolution has a certain tolerance to biases (Teunissen et al., 2000; Verhagen, 2012). Once the biases are estimated and fed back to the SD model, a zero mean error will be obtained. The remaining errors are then treated as Gaussian distributed random noise.

3.2.3 Constraint on the float solution

After the bias removal, the LAMBDA method will be modified to address the a-priori available constraint $\|\mathbf{x}\| = 1$ in the LOS model, aiming at improving ambiguity resolution performance.

By parameterizing the unit length of \mathbf{x} in a linearized version, the constraint can be treated as a pseudo-observation and added to the observation model of Eq.(3.24). More specifically, the length of the LOS vector can be represented as

$$l = \sqrt{x^2 + y^2 + z^2} \quad (3.30)$$

where l is equal to 1. It can be linearized by giving an initial guess of $\mathbf{x}_0 = (x_0, y_0, z_0)^T$ and using the Taylor expansion

$$l = l_0 + \mathbf{g}_0^T \begin{bmatrix} \delta x \\ \delta y \\ \delta z \end{bmatrix} \quad (3.31)$$

where $\mathbf{g}_0 = \left[\frac{x_0}{l_0}, \frac{y_0}{l_0}, \frac{z_0}{l_0} \right]^T$. A pseudo-observation \tilde{l} is assumed equal to the true length l plus uncertainty ε_l , then, \tilde{l} is equal to $l_0 + \mathbf{g}_0^T \delta \mathbf{x} + \varepsilon_l$. The observed-minus-computed pseudo-observation $\delta \tilde{l} = \tilde{l} - l_0$ can be added to the model $\delta \mathbf{y} =$

General form $\delta\tilde{\mathbf{y}} = \tilde{\mathbf{B}}\delta\mathbf{x} + \tilde{\mathbf{A}}\mathbf{a} + \tilde{\boldsymbol{\varepsilon}}$

$$\mathbf{Q}_{\tilde{\mathbf{y}}\tilde{\mathbf{y}}} = \text{diag}(\mathbf{Q}_{\mathbf{y}\mathbf{y}}, \sigma_l^2)$$

Iteration required:

$\mathbf{x}_0, \mathbf{a}_0 = \text{zeros}$

for $i = 1$: maximum number of iterations

$$l_0 = \|\mathbf{x}_0\|, \mathbf{g}_0^T = \frac{\mathbf{x}_0}{l_0}, \tilde{\mathbf{B}} = \begin{bmatrix} \mathbf{B} \\ \mathbf{g}_0^T \end{bmatrix}, \tilde{\mathbf{A}} = \begin{bmatrix} \mathbf{A} \\ \mathbf{0} \end{bmatrix}$$

$$\delta\tilde{\mathbf{y}} = \begin{bmatrix} \mathbf{y} \\ l \end{bmatrix} - \begin{bmatrix} \mathbf{B}\mathbf{x}_0 \\ l_0 \end{bmatrix}$$

$$\mathbf{Q}_{\tilde{\mathbf{v}}\tilde{\mathbf{v}}} = \begin{pmatrix} \mathbf{Q}_{\mathbf{x}\mathbf{x}} & \mathbf{Q}_{\mathbf{x}\tilde{\mathbf{a}}} \\ \mathbf{Q}_{\tilde{\mathbf{a}}\mathbf{x}} & \mathbf{Q}_{\tilde{\mathbf{a}}\tilde{\mathbf{a}}} \end{pmatrix} = \left(\begin{bmatrix} \tilde{\mathbf{B}} \\ \tilde{\mathbf{A}} \end{bmatrix} \mathbf{Q}_{\tilde{\mathbf{y}}\tilde{\mathbf{y}}}^{-1} \begin{bmatrix} \tilde{\mathbf{B}} & \tilde{\mathbf{A}} \end{bmatrix} \right)^{-1}$$

$$\begin{bmatrix} \delta\tilde{\mathbf{x}} \\ \delta\tilde{\mathbf{a}} \end{bmatrix} = \mathbf{Q}_{\tilde{\mathbf{v}}\tilde{\mathbf{v}}} \begin{bmatrix} \tilde{\mathbf{B}} \\ \tilde{\mathbf{A}} \end{bmatrix} \mathbf{Q}_{\tilde{\mathbf{y}}\tilde{\mathbf{y}}}^{-1} (\delta\tilde{\mathbf{y}} - \tilde{\mathbf{A}}\mathbf{a}_0)$$

$$\tilde{\mathbf{x}} = \mathbf{x}_0 + \delta\tilde{\mathbf{x}}; \quad \tilde{\mathbf{a}} = \mathbf{a}_0 + \delta\tilde{\mathbf{a}}$$

$$\mathbf{x}_0 = \tilde{\mathbf{x}}; \quad \mathbf{a}_0 = \tilde{\mathbf{a}}$$

if $\text{norm}(\delta\tilde{\mathbf{x}}) < \text{Threshold}$; break; end

end

Figure 3.4: Schematic iteration process to obtain the constrained float solutions

$\mathbf{B}\delta\mathbf{x} + \mathbf{A}\mathbf{a} + \boldsymbol{\varepsilon}$. To this end, an extended constrained observation model becomes

$$\begin{bmatrix} \delta\mathbf{y} \\ \delta\tilde{l} \end{bmatrix} = \begin{bmatrix} \mathbf{B} \\ \mathbf{g}_0^T \end{bmatrix} \delta\mathbf{x} + \begin{bmatrix} \mathbf{A} \\ \mathbf{0} \end{bmatrix} \mathbf{a} + \begin{bmatrix} \boldsymbol{\varepsilon} \\ \varepsilon_l \end{bmatrix} \quad (3.32)$$

where $\delta\mathbf{y} = \mathbf{y} - \mathbf{B}\mathbf{x}_0$ includes the observed-minus-computed SD pseudorange and carrier phase measurements. The error ε_l is assumed to be normal distributed and has variance σ_l^2

$$\begin{aligned} E \begin{bmatrix} \delta\mathbf{y} \\ \delta\tilde{l} \end{bmatrix} &= \begin{bmatrix} \mathbf{B} \\ \mathbf{g}_0^T \end{bmatrix} \delta\mathbf{x} + \begin{bmatrix} \mathbf{A} \\ \mathbf{0} \end{bmatrix} \mathbf{a} \\ D \begin{bmatrix} \delta\mathbf{y} \\ \delta\tilde{l} \end{bmatrix} &= \begin{bmatrix} \mathbf{Q}_{\mathbf{y}\mathbf{y}} & \mathbf{0} \\ \mathbf{0} & \sigma_l^2 \end{bmatrix}. \end{aligned} \quad (3.33)$$

As can be seen, the extended constrained observation model Eq.(3.33) is of the same type as Eq.(3.24). The matrix $\begin{bmatrix} \mathbf{B} & \mathbf{g}_0^T \end{bmatrix}^T$ plays the role of matrix \mathbf{B} and $\delta\mathbf{x}$ takes the place of \mathbf{x} . The measurement covariance includes two independent error contributions $\mathbf{Q}_{\mathbf{y}\mathbf{y}}$ and σ_l^2 .

After a sufficient number of iterations in Figure 3.4, one will obtain the constrained float ambiguity $\tilde{\mathbf{a}}$ as well as its covariance matrix $\mathbf{Q}_{\tilde{\mathbf{a}}\tilde{\mathbf{a}}}$ (part of $\mathbf{Q}_{\tilde{\mathbf{v}}\tilde{\mathbf{v}}}$ in Figure 3.4). The LAMBDA method can be implemented afterwards. However, unlike in the unconstrained model, the distribution of $\tilde{\mathbf{a}}$, governed by $\mathbf{Q}_{\tilde{\mathbf{a}}\tilde{\mathbf{a}}}$, is no longer ellipsoidal, but be an irregular shape. This will result in an inefficient search process and a high possibility of reaching a local instead of a global minimum.

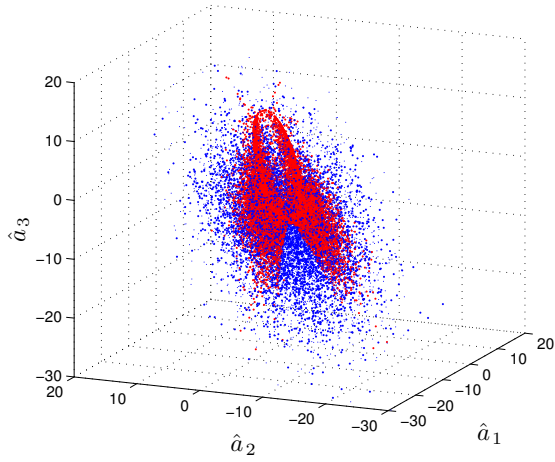


Figure 3.5: Distribution of the constrained (red) and unconstrained (blue) float ambiguities

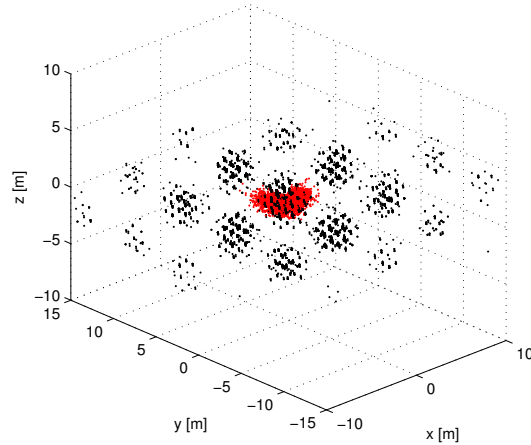


Figure 3.6: Distribution of the constrained float LOS vector (red) and the fixed LOS vectors (black)

Figure 3.5 displays the distributions of unconstrained and constrained float ambiguities by generating 10^4 multivariate normal distributed observations. As can be seen, the shape of the constrained float ambiguities is a scissor-like irregular shape, which lies inside the original unconstrained ellipsoidal shape. If mapping the float ambiguities to the integers based on this non-ellipsoidal shaped ambiguity covariance, a local minimum is usually obtained, easily resulting in an incorrect fixing of the LOS vector. Figure 3.6 illustrates the constrained float and fixed LOS vectors. As shown, even if the float LOS vector fulfils the unit length constraint, fixed solutions still present a high probability of deviation from the correct values, due to the wrong fixing of ambiguities.

The LOS vector can also be expressed by the bearing angles of elevation and azimuth $[el, az]^T$ as $\mathbf{x} = [\cos(el)\cos(az), \cos(el)\sin(az), \sin(el)]^T$. Thus, the LOS can be linearized with respect to these angles. This is mathematically identical to the

above linearization but more computationally intensive. Similar results would be obtained.

Intrinsically, this linearization is similar to the aforementioned LC-LAMBDA (Giorgi and Teunissen, 2012) and LWC-LAMBDA (Teunissen, 2010) methods, which linearize the ambiguity objective function instead of linearizing l . The same conclusion can be obtained that all constrained LAMBDA solutions by means of linearization have problems in finding the global optimum when l is short. According to Giorgi and Teunissen (2012), the underlying reason is that the nonlinearity of the constraint is due to the curved sphere with the radius equal to l , which has larger curvature and higher local nonlinearity for shorter l . Therefore, the irregularity of the constrained ambiguity covariance shape will become more severe for shorter constraint length.

3.2.4 Constraint on the integer mapping process

The other type of methods to deal with the constraint in this chapter is to include the constraint in the integer mapping (or search) process instead of modifying the observation model, so that the ellipsoidal ambiguity covariance can be maintained. Both the validation method and the subset ambiguity bounding method belong to this type. The equality length constraint l is replaced by inequality boundaries $[l - \delta l, l + \delta l]$. As a consequence, the search process can be modified in a way that both targets are hit: minimizing the standard ambiguity objective function $J(\mathbf{a}) = \|\hat{\mathbf{a}} - \mathbf{a}\|_{\mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{a}}}}^2$ by the integer least squares, and accepting only those candidates that yield the conditional LOS vector $\hat{\mathbf{x}}(\mathbf{a})$ with a length in the predefined threshold.

Validation method

The simplest way to do this is to treat the constraint as a validation to accept or reject ambiguity candidates. The all-ambiguity-set is used to calculate the conditional LOS vector $\hat{\mathbf{x}}(\mathbf{a})$, which is a very precise estimator and its norm can be expected to be very close to l , provided the candidate \mathbf{a} is the correctly fixed ambiguity. As a consequence, $\|\hat{\mathbf{x}}(\mathbf{a})\|$ is bounded by $[l - \delta l, l + \delta l]$ in the validation method. This process is expressed as follows

$$\hat{\mathbf{x}}(\mathbf{a}) = \hat{\mathbf{x}} - \mathbf{Q}_{\hat{\mathbf{x}}\hat{\mathbf{a}}} \mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{a}}}^{-1} (\hat{\mathbf{a}} - \mathbf{a}) \quad (3.34)$$

$$\Omega_{\mathbf{a}} = \left\{ \mathbf{a} \in \mathbb{Z}^p \mid \begin{array}{l} \|\hat{\mathbf{a}} - \mathbf{a}\|_{\mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{a}}}} \leq \chi^2 \\ l - \delta l \leq \|\hat{\mathbf{x}}(\mathbf{a})\| \leq l + \delta l \end{array} \right\} \quad (3.35)$$

$$\check{\mathbf{a}} = \min_{\mathbf{a} \in \Omega_{\mathbf{a}} \cap \mathbb{Z}^p} \|\hat{\mathbf{a}} - \mathbf{a}\|_{\mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{a}}}}^2. \quad (3.36)$$

The size of the initial search space χ_0^2 is obtained by rounding or bootstrapping on the unconstrained float ambiguities. If δl is set to a too small value, it is possible that χ_0^2 does not include any candidate that fulfills Eq.(3.35). The searching results of the fixed ambiguities in the set $\Omega_{\mathbf{a}}(\chi_0^2)$ will thus be empty. Therefore, a proper expansion needs to be used to scale up χ_0^2 until, at step s , the set of $\Omega_{\mathbf{a}}(\chi_s^2)$ is non-empty. On the other hand, a too-big δl loses the benefit of using the constraint.

Before exploring a proper δl , the subset ambiguity bounding method will be primarily introduced in the following as comparison with this validation method. The threshold δl will be derived afterwards.

Subset ambiguity bounding method

In contrast, the subset ambiguity bounding method divides ambiguities into two sets: the primary set with the first three ambiguities and the second set with the rest of ambiguities. The mapping function is established which maps only the primary set into the conditional LOS vector $\hat{\mathbf{x}}(\mathbf{a})$, so that the length constraint can be fulfilled in the early stage of search

$$\mathbf{a} = [\mathbf{a}_p \quad \mathbf{a}_s], \mathbf{a}_p \in \mathbb{Z}^3, \mathbf{a}_s \in \mathbb{Z}^{p-3}. \quad (3.37)$$

The conditional LOS vector $\hat{\mathbf{x}}(\mathbf{a})$ is now calculated from carrier phase observations directly

$$\hat{\mathbf{x}}(\mathbf{a}_p) = \mathbf{G}_p^{-1}(\Delta\Phi_p - \mathbf{A}_p\mathbf{a}_p) \quad (3.38)$$

where \mathbf{G}_p corresponds to the first three rows of the \mathbf{G} matrix and represents three of the antenna baselines, $\mathbf{A}_p = \lambda\mathbf{I}_3$ contains the wavelength of the carrier frequency. Only three ambiguities are grouped into the primary set in order to assure the corresponding \mathbf{G}_p matrix is a full rank matrix and invertible.

Since the length of the LOS vector \mathbf{x} can also be written as

$$l^2 = \|\mathbf{x}\|^2 = \mathbf{x}^T \mathbf{x}, \quad (3.39)$$

substituting Eq.(3.38) into (3.39) and replacing l^2 by lower and upper boundaries $(l - \delta l)^2$ and $(l + \delta l)^2$, the constraint inequality can be written as

$$(l - \delta l)^2 \leq (\Delta\Phi_p - \mathbf{A}_p\mathbf{a}_p)^T \mathbf{G}_p^{-T} \mathbf{G}_p^{-1} (\Delta\Phi_p - \mathbf{A}_p\mathbf{a}_p) \leq (l + \delta l)^2. \quad (3.40)$$

This can be translated to the same form as the inequality $\|\hat{\mathbf{a}} - \mathbf{a}\|_{\mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{a}}}}^2 \leq \chi^2$,

$$\Omega_{\mathbf{a}_p, -\delta l} = \left\{ \mathbf{a}_p \in \mathbb{Z}^3 \mid \|\Delta\Phi_p - \mathbf{A}_p\mathbf{a}_p\|_{\mathbf{G}_p\mathbf{G}_p^T}^2 \leq (l - \delta l)^2 \right\} \quad (3.41)$$

$$\Omega_{\mathbf{a}_p, +\delta l} = \left\{ \mathbf{a}_p \in \mathbb{Z}^3 \mid \|\Delta\Phi_p - \mathbf{A}_p\mathbf{a}_p\|_{\mathbf{G}_p\mathbf{G}_p^T}^2 \leq (l + \delta l)^2 \right\}. \quad (3.42)$$

To solve the inequality of Eq.(3.41) and (3.42), the same sequential conditional adjustment used in the LAMBDA can be applied (de Jonge and Tiberius, 1996). Specifically, by KHK^T -decomposition to the matrix $\mathbf{G}_p^{-T} \mathbf{G}_p^{-1}$ using the lower triangular matrix \mathbf{K} and the diagonal matrix \mathbf{H} , the left-hand side of the quadratic inequality, e.g., in Eq.(3.41) can be written as

$$\sum_{i=1}^3 h_i \left[(\lambda a_i - \Delta\Phi_{pi}) + \sum_{j=1}^{i-1} k_{ij} (\lambda a_j - \Delta\Phi_{pj}) \right]^2 \leq (l - \delta l)^2 \quad (3.43)$$

where h_i and k_{ij} are the diagonal elements in \mathbf{H} and the lower triangular elements in \mathbf{K} , respectively. The sequential conditional adjustment is then performed by rewriting Eq.(3.43) in three sequential intervals for searching each of the ambiguities

$$\begin{aligned} (\lambda a_{p1} - \Delta\Phi_{p1})^2 &\leq \frac{(l - \delta l)^2}{h_1} \\ (\lambda a_{p2} - \Delta\Phi_{p2|p1})^2 &\leq \frac{(l - \delta l)^2 - h_1(\lambda a_{p1} - \Delta\Phi_{p1})^2}{h_2} \\ (\lambda a_{p3} - \Delta\Phi_{p3|p1,p2})^2 &\leq \frac{(l - \delta l)^2 - h_1(\lambda a_{p1} - \Delta\Phi_{p1})^2 - h_2(\lambda a_{p2} - \Delta\Phi_{p2|p1,p2})^2}{h_3} \end{aligned} \quad (3.44)$$

where the conditional estimates for $\Delta\Phi_{p2}$ and $\Delta\Phi_{p3}$ are defined as $\Delta\Phi_{p2|p1}$ and $\Delta\Phi_{p3|p1,p2}$, which are conditioned on the previous estimated integers of a_{p1} and a_{p1}, a_{p2}

$$\begin{aligned}\Delta\Phi_{p2|p1} &= \Delta\Phi_{p2} - k_{21}(\lambda a_{p1} - \Delta\Phi_{p1}) \\ \Delta\Phi_{p3|p1,p2} &= \Delta\Phi_{p3} - k_{31}(\lambda a_{p1} - \Delta\Phi_{p1} - k_{32}(\lambda a_{p2} - \Delta\Phi_{p2})).\end{aligned}\quad (3.45)$$

Similar expressions can also be written for the upper bound of $(l + \delta l)^2$ in Eq. (3.42).

Note that simultaneously bounding each of the sequential intervals by both the lower and upper boundaries can result in an empty set. To assure the non-emptiness, \mathbf{a}_p are firstly searched in the lower boundary $(l - \delta l)^2$, leading to a set of ambiguity candidates in $\Omega_{\mathbf{a}_p, -\delta l}$, and then the upper boundary $(l + \delta l)^2$ is sequentially performed in a way that only the candidates in set $\Omega_{\mathbf{a}_p, +\delta l}$ without $\Omega_{\mathbf{a}_p, -\delta l}$ are collected for the further search in the integer least squares.

The entire search procedure can then be expressed by the following equations

$$\mathbb{C}^3 = \Omega_{\mathbf{a}_p, +\delta l} \setminus \Omega_{\mathbf{a}_p, -\delta l} \quad (3.46)$$

$$\Omega_{\mathbf{a}_p} = \left\{ \mathbf{a}_p \in \mathbb{C}^3 \cap \mathbb{Z}^3 \mid \|\hat{\mathbf{a}}_p - \mathbf{a}_p\|_{\mathbf{Q}_{\hat{\mathbf{a}}_p \hat{\mathbf{a}}_p}}^2 \leq \chi^2 \right\} \quad (3.47)$$

$$\Omega_{\mathbf{a}_s | \mathbf{a}_p} = \left\{ \mathbf{a}_s \in \mathbb{Z}^{p-3} \mid \begin{array}{l} \|\hat{\mathbf{a}}_s - \mathbf{a}_s\|_{\mathbf{Q}_{\hat{\mathbf{a}}_s \hat{\mathbf{a}}_s}} \leq \chi^2, \\ \mathbf{a}_s \text{ conditioned on } \mathbf{a}_p \in \Omega_{\mathbf{a}_p} \end{array} \right\} \quad (3.48)$$

$$\mathbb{C}^p = \Omega_{\mathbf{a}_p} \cup \Omega_{\mathbf{a}_s | \mathbf{a}_p} \quad (3.49)$$

$$\hat{\mathbf{a}} = \min_{\mathbf{a} \in \mathbb{C}^p \cap \mathbb{Z}^p} \|\hat{\mathbf{a}} - \mathbf{a}\|_{\mathbf{Q}_{\hat{\mathbf{a}} \hat{\mathbf{a}}}}^2. \quad (3.50)$$

Eq.(3.46) means that an integer set \mathbb{C}^3 for the first three ambiguities only accepts the candidates in set $\Omega_{\mathbf{a}_p, +\delta l}$ exclusive $\Omega_{\mathbf{a}_p, -\delta l}$. The standard quadratic form $\|\hat{\mathbf{a}}_p - \mathbf{a}_p\|_{\mathbf{Q}_{\hat{\mathbf{a}}_p \hat{\mathbf{a}}_p}}^2$ can apply, and now over a smaller region $\mathbb{C}^3 \subset \mathbb{Z}^3$, instead over the complete space \mathbb{Z}^3 . Therefore, the constraint is fulfilled in the smaller region that can exclude some wrong candidates in the early stage of search, leading to the conditional search space in Eq.(3.48) for the rest subset of ambiguities in the set of $\Omega_{\mathbf{a}_s | \mathbf{a}_p}$. To this end, minimizing the objective function (3.50) will be intrinsically different to the minimization of the standard unconstrained objective function because the minimizer is searched in a smaller and more precise integer region $\mathbb{C}^p \subset \mathbb{Z}^p$.

If the integer set $\mathbb{C}^p(\chi_0^2)$ is empty for an initial search volume χ_0^2 , the expansion approach can also be used by scaling up χ_0^2 until, at step s , the set of $\mathbb{C}^p(\chi_s^2)$ is non-empty. However, with a proper choice of δl , it has been numerically demonstrated that this expansion is unnecessary most of the time.

3.2.5 Threshold

The strategy of utilizing the constraint either by the validation method or by the subset ambiguity bounding method can be explained in Figure 3.7, which depicts the float LOS distribution and the unconstrained fixed LOS distributions based on 10^4 epochs of estimates. The blue dots represent the float LOS distribution, which shows an elongated ellipse due to the correlations between its coordinates. The center of the ellipse is the correct LOS solution. The unconstrained fixed LOS distributions are shown as either red or yellow dots: red if the ambiguities are wrongly fixed, yellow if they are correctly fixed. These unconstrained solutions are obtained by

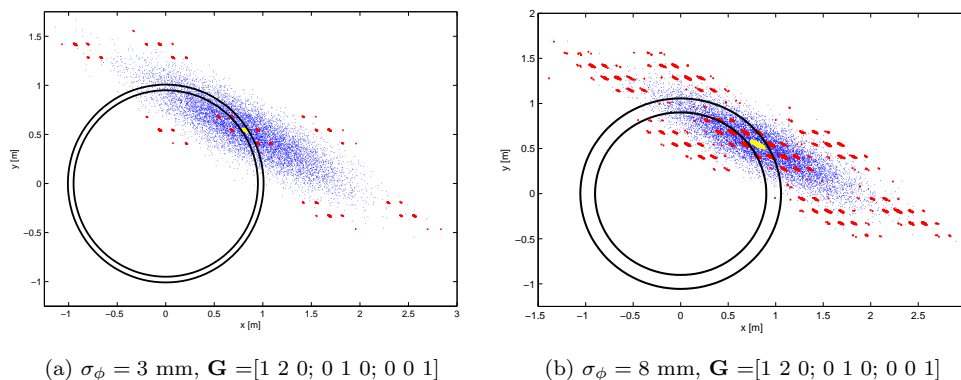


Figure 3.7: Two-dimensional float LOS distributions (blue) and the corresponding unconstrained fixed LOS solutions in yellow or red: yellow if ambiguities are correctly fixed, while red if ambiguities are wrongly fixed. The size of boundary circular rings $1 \pm \delta l$ (black) is adaptive to the model. In (a), 69.2% out of the 10^4 solutions is correctly fixed (yellow), while 30.8% is wrongly fixed (red); In (b), the correctly and wrongly fixed solutions are 13.9% (yellow) and 86.1% (red), respectively. Lower and upper boundaries show the ability to exclude the wrong solutions and remain the correct solutions.

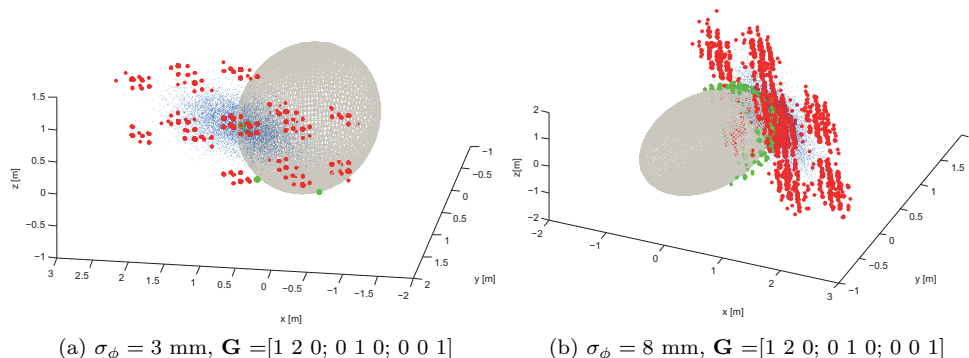


Figure 3.8: Three-dimensional float LOS distributions (blue) and unconstrained (red) and constrained (green) fixed LOS distributions. The constrained results are scattered on the surface of the lower boundary sphere. The upper boundary sphere does not show in the figure for clarity. In (b), note that different scaling is used in three axis.

only applying the minimization in the standard integer least-squares (ILS) as in the case of the original LAMBDA method. It is clear that, without the constraint, only a part of the resultant ILS minimizers can lead to the correct fixed LOS in yellow, while a high percentage of the ILS minimizers cause the fixed LOS distributed in the non-physical locations in red. Therefore, the lower and upper circular boundary rings can play a dramatic role in excluding wrong minimizers. It is easy to reject the wrong minimizers that lead to the LOS distribution far away from the boundary ring. However, it is difficult to exclude the ones that are wrong solutions but still make the resultant LOS drop into the boundary ring. Those minimizers are called false alarm minimizers. The width of the boundary ring determines the percentage of the false

alarm minimizers, which needs to be adaptive to the quality of the model. Making the boundary ring too wide will increase the probability of the wrong minimizers to be turned into the false alarm minimizers, while making it too narrow risks missing the correct solution.

Out of the 10^4 float solutions (blue) in Figure 3.7a and 3.7b, only 69.2% and 13.9%, respectively, have been correctly fixed without the constraint (yellow). After the subset ambiguity bounding, the success rates increase to 99.2% and 39.2%, respectively, as shown in Figure 3.8a and 3.8b where the green dots represent the constrained fixed LOS results. It is clear that these constrained LOS vectors are scattered on the surface of the lower boundary sphere, indicating that the constraint is fulfilled in the ILS.

The constrained LOS distribution is governed by its conditional covariance matrix, which is $\mathbf{Q}_{\hat{\mathbf{x}}(\mathbf{a})\hat{\mathbf{x}}(\mathbf{a})}$ for the validation method and $\mathbf{Q}_{\hat{\mathbf{x}}(\mathbf{a}_p)\hat{\mathbf{x}}(\mathbf{a}_p)}$ for the bounding method

$$\begin{cases} \mathbf{Q}_{\hat{\mathbf{x}}(\mathbf{a})\hat{\mathbf{x}}(\mathbf{a})} = \frac{\sigma_\phi^2}{m} (\mathbf{G}^T \mathbf{W} \mathbf{G})^{-1} & \text{Validation} \\ \mathbf{Q}_{\hat{\mathbf{x}}(\mathbf{a}_p)\hat{\mathbf{x}}(\mathbf{a}_p)} = \sigma_\phi^2 (\mathbf{G}_p^T \mathbf{W}_p \mathbf{G}_p)^{-1} & \text{Bounding} \end{cases} \quad (3.51)$$

where $\mathbf{W} = (\mathbf{D}\mathbf{D}^T)^{-1}$ with $\mathbf{D} = [\mathbf{I}_n, -\mathbf{e}_n]$ as the SD operator for n baselines, and $\mathbf{W}_p = (\mathbf{D}_p\mathbf{D}_p^T)^{-1}$ with $\mathbf{D}_p = [\mathbf{I}_3, -\mathbf{e}_3]$ for three baselines, and m is the number of carrier frequencies. This equation shows that the conditional LOS covariance depends upon the measurements noise σ_ϕ^2 , the way in which the noise is attenuated by the antenna baseline matrix \mathbf{G} or \mathbf{G}_p , and the number of frequencies m in case of the validation method. It is clear that the validation method has more precise covariance matrix due to the usage of all baselines and all frequencies, whereas the bounding method is only based on three baselines and a single frequency.

Eq.(3.51) derives the conditional LOS accuracy in three dimensions. To determine the boundary threshold δl , it needs to be formulated to the length accuracy in one dimension

$$\begin{cases} \sigma_{\|\hat{\mathbf{x}}(\mathbf{a})\|} = \sqrt{\text{tr}(\mathbf{Q}_{\hat{\mathbf{x}}(\mathbf{a})\hat{\mathbf{x}}(\mathbf{a})})} & \text{Validation} \\ \sigma_{\|\hat{\mathbf{x}}(\mathbf{a}_p)\|} = \sqrt{\text{tr}(\mathbf{Q}_{\hat{\mathbf{x}}(\mathbf{a}_p)\hat{\mathbf{x}}(\mathbf{a}_p)})} & \text{Bounding} \end{cases} \quad (3.52)$$

where $\text{tr}()$ denotes trace.

The constraint bounding is essentially a hypothesis testing procedure using the length of the conditional LOS vector as test statistic. If there is no correlation between the coordinates of the fixed LOS vector, the LOS length is Gaussian distributed and the threshold δl can then to be chosen as $3\sigma_{\|\hat{\mathbf{x}}(\mathbf{a})\|}$ for validation and $3\sigma_{\|\hat{\mathbf{x}}(\mathbf{a}_p)\|}$ for bounding, so that the correct solution will pass the test at a confidence as high as 99.7%. However, in fact, the LOS coordinates are highly correlated, indicating that the 3-sigma boundary ring is too large to exclude the wrong solutions. Figure 3.9 depicts the LOS distributions bounded by both 1-sigma and 3-sigma rings. Variable phase noise and baseline geometries are assumed in Figure 3.9 (a-c), which leads to different distribution shapes and sizes. Compared to (a), (b) is more noisy while (c) has an inferior baseline geometry with smaller angular separations. Boundary rings show different widths in different cases. However, it is clear that out of 10^4 epochs of estimates, the correctly fixed solutions (yellow) all completely fall into the 3-sigma and partly fall into the 1-sigma boundary ring in all cases. On one hand, this verifies the statement that δl can be determined according to the conditional LOS covariance. On the other hand, a compromise between maximally

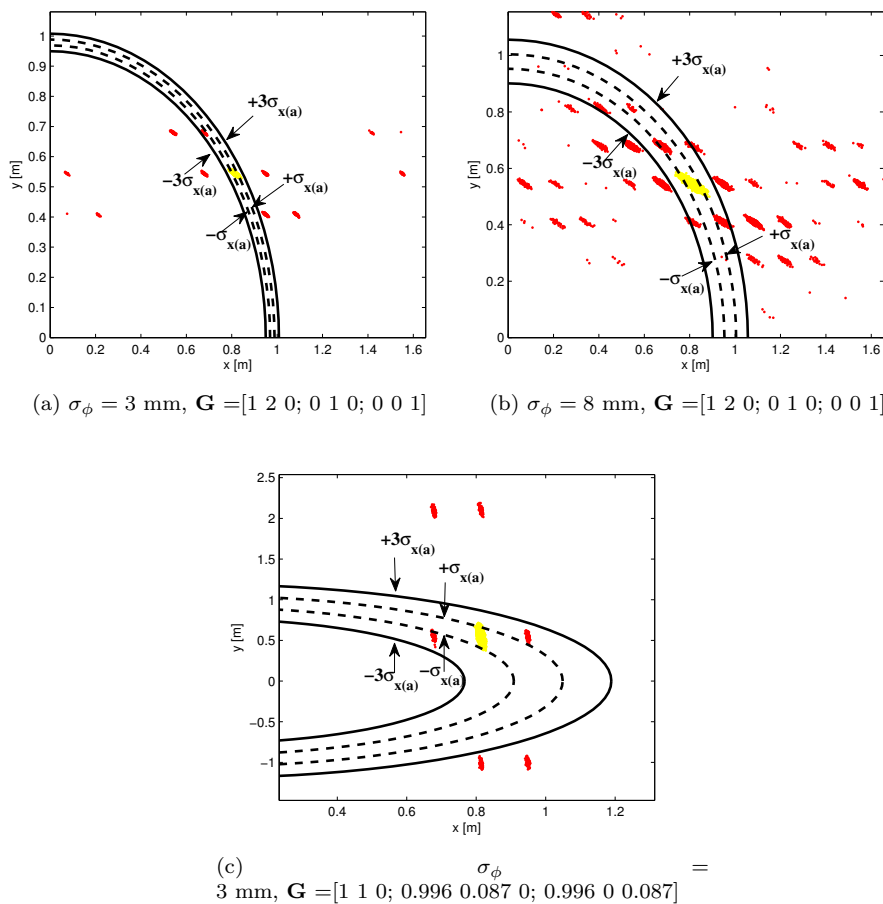
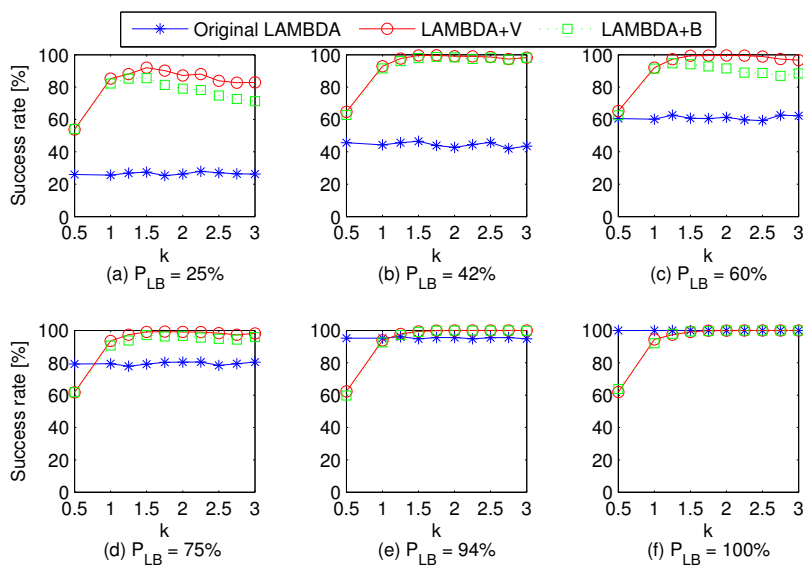


Figure 3.9: Two-dimensional unconstrained fixed solutions (yellow or red): yellow if ambiguities are correctly fixed, red if they are wrongly fixed. The 3-sigma boundary ring (black) is relatively wide to exclude the wrong solutions, while the 1-sigma boundary ring (black dash) is rather narrow that risks missing the correct solution.

including the correct solution and rejecting the wrong solutions requires δl to be chosen between 1-sigma and 3-sigma.

The threshold δl is assumed k times $\sigma_{\|\hat{\mathbf{x}}(\mathbf{a})\|}$ for the validation method (LAMBDA+V) and k times $\sigma_{\|\hat{\mathbf{x}}(\mathbf{a}_p)\|}$ for the subset ambiguity bounding method (LAMBDA+B). Simulations are performed with k ranging from 0.5 to 3 for different configurations in the model given different code and phase noise variances, variable baseline numbers and geometries, dual or triple frequencies. If any of these parameters change, the quality of the model will change. The bootstrapping lower bound success rate P_{LB} can serve as an indicator to characterize the quality of the model at different configurations (Verhagen, 2005).

As shown in Figure 3.10, the bootstrapping success rate is ranging from 25% to 100%. For all ranges of k , the empirical success rate is calculated accordingly. Being independent of δl , the empirical success rates for the original LAMBDA (in blue) keep to reaching the same value, although a small oscillation exists due to the use of different sets of random data for different k . With the constraint in either validation

Figure 3.10: The effect of δl on the success rate

or bounding methods, the largest empirical success rate then occurs in Figure 3.10 (a) to (c) when k is close to 1.5, while from (b) to (d), the largest success rate keeps approaching 1 when k is larger than 2. Thus, for low quality models, only a suitable point of k assures the best use of constraint. In contrast, for high quality models, the performance is better with larger k . The rule-of-thumb for k is then chosen to 1.75 for $P_{LB} \geq 80\%$ and 3 for $P_{LB} < 80\%$. Note that the validation method provides a slightly better performance than the subset ambiguity bounding method since the usage of all-ambiguity-set gives more precise covariance in $\mathbf{Q}_{\hat{\mathbf{x}}(\mathbf{a})\hat{\mathbf{x}}(\mathbf{a})}$ over the covariance in $\mathbf{Q}_{\hat{\mathbf{x}}(\mathbf{a}_p)\hat{\mathbf{x}}(\mathbf{a}_p)}$ for subset ambiguities.

The threshold δl is then equal to

$$\delta l = \begin{cases} \frac{k\sigma_\phi}{m\sqrt{\text{tr}(\mathbf{G}^T\mathbf{W}\mathbf{G})}} & \text{Validation} \\ \frac{k\sigma_\phi}{\sqrt{\text{tr}(\mathbf{G}_p^T\mathbf{W}_p\mathbf{G}_p)}} & \text{Bounding} \end{cases} \quad (3.53)$$

with the rule-of-thumb for k

$$k = \begin{cases} 1.75 & P_{LB} < 80\% \\ 3 & P_{LB} \geq 80\% \end{cases} \quad (3.54)$$

For the subset ambiguity bounding method, once more than three antenna baselines or more than one frequency are involved, the number of ambiguities will be larger than three. Only three of them need to be chosen for constraint bounding. They will be chosen based on the following two criteria according to Eq.(3.53)

- (1) \mathbf{a}_p are on the frequency that provides smaller phase noise;
- (2) \mathbf{a}_p are for those three baselines that lead to larger values of the entries in $\mathbf{G}_p^T\mathbf{W}_p\mathbf{G}_p$. This can be achieved by choosing three baselines with larger lengths or better geometrical arrangements of their relative positions.

3.3 Antenna geometry aspects

The antenna baseline geometry plays an important role in the determination of δl . Moreover, it crucially influences the ambiguity accuracy and the LOS accuracy. This section emphasizes on the discussion of antenna geometries, including proposing an insightful LOS dilution of precision based on antenna baseline geometrical positions, proposing an easy-to-use measure for the constraint ambiguity resolution capability, and evaluating and suggesting a better geometry for the LOS estimation.

3.3.1 LOS dilution of precision

Following the derivations in Appendix A, the covariance matrices $\mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{a}}}$, $\mathbf{Q}_{\hat{\mathbf{x}}\hat{\mathbf{x}}}$, $\mathbf{Q}_{\hat{\mathbf{x}}(\mathbf{a})\hat{\mathbf{x}}(\mathbf{a})}$ and $\mathbf{Q}_{\hat{\mathbf{x}}(\mathbf{a}_p)\hat{\mathbf{x}}(\mathbf{a}_p)}$ that determine the float ambiguity accuracy, the float LOS accuracy, and the conditional fixed LOS accuracy based on all ambiguities or subset of ambiguities can be expressed as

$$\begin{aligned}\mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{a}}} &= \left(\frac{1}{\sigma_\phi^2} \mathbf{\Lambda}_1 \otimes \mathbf{W} - \frac{\sigma_\rho^2}{\sigma_\phi^2(\sigma_\phi^2 + m\sigma_\rho^2)} (\mathbf{\Lambda}_2^T \mathbf{\Lambda}_2) \otimes (\mathbf{W}\mathbf{G}(\mathbf{G}^T\mathbf{W}\mathbf{G})^{-1}\mathbf{G}^T\mathbf{W}) \right)^{-1} \\ \mathbf{Q}_{\hat{\mathbf{x}}\hat{\mathbf{x}}} &= \sigma_\rho^2 (\mathbf{G}^T\mathbf{W}\mathbf{G})^{-1} \\ \mathbf{Q}_{\hat{\mathbf{x}}(\mathbf{a})\hat{\mathbf{x}}(\mathbf{a})} &= \frac{\sigma_\phi^2}{m} (\mathbf{G}^T\mathbf{W}\mathbf{G})^{-1} \\ \mathbf{Q}_{\hat{\mathbf{x}}(\mathbf{a}_p)\hat{\mathbf{x}}(\mathbf{a}_p)} &= \sigma_\phi^2 (\mathbf{G}_p^T\mathbf{W}_p\mathbf{G}_p)^{-1}\end{aligned}\quad (3.55)$$

where $\mathbf{\Lambda}_1 = \text{diag}(\lambda_1^2, \dots, \lambda_m^2)$, $\mathbf{\Lambda}_2 = (\lambda_1, \dots, \lambda_m)$ are the matrices having the wavelengths as entries, σ_ρ and σ_ϕ are standard deviations for the undifferenced pseudo-range and carrier phase measurements, and σ_ϕ is at least two orders of magnitude smaller than σ_ρ .

As can be seen, the float LOS accuracy, expressed in $\mathbf{Q}_{\hat{\mathbf{x}}\hat{\mathbf{x}}}$, is mainly determined by the code noise while the conditional fixed LOS accuracy of $\mathbf{Q}_{\hat{\mathbf{x}}(\mathbf{a})\hat{\mathbf{x}}(\mathbf{a})}$ and $\mathbf{Q}_{\hat{\mathbf{x}}(\mathbf{a}_p)\hat{\mathbf{x}}(\mathbf{a}_p)}$ will increase to the precision that is in accordance with the high precision of phase measurements. In addition, having more frequencies improves the accuracy of $\mathbf{Q}_{\hat{\mathbf{x}}(\mathbf{a})\hat{\mathbf{x}}(\mathbf{a})}$. Note that the specific choice of the frequency values influences the ambiguity accuracy $\mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{a}}}$, but does not play a role in the determination of the LOS accuracy.

All these covariance matrices depend on the baseline matrix \mathbf{G} . In order to obtain accurate LOS solutions, smaller values of the entries in $(\mathbf{G}^T\mathbf{W}\mathbf{G})^{-1}$ are expected. This can be achieved by increasing the number of antenna baselines, lengthening the baselines or properly arranging the relative orientations between the baselines. Similar to the GNSS-based positioning that is performed by using four or more satellites, the LOS estimation for a single signal source but having multiple antennas can play the same role of ensuring a good geometry diversity and observability. Baseline information in the matrix $(\mathbf{G}^T\mathbf{W}\mathbf{G})^{-1}$ can thus reflect the LOS Dilution of precision (LOSDOP), which is defined regardless of the measurement noise and purely dependent on the number of baselines and their geometries

$$LOSDOP = \sqrt{\text{tr}(\mathbf{G}^T\mathbf{W}\mathbf{G})^{-1}}. \quad (3.56)$$

A smaller LOSDOP assures a better observability and higher precision of the LOS estimation.

3.3.2 Constrained ambiguity dilution of precision

Ambiguity dilution of precision (ADOP) was primarily introduced in Teunissen (1997) and Teunissen and Odijk (1997) as an easy-to-compute scalar diagnostic to measure the expected success rate of ambiguity resolution. It differs to the position-related DOP measures such as the position (PDOP), the vertical (VDOP), the horizontal dilution of precision (HDOP) or the aforementioned LOSDOP. These latter DOP measures are all based on the *trace* of the variance matrix. The ADOP uses the *determinant* of $\mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{a}}}$ instead of *trace* due to the following two main reasons. Firstly, both the variances and the correlations of ambiguities can be taken into account by the determinant. This is of great importance as the ambiguities can be highly correlated, especially in case of short observation times. Secondly, the determinant of the ambiguity covariance is invariant under ambiguity transformations. This is also an important feature as a decorrelation transformation procedure is included in the LAMBDA method, which maximumly decorrelates the covariances between ambiguities and returns new ambiguities that show a dramatic improvement in correlation and precision. Thus, the ADOP does not change during this ambiguity decorrelation procedure if the determinant instead of trace is used. The ADOP is defined as (Teunissen, 1997)

$$ADOP = |\mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{a}}}|^{\frac{1}{2mn}} \text{ [cycle]} \quad (3.57)$$

where $|\cdot|$ means determinant, mn is equal to the number of ambiguities with m as the number of frequencies and n as the number of baselines.

The ADOP can be linked to an approximated value of the success rate P_{ADOP} (Teunissen, 1998)

$$P_{ADOP} = \left[2\Phi\left(\frac{1}{2ADOP}\right) - 1 \right]^{mn} \quad (3.58)$$

with $\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}v^2\right\} dv$. Figure 3.11 shows the P_{ADOP} as function of the ADOP for varying number of ambiguities. It can be seen that the ADOP-based success rate decreases as the ADOP increases, and this decrease is steeper when more ambiguities need to be estimated (Teunissen, 2011). For ADOP values smaller than 0.12 cycles, the approximated success rate P_{ADOP} can be expected higher than 0.99.

For the LOS estimation, the determinant of $\mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{a}}}$ is derived in Appendix B in a closed-form expression

$$|\mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{a}}}| = m^3(n+1)^m \frac{\sigma_\rho^6 \sigma_\phi^{2mn-6}}{\left(\prod_{i=1}^m \lambda_i\right)^{2n}}. \quad (3.59)$$

This expression shows all factors that contribute to the ambiguity accuracy. Substituting Eq.(3.59) into (3.58), the ADOP can be obtained, which gets smaller for smaller code variance, or larger number of frequencies and antenna baselines. The ADOP is proportional to the phase variance under the condition that $mn \geq 3$. This condition should be fulfilled to avoid the ADOP changing to an extremely large value when ambiguities are unsolvable.

The ADOP of Eq.(3.59) only quantifies the advantage of having a higher number of baselines, but it is independent on the baseline relative geometries (baseline lengths and relative orientations). This conclusion holds true only in the situation

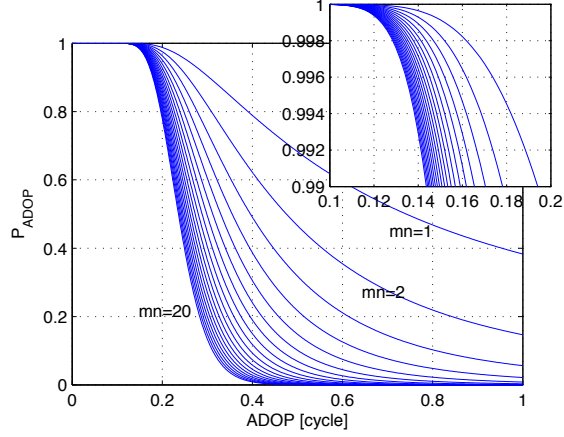


Figure 3.11: P_{ADOP} versus ADOP for varying number of ambiguities $1 \leq mn \leq 20$ (Teunissen, 2011; Odijk et al., 2008)

without the LOS constraint validation or bounding as it is calculated by the unconstrained float ambiguities $\mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{a}}}$. When the conditional LOS solution is used to validate or bound the search of integers, the accuracy of the conditional LOS solution in $\mathbf{Q}_{\hat{\mathbf{x}}(\mathbf{a})\hat{\mathbf{x}}(\mathbf{a})}$ or $\mathbf{Q}_{\hat{\mathbf{x}}(\mathbf{a}_p)\hat{\mathbf{x}}(\mathbf{a}_p)}$ has to be taken into account. The baseline geometry that is embedded inside $\mathbf{Q}_{\hat{\mathbf{x}}(\mathbf{a})\hat{\mathbf{x}}(\mathbf{a})}$ or $\mathbf{Q}_{\hat{\mathbf{x}}(\mathbf{a}_p)\hat{\mathbf{x}}(\mathbf{a}_p)}$ will then play an important role in the definition of the constrained ADOP.

It is not trivial to express the nonlinear quadratic constrained ADOP. Even if the constraint has been replaced by inequality boundaries, the inequality function is still quadratic. However, it is relatively easy to address when the constraint can be expressed in a linear function according to Teunissen (2011).

Assuming the length-constrained variance matrix is $\mathbf{Q}_{\hat{\mathbf{i}}\hat{\mathbf{i}}}$, and its correlations with ambiguities are described by matrices $\mathbf{Q}_{\hat{\mathbf{i}}\hat{\mathbf{a}}}$ or $\mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{i}}}$, the determinant factorization property can be used for the derivation of the determinant of the constrained ambiguity covariance $|\mathbf{Q}_{\hat{\mathbf{a}}(l)\hat{\mathbf{a}}(l)}|$

$$\left| \begin{bmatrix} \mathbf{Q}_{\hat{\mathbf{i}}\hat{\mathbf{i}}} & \mathbf{Q}_{\hat{\mathbf{i}}\hat{\mathbf{a}}} \\ \mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{i}}} & \mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{a}}} \end{bmatrix} \right| = |\mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{a}}}| |\mathbf{Q}_{\hat{\mathbf{i}}(\mathbf{a})\hat{\mathbf{i}}(\mathbf{a})}| = |\mathbf{Q}_{\hat{\mathbf{a}}(l)\hat{\mathbf{a}}(l)}| |\mathbf{Q}_{\hat{\mathbf{i}}\hat{\mathbf{i}}}|. \quad (3.60)$$

Thus, we get

$$|\mathbf{Q}_{\hat{\mathbf{a}}(l)\hat{\mathbf{a}}(l)}| = \frac{|\mathbf{Q}_{\hat{\mathbf{i}}(\mathbf{a})\hat{\mathbf{i}}(\mathbf{a})}|}{|\mathbf{Q}_{\hat{\mathbf{i}}\hat{\mathbf{i}}}|} |\mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{a}}}|. \quad (3.61)$$

Redefine the unconstrained ADOP as $ADOP_{\infty}$ and the constrained ADOP with boundaries as $ADOP_{\delta l}$. Since $ADOP_{\infty} = |\mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{a}}}|^{\frac{1}{2mn}}$ while $ADOP_{\delta l} = |\mathbf{Q}_{\hat{\mathbf{a}}(l)\hat{\mathbf{a}}(l)}|^{\frac{1}{2mn}}$, the $ADOP_{\delta l}$ can be obtained as

$$ADOP_{\delta l} = |\mathbf{Q}_{\hat{\mathbf{a}}(l)\hat{\mathbf{a}}(l)}|^{\frac{1}{2mn}} = \left(\frac{|\mathbf{Q}_{\hat{\mathbf{i}}(\mathbf{a})\hat{\mathbf{i}}(\mathbf{a})}|}{|\mathbf{Q}_{\hat{\mathbf{i}}\hat{\mathbf{i}}}|} \right)^{\frac{1}{2mn}} ADOP_{\infty}. \quad (3.62)$$

This equation shows that the effect of the length constraint l on the ADOP is driven by the ratio $|\mathbf{Q}_{\hat{\mathbf{i}}(\mathbf{a})\hat{\mathbf{i}}(\mathbf{a})}|/|\mathbf{Q}_{\hat{\mathbf{i}}\hat{\mathbf{i}}}|$ and thus by the gain in precision of estimating

l when knowing \mathbf{a} . If knowledge of \mathbf{a} improves the capability of estimating l , then so will knowledge of l improve the ADOP (Teunissen, 2011). By linearization of the length constraint l by giving an initial approximation with $\mathbf{g}_0 = \begin{bmatrix} \frac{x_0}{l_0} & \frac{y_0}{l_0} & \frac{z_0}{l_0} \end{bmatrix}^T$,

$$l = \mathbf{g}_0^T \mathbf{x} + \varepsilon_l, \quad (3.63)$$

the variance matrices for the float and conditional length constraint $\mathbf{Q}_{\hat{l}\hat{l}}$ and $\mathbf{Q}_{\hat{l}(\mathbf{a})\hat{l}(\mathbf{a})}$ can then be linearly linked to the unconstrained float and conditional LOS covariance $\mathbf{Q}_{\hat{\mathbf{x}}\hat{\mathbf{x}}}$ and $\mathbf{Q}_{\hat{\mathbf{x}}(\mathbf{a})\hat{\mathbf{x}}(\mathbf{a})}$ given the constraint threshold δl

$$\mathbf{Q}_{\hat{l}\hat{l}} = \mathbf{g}_0 \mathbf{Q}_{\hat{\mathbf{x}}\hat{\mathbf{x}}} \mathbf{g}_0^T + \delta l^2 \quad (3.64)$$

$$\mathbf{Q}_{\hat{l}(\mathbf{a})\hat{l}(\mathbf{a})} = \mathbf{g}_0 \mathbf{Q}_{\hat{\mathbf{x}}(\mathbf{a})\hat{\mathbf{x}}(\mathbf{a})} \mathbf{g}_0^T + \delta l^2. \quad (3.65)$$

Substituting the derivations for $\mathbf{Q}_{\hat{\mathbf{x}}\hat{\mathbf{x}}}$ and $\mathbf{Q}_{\hat{\mathbf{x}}(\mathbf{a})\hat{\mathbf{x}}(\mathbf{a})}$ in Eq.(3.55), the closed-form expression for the constrained ADOP is

$$ADOP_{\delta l} = \left(\frac{\frac{\sigma_\phi^2}{m} \mathbf{g}_0 (\mathbf{G}^T \mathbf{W} \mathbf{G})^{-1} \mathbf{g}_0^T + \delta l^2}{\sigma_\rho^2 \mathbf{g}_0 (\mathbf{G}^T \mathbf{W} \mathbf{G})^{-1} \mathbf{g}_0^T + \delta l^2} \right)^{\frac{1}{2mn}} \cdot ADOP_\infty \quad (3.66)$$

where δl is given in Eq.(3.53) for either the validation or the subset ambiguity bounding method, and $ADOP_\infty$ is given in Eq.(3.59) and (3.58). As described in section 3.2.3, after linearization, the constrained ambiguity covariance becomes non-elliptical. Although this non-elliptical shape impedes the subsequent searching process from reaching a global optimum, it does improve the precision of the ambiguity covariance. Thus, it is reasonable to use a suitable linearization to determine $ADOP_{\delta l}$ and predict the expected success rate. The $ADOP_{\delta l}$ is always smaller than $ADOP_\infty$. This demonstrates that the model is stronger by the addition of a constraint. By numerical simulations, it is verified that the choice of \mathbf{g}_0 hardly changes the value of $ADOP_{\delta l}$. Therefore, the $ADOP_{\delta l}$ provides an easy-to-use rule-of-thumb for the ambiguity resolution capabilities with constraint boundaries.

3.3.3 Antenna geometry

As shown in sections 3.2.5, 3.3.1 and 3.3.2, the baseline geometry plays an important role in the determination of δl , $LOSDOP$ and $ADOP_{\delta l}$. The baseline coordinate matrix \mathbf{G} consequently affects both the ambiguity resolution performance and the LOS accuracy.

In order to illustrate the influence of the baseline number, baseline lengths and relative orientations between baselines on δl , $LOSDOP$ and $ADOP_{\delta l}$, we consider four different baseline geometries: **G1** with “good” angular separations and 1 m baseline length; as compared to **G1**, **G2** has halved baseline length of 0.5 m; **G3** has “badly” distributed baselines with much smaller angular separations; and **G4** reduces the baseline lengths with different ratios, ranging from 0.8 m to 0.2 m. These baseline geometries are shown in Figure 3.12 and their spherical coordinates are given in Table 3.5.

Table 3.6 reports the value of the $LOSDOP$, $ADOP_\infty$, $ADOP_{\delta l}$ and the empirical success rate for these four configurations.

It is shown that the $LOSDOP$ is smallest for **G1** and it becomes larger the shorter the baseline lengths (**G2**, **G4**) or the smaller the angles between baselines (**G3**).

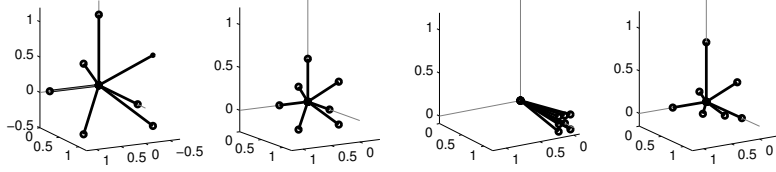
Figure 3.12: Antenna baseline geometries: from left to right **G1**, **G2**, **G3**, **G4**

Table 3.5: Spherical coordinates of four baseline configurations

		\mathbf{g}_1	\mathbf{g}_2	\mathbf{g}_3	\mathbf{g}_4	\mathbf{g}_5	\mathbf{g}_6	\mathbf{g}_7
G1	r [m]	1	1	1	1	1	1	1
	az [°]	0	90	0	45	45	-45	-45
	el [°]	0	0	90	-45	45	-45	45
G2	r [m]	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	az [°]	0	90	0	45	45	-45	-45
	el [°]	0	0	90	-45	45	-45	45
G3	r [m]	1	1	1	1	1	1	1
	az [°]	0	5	0	5	5	-5	-5
	el [°]	0	5	0	-5	5	-5	5
G4	r [m]	0.8	0.7	0.6	0.5	0.4	0.3	0.2
	az [°]	0	90	0	45	45	-45	-45
	el [°]	0	0	90	-45	45	-45	45

This indicates that a good baseline geometry in the sense of better observability and higher LOS precision should be the one that has its baseline elements highly angularly separated with longer lengths. However, in the sense of better ambiguity resolution performance, this does not hold true. Compared to **G1**, the baseline lengths are all halved in **G2** while they are diversely shortened with ratios ranging from 0.8 to 0.2 in **G4**. Although both **G2** and **G4** have shorter lengths, their success rates, either unconstrained or constrained, go opposite directions. The success rates of LAMBDA+V for **G2** are around 2% (11.8%-9.2%)~18% (73.3%-55.9%) lower than **G1**, however, they remarkably increase by up to 27% (99.9%-73.3%) in **G4**. The success rates of the unconstrained LAMBDA for **G4** are up to 77% (83.8%-6.3%) larger than **G1**. This means diverse baseline lengths can benefit the ambiguity resolution much more than longer ones.

Compared to **G1**, **G3** has badly distributed baselines with very small angular separations. The unconstrained success rates are reduced as expected. However, an unexpected phenomenon can be seen that the constrained success rates for **G3** are comparable or even 10% higher than **G1**. The reason is that bad angular separations result in wrong solutions scattered far away from the correct solution and can thus be easily distinguished by a proper choice of δl . This can be seen from Figure 3.9c. Thus, the impact of bad angular separations can be dramatically compensated once δl is chosen properly.

The ADOP, as the indicator of the ambiguity resolution capability, is also shown in Table 3.6. The ADOP_∞ is unconstrained and can only quantify the advantage of having a higher number of baselines, but it is independent on the baseline length or angular separations according to Eq.(3.59). The same value of ADOP_∞ is thus obtained for different baseline geometries. The $\text{ADOP}_{\delta l}$ is largely decreased than

Table 3.6: $LOSDOP$, $ADOP_\infty$, $ADOP_{\delta l}$ and empirical success rate P_E for different number of baselines and baseline geometries based on 10^3 simulated epochs. Single-epoch, single-frequency performances for the original LAMBDA, LAMBDA with validation (LAMBDA+V) and LAMBDA with subset ambiguity bounding (LAMBDA+B) are evaluated. Simulation conditions: carrier frequency is in S-band at 2271.06 MHz, code and phase noise are $\sigma_\rho = 0.3$ m and $\sigma_\phi = 3$ mm, respectively.

G	n	$LOSDOP$	Original LAMBDA		LAMBDA+V		LAMBDA+B	
			$ADOP_\infty$	P_E [%]	$ADOP_{\delta l}$	P_E [%]	$ADOP_{\delta l}$	P_E [%]
G1	3	2.4495	3.0312	0.3	0.9075	11.8	0.9075	11.8
	4	2.2728	0.9303	2.1	0.3802	47.6	0.3865	47.3
	5	1.8708	0.4561	3.1	0.2275	61.3	0.2390	57.4
	6	1.6018	0.2829	4.5	0.1576	70.2	0.1682	65.0
	7	1.5207	0.2009	6.3	0.1215	73.3	0.1294	68.5
G2	3	4.8990	3.0312	0.2	0.9075	9.3	0.9075	9.4
	4	4.5456	0.9303	2.4	0.3802	44.2	0.3865	43.1
	5	3.7417	0.4561	3.4	0.2275	52.6	0.2390	50.3
	6	3.2036	0.2829	4.4	0.1576	56.2	0.1682	53.7
	7	3.0414	0.2009	4.3	0.1215	55.9	0.1294	54.6
G3	3	22.9475	3.0312	0.2	0.9771	16.1	0.9771	16.1
	4	19.8795	0.9303	0.5	0.4170	42.0	0.4313	28.2
	5	13.3026	0.4561	0.7	0.2423	72.4	0.2691	64.2
	6	8.9282	0.2829	1.4	0.1608	80.2	0.1869	78.8
	7	7.4562	0.2009	2.6	0.1233	80.1	0.1438	78.2
G4	3	3.5724	3.0312	0.4	0.8999	7.1	0.8999	7.1
	4	3.1051	0.9303	2.9	0.3820	43.3	0.3941	41.0
	5	3.0205	0.4561	17.9	0.2246	74.2	0.2315	74.7
	6	2.5507	0.2829	46.9	0.1574	98.1	0.1657	97.8
	7	2.4793	0.2009	83.8	0.1301	99.9	0.1368	99.3

the $ADOP_\infty$, indicating that the capability of reliably solving the ambiguities is remarkably increased with the help of constraint. The slightly higher success rates of the validation method as opposed to the subset ambiguity bounding method can also be reflected in the slightly changed $ADOP_{\delta l}$ values.

3.4 Verification

3.4.1 Numerical simulations

Numerical simulations are set up to investigate the general performance of the constrained ambiguity resolution as a function of antenna baselines, frequencies, and the code and phase noise. Parameters are given in Table 3.7. Different levels of noise are assumed on the undifferenced code (from 1.5 m to 0.1 m) and phase measurements (from 0.009 m to 0.003 m). A set of 10^3 multivariate normal distributed observations was generated, and each data set was created with a different number of antenna baselines varying from 3 to 6 and using different ultra-BOC structures with 1 or 2 separate tones apart from the central carrier. For all these trials, only single epoch observations are used to examine an instantaneous ambiguity resolution performance.

The large values of the code and phase noise $\sigma_\phi = 0.009$ m and $\sigma_\rho = 1.5$ m are conservative values that could include other error sources than thermal noise as for such as multipath. For a single epoch processing, multipath can be considered as

Table 3.7: Numerical simulation set-up for investigating the performance of different IAR methods

Number of baselines	3, 4, 5, 6
Baseline geometries	$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0.3 & 0.8 & 0.5 \\ 1 & 2 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0.7 & 1.2 & 0 \\ 0.3 & 0.8 & 0.5 \\ 1 & 2 & 0 \end{bmatrix}$
Carrier frequency [MHz]	222×10.23
1st separate tone [MHz]	206×10.23
2nd separate tone [MHz]	201×10.23
Code noise σ_ρ [m]	1.5, 1.0, 0.5, 0.1
Phase noise σ_ϕ [m]	0.009, 0.003
Epochs simulated	10^3

random noise and lumped into the thermal noise as time-correlation property of the multipath is not relevant in single epoch (Buist et al., 2011).

Figure 3.13 illustrates the performances under different parameter combinations. Rather than the empirical success rate, the empirical failure rate is depicted for clarity. Generally, the improvement from the original LAMBDA to the constrained LAMBDA by validation or bounding can reach up to 80%. The ultra-BOC signal structure benefits from its nature of introducing more phase observations on separate tones and enables lower failure rate. For example, in the case of $\sigma_\phi = 0.003$ m with 4 or more antenna baselines, as σ_ρ varying from 1.5 m to 0.1 m, all the constrained failure rates are below 10% with a single tone and below 1% with two tones. Zero failure rates are obtained as more antennas are involved. Having more antennas generally strengthens the model and also increases the observability. When the number of antenna baselines is increased to 5 with properly arranged geometry in Table 3.7, the constrained failure rates are all zeros in case of $\sigma_\phi = 0.003$ m and are all less than 15% in case of $\sigma_\phi = 0.009$ m.

The LOS accuracy with respect to different specifications is also investigated in Figure 3.14. According to the closed-form expression in Eq.(3.55), the LOS accuracy in $\mathbf{x} = (x, y, z)^T$ can reach a precision in accordance with the phase measurements and it is independent on the noise of the code measurements. This can be demonstrated from Figure 3.14a where the LOS errors in terms of different levels of code noise (from 1.5 m to 0.1 m) are the same. The LOS error is also irrelevant to the choice of ambiguity resolution methods or ambiguity accuracy since the fixed LOS vector is calculated given the fixed deterministic ambiguities. As the number of frequencies or baselines increases, a smaller LOS error can be obtained. The error in y -axis is the smallest among the three axes. This is due to the fact that the y -axis (the second column) of the baseline matrix in Table 3.7 has more diversely distributed elements and less zeros than the x - and z -axis (the first and third columns). Since the LOS can also be expressed by elevation and azimuth bearing angles in polar coordinates, the LOS error in degree is also investigated in Figure 3.14b. As can be seen, both the elevation and azimuth error are less than 0.5° in case of $\sigma_\phi = 0.009$ m and less than 0.2° in case of $\sigma_\phi = 0.003$ m.

3.4.2 Field tests

A field test is implemented for verifying the validation method and the subset ambiguity bounding method. As shown in Figure 3.16, five Novatel 702GG antennas

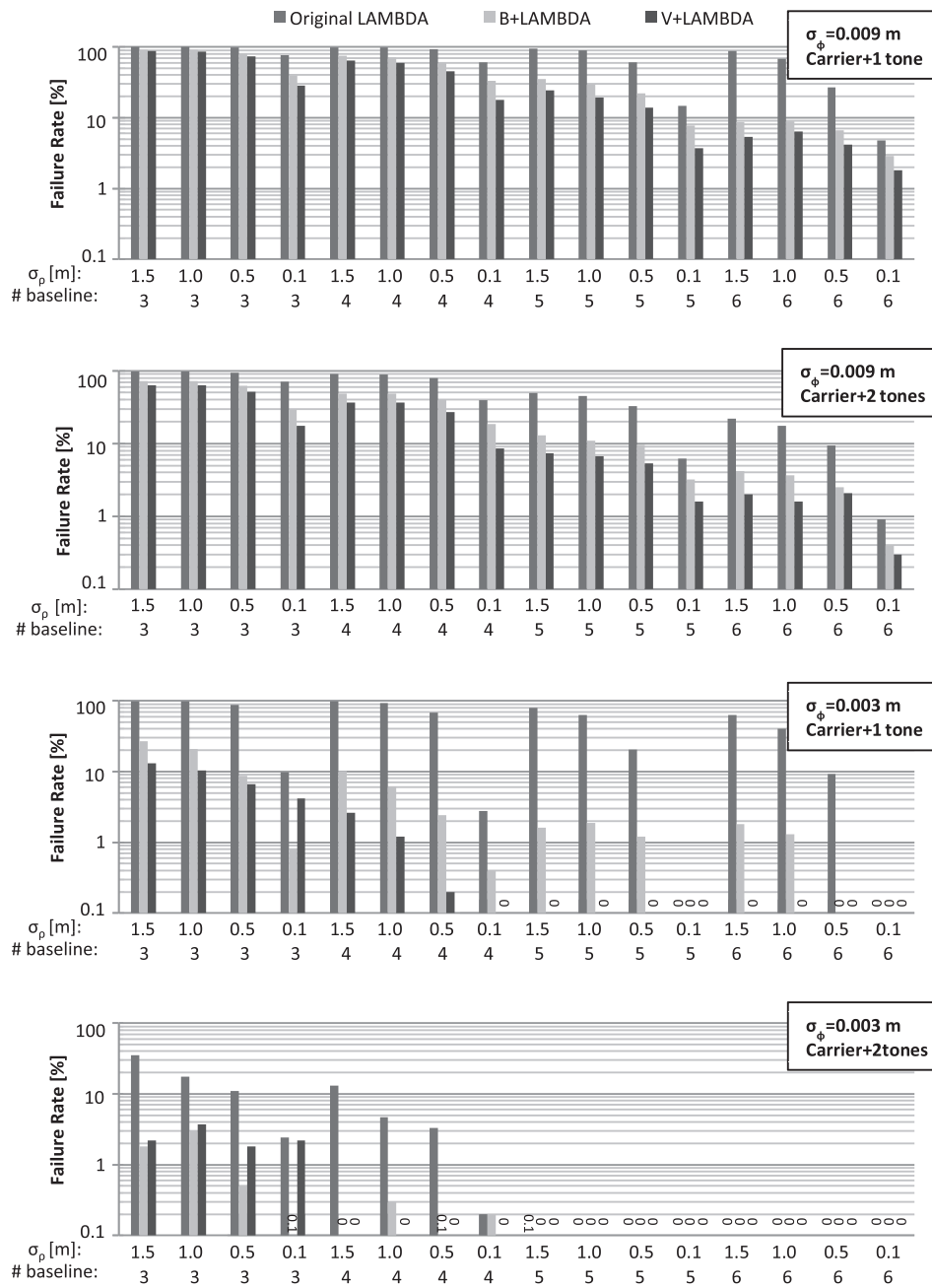
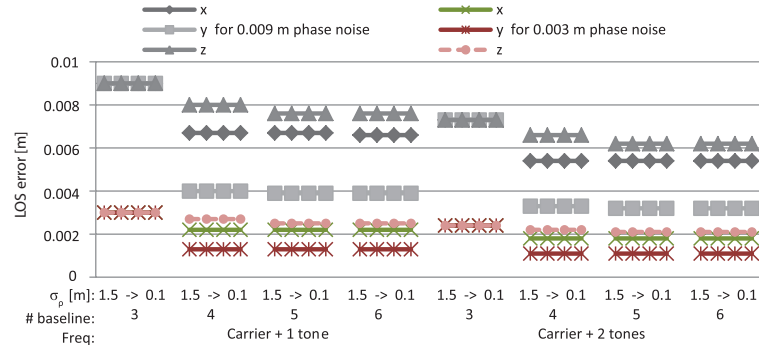
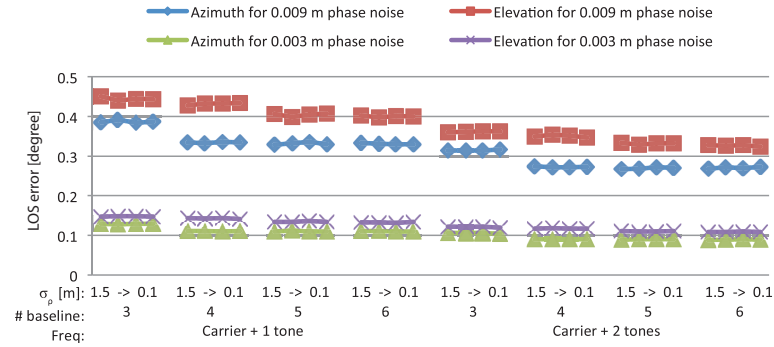


Figure 3.13: Empirical failure rate for different code and phase noise, different number of baselines and frequencies.



(a) The LOS error in meter



(b) The LOS error in degree

Figure 3.14: The LOS error in Cartesian coordinates or polar coordinates with respect to different code and phase noise, different number of baselines and frequencies.

with connections to five Novatel OEMV dual-frequency receivers are used, representing the arrangement of one spacecraft in a two-spacecraft formation, while one of the GPS satellites is treated as the other spacecraft. Although the RF-based inter-satellite ranging system in space has differences to the GPS-based geodetic application on ground, it is still acceptable to have a representative verification as the following issues can be satisfied:

- Experiments are implemented in open-sky in order to be close to the space environment;
- Static experiments can be used since the relative velocity of two spacecraft in the formation is rather low;
- Antennas are arranged close to each other in order to fit the dimension of the spacecraft;
- Only the observations from a single GPS satellite shall be used in the model in order to represent two-spacecraft formation;

- Signals coming from different GPS satellites shall be investigated separately so that the impact of the signal direction of arrivals can be evaluated;
- Driving multiple receivers with an external common oscillator shall have a similar performance as using a multi-antenna receiver on the spacecraft after the initial clock biases are calibrated out a priori.

Following these rules, the field test set-up was established on the roof of the Geomatics Engineering building at the University of Calgary, as shown in Figure 3.16. Data were collected for around 40 minutes on Sep. 21st, 2012. Antennas were arranged in a relatively small dimension, but different heights and variable locations were still assured in order to have good observability and geometry of diversity. Figure 3.15 illustrates the schematic diagram of the field test. An external oven-controlled crystal oscillator (OCXO) at 10 MHz is used. The OCXO signals are sent through a splitter to feed the primary and all other secondary receivers. The 1PPS output signal from the primary receiver has also been physically fed to other secondary receivers in order to synchronize clocks. Driving multiple receivers with an external common clock eliminates the clock drift over time, while the initial phase bias remains and can be treated as a constant bias as explained in chapter 2. Given the fact that the line bias is also constant, the combined bias will then be filtered out prior in a low pass filter, as formulated in section 3.2.2.

In this field test, the east-north-up frame is treated as the body fixed frame. Antenna 3 is assumed as the reference antenna. Baseline coordinates of other antennas with respect to antenna 3 were precisely calibrated and yielded

$$\mathbf{G} = \begin{bmatrix} \mathbf{g}_{13} \\ \mathbf{g}_{23} \\ \mathbf{g}_{43} \\ \mathbf{g}_{53} \end{bmatrix} = \begin{bmatrix} -0.709 & -1.124 & 0.361 \\ 0.394 & -1.066 & -0.358 \\ 0.992 & 0.057 & -0.137 \\ 1.147 & -1.244 & -0.168 \end{bmatrix}. \quad (3.67)$$

In space, multipath will still occur, but signals are mainly reflected from the near surfaces at the vehicle itself. The wall located on the West side of the building in Figure 3.16 can represent this circumstance, especially for antenna 2 and 5 whose heights are lower than the wall. As the satellite moves towards lower elevations, multipath on antenna 2 and 5 may thus happen. Since this field test aims to a single-epoch ambiguity resolution verification, the small multipath can be lumped into the thermal noise as the time-correlation is irrelevant in single epoch, while the large multipath may cause signal interruptions or corruptions that manifest themselves in cycle slips or losses of lock. Cycle slips and losses of lock would not affect the single-epoch resolution as in a filter. However, their occurrences will reduce the number of usable antennas at the given epoch. The field test will also demonstrate that as long as four of the antennas have uncorrupted observables, cycle slips and losses of lock can be tolerated on the other antennas. In addition, cycle slips and losses of lock may interrupt the clock bias estimation in the low pass filter. Whether the bias will change after interruptions will also be examined in the field test.

Cycle slips represent a sudden integer jump in the observations, which are detected by high-order phase differencing (Dai, 2012). Loss of lock is defined as the moment when the phase tracking loop is broken and the phase observable shows zero in the RINEX file. Taking PRN 9 for example, Table 3.8 shows the worst data quality on Antenna 2, which might be caused by multipath. The losses of lock on the L2 frequency are more frequent than on L1 as the signal strength is 3 dB lower on L2.

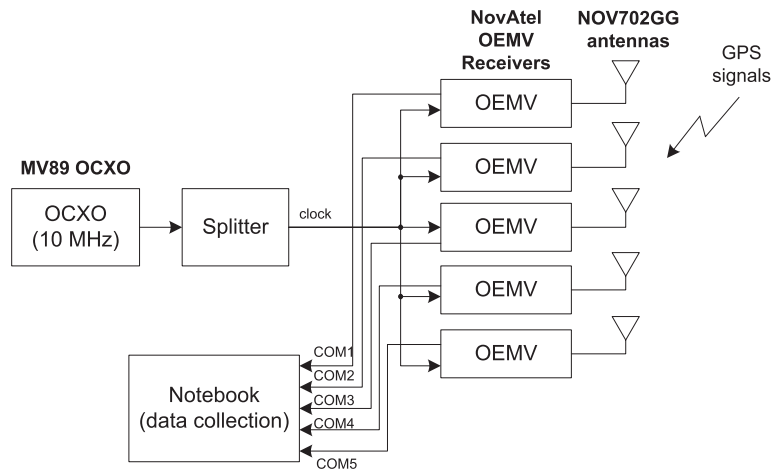


Figure 3.15: Architecture for the field test



Figure 3.16: Field test set-up

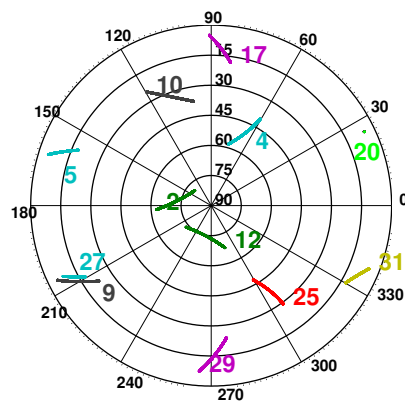


Figure 3.17: Sky plot of the field test

Table 3.8: Statistics of cycle slips and losses of lock for Satellite RPN 9 moving from 20° elevation to 4° elevation.

Antenna nr.	Total epochs	Cycle Slip		Loss of Lock	
		L1	L2	L1	L2
1	2631	0	0	0	0
2	2613	0	4	46	189
3	2584	0	1	0	0
4	2630	0	0	0	0
5	2630	0	0	0	35

Bias estimation

Constant biases remain in the SD code and phase observables, which reveals the differences in the cable lengths from antennas to receivers, as well as the differences in the initial phase biases given by different receivers. They are estimated in the low pass filter.

Taking PRN 9 for example, Figure 3.18a and 3.18b depict the code and phase bias estimation. As can be seen, all biases converge to steady states after being processed in the low pass filter. Code biases are at meter level, while phase biases have a centimeter magnitude after their integer parts are absorbed by ambiguities, leaving only the fractional parts. The theoretical maximum phase bias should be therefore no larger than half of the wavelength.

The steady state, however, could be interrupted by cycle slips or losses of lock, e.g., SD on antenna 2 & 3 (yellow) and antenna 5 & 3 (green) for L2 frequency in Figure 3.18b. Biases are re-estimated once there is a cycle slip or loss of lock in the algorithm. However, after re-estimation, the estimated phase bias still converges to the steady state which is at the same level as the previous bias before disturbance. This means cycle slips and losses of lock do not change the values of the fractional part of the initial phase biases. The consistency in Figure 3.18b is not perfect because the number of the observables between two sequential disturbances is insufficient for the low pass filter to reach the ultimate steady state. The phase bias re-estimation for the antenna 2 & 3 on L1 frequency (dark green) is an exception because the bias is close to half of the wavelength. After the integer part is absorbed by the integer ambiguity, the remaining fractional part changes its sign from the -0.5 cycle to the +0.5 cycle. The true ambiguity is accordingly changed by a cycle after re-estimation.

Once the biases are calibrated and fed back to the SD measurements, zero-mean residuals are obtained, as depicted in Figure 3.19a and 3.19b. The remaining 2-m and 2-cm fluctuations on the SD code and phase residuals are due to multipath that cannot be cancelled out by differencing. The SD noise value is amplified by a factor of $\sqrt{2}$ as compared to the noise of the undifferenced observables. Therefore, the standard derivations for undifferenced code and phase measurements are 0.8 m and 8 mm, respectively, including the effects of multipath.

Performance

After the bias removal, the constrained integer ambiguity resolutions by validation or bounding on single-epoch observations are implemented for each of the satellites separately. Results are presented in Table 3.9.

It is clear that ambiguities can be resolved with high success rate using only single

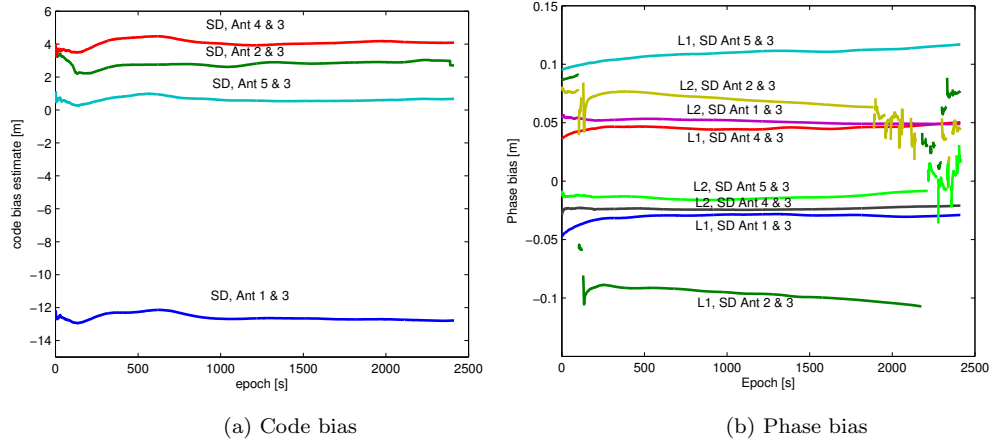


Figure 3.18: SD bias estimation for PRN9

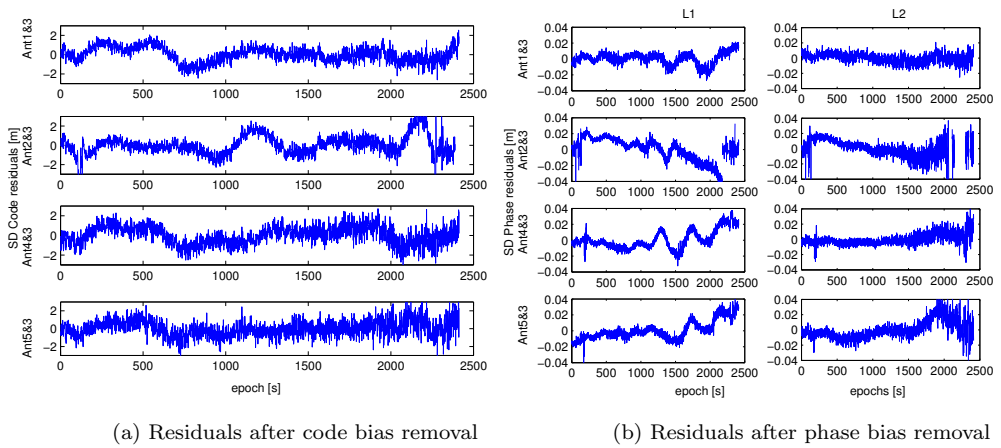


Figure 3.19: SD residuals after bias estimate feedback for PRN 9

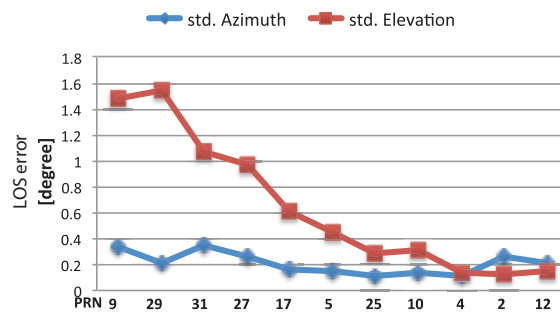


Figure 3.20: The LOS error in the field test

Table 3.9: Statistics of the integer ambiguity resolution for the field test

PRN	Epochs	General performance			Performance improvement under disturbance		
		original LAMBDA [%]	LAMBDA +V [%]	LAMBDA +B [%]	disturbance: cycle slips & losses of lock [%]	LAMBDA +V under disturbance [%]	LAMBDA +B under disturbance [%]
9	2412	79.02	98.67	97.93	11.40	10.07	9.33
29	2177	70.17	99.06	98.99	17.50	16.56	16.49
31	1629	52.54	96.62	95.70	6.02	2.64	1.72
27	1254	92.50	100.00	100.00	2.55	2.55	2.55
17	1972	48.89	95.27	94.83	0.20	0	0
5	1916	79.33	99.90	99.84	0	0	0
25	2518	91.42	99.92	99.92	0	0	0
10	2518	89.91	99.96	99.25	0	0	0
4	2518	100.00	100.00	100.00	0	0	0
2	2518	100.00	100.00	100.00	0	0	0
12	2518	100.00	100.00	100.00	0	0	0

epochs. By making use of the unit length constraint, the performance is dramatically improved by either validation or subset ambiguity bounding. Referring to the field test set-up in Figure 3.16 and the sky plot in Figure 3.17, Satellite PRN 9, 29, 31 and 27 have elevations less than 20° and face to the wall. Their measurements are vulnerable to multipath, resulting in disturbances such as cycle slips and losses of lock. The percentages of disturbances for these four satellites are 11.4%, 17.5%, 6.02% and 2.55%, respectively, out of the total epochs. The original LAMBDA method fails to resolve ambiguities when disturbances happen. However, the LAMBDA with constraint validation can improve the performance by 10.07%, 16.56%, 2.64% and 2.55%, respectively, using only the remaining uncorrupted observables. It implies that the algorithm has tolerance to signal interruption or corruption as long as four antennas are remaining with uncorrupted observables. It can also be seen that the validation method provides a slightly better performance than the subset ambiguity bounding method as expected.

For Satellite PRN 17 and 5 that face against the wall at low elevations, signal blockages may occur on antenna 2. In this circumstance, the constrained success rates have been improved by up to 45% compared to the original LAMBDA method. For Satellite PRN 25 and 10 at the elevations between 25° and 45° , neither cycle slips nor losses of lock occur. The constrained success rates are higher than 99%. For Satellite PRN 4, 2 and 12 at the elevations higher than 45° , both the original LAMBDA method and the constrained LAMBDA with validation or subset ambiguity bounding can reach a success rate of 100% in single epoch. This demonstrates that a fast and reliable estimation is easy to achieve in open sky.

Figure 3.20 shows the elevation and azimuth error for each satellite. The elevation error is around 1.2° in case of high occurrences of signal disturbances, while the error reduces to 0.1° as better quality signals are obtained. Bad signal quality also impacts the azimuth error as well, e.g., PRN 9 and 29. However, the overall azimuth error can reach below 0.4° . For PRN 2 and 12, a slightly larger azimuth error is visible. This is due to the fact that the azimuth error is highly dependent on the signal direction of arrival. A small error in the x - or y -axis of the LOS vector in the Cartesian coordinates can result in a large azimuth error in polar coordinates when the signal comes from the zenith direction.

3.5 Chapter summary

In this chapter, a line-of-sight (LOS) estimation algorithm has been developed for the radio frequency (RF)-based inter-satellite ranging system. The most challenging aspect of the LOS estimation is the carrier phase integer ambiguity resolution. Fast, reliable and robust ambiguity resolution methods have been achieved by taking advantage of a nonlinear quadratic LOS length constraint in this chapter. Two types of methods, namely, the validation method and the subset ambiguity bounding method, of applying the constraint into the LAMBDA method were proposed, so that both targets were met: minimizing the standard ambiguity objective function in a smaller and more precise search space to assure computational efficiency, and accepting the constrained conditional LOS vectors within a pre-defined threshold to improve the success rate of ambiguity resolution. The performance of the validation and the subset ambiguity bounding methods were verified by numerical simulations and field tests as well. Both of these two methods shown remarkable improvements with up to 80% higher success rates than the original LAMBDA. The validation method provided a slightly better performance than the subset ambiguity bounding method as they differ in utilizing all-ambiguity-set and subset-ambiguity, respectively. The rule-of-thumb for the pre-defined threshold has also been derived in the closed-form expression, providing an insightful guidance on how to choose boundaries according to the noise level and antenna geometry.

This chapter also demonstrated that both the ambiguity accuracy and the associated LOS accuracy explicitly depend on the number, length and relative orientations of the antenna baselines employed. In the sense of better observability and higher LOS accuracy, the baseline geometry is recommended to be with longer lengths and higher angular separations. Nonetheless, in the sense of better ambiguity resolution performance, diverse baseline lengths are more beneficial than the longer ones, and the impact of bad angular separations can be compensated dramatically by the properly determined boundary threshold.

The chapter also proposed a constrained ambiguity dilution of precision measure ($ADOP_{\delta l}$), which can serve as an indicator to characterize the expected success rate of ambiguity resolution. The $ADOP_{\delta l}$ was analytically derived, which provides an easy-to-use and insightful rule-of-thumb for the ambiguity resolution capability and allows for directly capturing the impact of various factors as well.

The proposed constrained ambiguity resolution methods in this chapter were based on single-epoch measurements. Only random noise is assumed in the model. Small multipath can be tolerated when it is lumped together with the thermal noise in a single epoch. For large multipath, multi-epoch processing will apply and acceptable success rates are expected to be achieved in only a few epochs with the help of constraint. Multipath will then be treated as a coloured noise with time correlations. The mitigation of multipath in the pseudorange and carrier phase measurements will be discussed in chapter 4 and 5, respectively. The multipath-robustness of the constrained LAMBDA method will be introduced in chapter 5.

Chapter 4

Code Multipath Effects and Mitigation Method

Multipath degrades the navigation accuracy. Moreover, in precise applications, multipath errors can dominate the total error budget. Multipath errors in pseudorange and carrier phase observations are referred to as code multipath and phase multipath, respectively. This chapter focuses on the code multipath mitigation methods, while the phase multipath reduction will be described in chapter 5. A promising method to particularly eliminate the short-delay code multipath is proposed in this chapter for space applications where multipath are mainly reflected from the dimension-limited spacecraft structures with short delays.

The chapter starts with a categorisation of existing code multipath mitigation methods in section 4.1. Several receiver-internal techniques are discussed. Among them, the a-posteriori multipath estimation (APME) method outperforms others against the multipath at short delays. In section 4.2, APME is extended by further exploring correlations between the multipath and signal strength. The multipath envelope curve fitting method is then proposed that provides the best fit to the multipath error by using the combination of signal strength estimators. Both the estimation performance and the noise induced in the estimation process are discussed. In section 4.3, the software multipath simulator and receiver are designed to demonstrate this new method.

Throughout this chapter, the BPSK-R code is used since it has been chosen as the ranging code for the radio-frequency based relative navigation in chapter 2. The general multipath-resistant capability of the BPSK-R, as opposed to the BOC code, can be found in section 2.4. Chapter 5 will introduce the carrier phase multipath mitigation, as well as the impacts of multipath on the integer ambiguity resolution.

4.1 Problem statements and existing methods

Multipath is caused by signal reflection (specular multipath) or diffraction. Reflection occurs on relatively smooth surfaces, while diffraction occurs at the edges of the obstructing object. The errors caused by multipath propagation at different antenna locations are uncorrelated, causing the differencing techniques (e.g., single-differencing or double-differencing) to be ineffective. Multipath will bias the measurements as a function of the multipath delay, phase and relative amplitude with

respect to the direct LOS signal. The general unknown number of multipath components and their path geometries, the signal structures, the reflection and diffraction effects as well as their changing nature together with the antenna and receiver design make multipath mitigation very challenging (Smyrnaiois et al., 2013).

4.1.1 Multipath in space

In space, the multipath environment is relatively “clean”. Most space vehicles, e.g., spacecraft and International Space Station (ISS), have structures such as solar panels in vicinity of antennas and thus introduce short-delay multipath. Antennas are sometimes placed on long booms in order to avoid the blockage and assure a wide field of view. In this circumstance, the surface of space vehicles will introduce multipath as well. For the RF-based inter-satellite system that transmits signals via the inter-satellite link, signal reflections are thus mainly from the vehicle itself in the aforementioned scenarios. Some reflections may occasionally come from a third close-by spacecraft in the formation, but generally within a short time duration.

Since many space applications require high positioning and navigation accuracy, carrier phase based solutions are normally used and the phase multipath effects have attracted much attention (Hwu and Loh, 1999; Axelrad et al., 1999; Reichert and Axelard, 1999; Reichert, 1999; Grelier et al., 2011). In contrast, very little literature can be found on the code multipath mitigation in space. There are two possible reasons. Firstly, the multipath caused by reflections or diffractions on space vehicles have very short delays (several meters maximum) with respect to the LOS signal. To this end, the resulted code multipath error is already small (several meters maximum). Table 2.8 has shown the multipath error level with short delays. Secondly, few space applications use sole pseudorange-based navigation. Most of the time, pseudoranges are either smoothed by carrier phases or combined with carrier phases in the process of, e.g., the integer ambiguity resolution. In these cases, a rather long time of recursive smoothing/filtering is used. The code multipath can be partially reduced in this process. It is therefore not highly necessary to have a dedicated code multipath mitigation strategy in these applications.

However, as demonstrated in Joosten and Irsigler (2003), Kubo and Yasuda (2003), Verhagen et al. (2007) and Huisman et al. (2010), code multipath effects increase the failure rate in the resolution of phase integer ambiguities. In other words, a longer time for the ambiguity resolution is required to assure a reliable solution. Thus, mitigating the code multipath becomes important when a fast (or instantaneous) ambiguity resolution is required. In addition, in the coarse-mode RF metrology where pseudorange measurements can be solely used for the inter-satellite distance estimation, the code multipath mitigation is then necessary for the accuracy improvement.

4.1.2 Multipath mitigation method categorisation

Typical code multipath mitigation methods can be categorized as follows.

Carrier smoothing. Smoothing of code observations using precise carrier phase observations is a prominent method to reduce code multipath. The basic smoothing concept was first introduced by R. Hatch and is therefore termed as Hatch-filter (Hatch, 1986). Longer smoothing periods give better performance in general. According to Irsigler (2008), the slow-varying (low frequency) multipath, which generally occurs for short reflector-antenna distances, requires a long smoothing time.

However, a long smoothing process may cause undesired errors in some cases, e.g., in varying ionospheric conditions. In addition, van Nee (1995) showed that due to the non-zero mean of the code multipath, multipath effects can not be completely eliminated by simply averaging over time, even if the averaging time is sufficiently long.

Receiver-internal multipath mitigation. This category can be further subdivided into two classes of techniques. One is to modify the code tracking loop discriminator in order to make it resistant to multipath. The earliest technique in this subclass is the narrow early-minus-late correlator (nEML) (Dierendonck et al., 1992). By reducing the spacing between early and late correlators from 1 chip to 0.1 chips, a significant reduction of multipath error can be achieved. Further improvements to the nEML correlator include the double delta ($\Delta\Delta$) correlator (Garin et al., 1996; McGraw and Braash, 1999), the N^{th} derivative correlator (Pany et al., 2004) and the optimum discriminator shaping (Pany et al., 2005). They all have proven better multipath rejection than the nEML against the multipath having a medium or large delay (> 0.1 chips), whereas they are ineffective to short-delay multipath. In fact, the short-delay multipath cannot be mitigated by any method which is compatible with discriminator modification (Pany et al., 2005). This is due to the fact that the correlation distortion is impossible to avoid when the multipath delay is short and close to the tracking point.

The second class in the receiver-internal multipath mitigation methods incorporates estimations of the multipath error or multipath parameters (delay, phase and relative amplitude). The Early/Late slope technique (ELS) (Townsend and Fenton, 1994) and the a-posteriori multipath estimation technique (APME) (Slee-waegen and Boon, 2001) are examples of the multipath error estimation. The former is used to detect the deformed slopes on both sides of the correlation peak, while the latter makes use of the in-phase relationship between the signal strength and the multipath error. The APME outperforms especially for short-delay multipath mitigation. Some examples of multipath parameter estimation include the multipath estimation delay lock loop (MEDLL) (van Nee et al., 1994; Townsend et al., 1995b,a), the multipath mitigation technique (MMT) (Weill, 2002) and the vision correlator (Fenton and Jones, 2005). The MEDLL was proposed by NovAtel Inc. based on the maximum likelihood theory. This technique requires the correlation function to be sampled by a rather large amount of correlators and is computational expensive. To limit the associated computational burden, NovAtel's first MEDLL receivers worked with 12 correlators per channel and assumed the existence of one LOS and two multipath components. In this configuration, the remaining code errors after multipath mitigation are similar to those obtained by $\Delta\Delta$ method, and it is also inefficient against short-delay multipath (Irsigler, 2008). The MMT, developed by Weill (2002), uses an optimized maximal likelihood process that assures a more efficient implementation than the MEDLL. Rather than detecting the correlation function as the MEDLL and MMT, the NovAtel's latest invention, the vision correlator, is a technique to monitor the PRN chip transitions in the time domain since it has been found that chip transitions are distorted more than the correlation function. The mathematics behind the vision correlator is similar to the MMT based on an optimized maximal likelihood theory. It is able to detect and remove the multipath at delays as short as 10 m.

Antenna (or phased array) design and location selection. Special antennas or antenna arrays have been developed to mitigate multipath. The basic principle of a multipath-resistant antenna is to increase the directivity (or gain pat-

tern) for the upper hemisphere and reduces reflections below the antenna horizon. The choke ring antenna consisting of several concentric rings is widely used to mitigate multipath. In theory, the grooves between two rings are able to cancel out the primary and secondary waves of a reflected signal if the groove's depth is equal to $1/4$ of the carrier wavelength. Dual-depth choke rings have also been developed (Filippov et al., 1999). Since the large size of choke rings, they are normally used for reference stations. Besides, recognizing the change of polarizations from the right-hand (RHCP) to left-hand circular polarization (LHCP) upon reflections, the antenna can also be designed to maximize the RHCP to LHCP gain ratio which furthermore increases multipath rejection capability. If possible, selecting a proper antenna site is also an option to avoid reflections.

A phased array is an array of antenna elements which can be manipulated to allow the main beam to be pointed towards the LOS signal and nulls to be placed on other undesired directions. This process is also called *beamforming*. In beamforming, directions of multipath components have to be found primarily by direction finding algorithms such as the MUSIC (Multiple Signal Classification) algorithm (Schmidt, 1981) and the ESPRIT (Estimation of Signal Parameters via Rotational Invariance Technique) algorithm (Roy and Kailath, 1989) before putting nulls in these directions. Apart from direction finding algorithms, another difficulty in beamforming arises from the fact that there is a high degree of correlation between the LOS and multipath components, which results in severe degradation of beamforming performance (Daneshmand, 2013).

Multipath characterization and modelling. Characterizing multipath propagation has also attracted much attention in the multipath-related studies, mainly by modelling (Lau, 2005; Irsigler, 2008; Smyrniotis et al., 2013; Weiss et al., 2007) or ray-tracing methods (Hwu and Loh, 1999; Byun et al., 2002; Ercek et al., 2006; Fan and Ding, 2006). A complete multipath characterization needs to take into account the different types of multipath (reflection vs. diffraction), the variant multipath periodic variations caused by a changing transmitter-antenna-reflector geometry, the different antenna gain patterns with respect to RHCP and LHCP polarizations, and the impact on the signal tracking process within the receiver. Multipath characterization and modelling results are mainly used for multipath prediction within a given environment. Good agreements between the modelled and real multipath have been reported in Hwu and Loh (1999), Lau (2005) and Smyrniotis et al. (2013). The potential of using multipath modelling for error corrections is mentioned in Lau (2005). However, this can only be done when all aforementioned impacting factors are carefully and precisely considered in the model. Small discrepancies, especially the discrepancy in frequency estimation, may cause serious problems in the correction process.

The chapter focuses on the short-delay code multipath in space applications. Several receiver-internal mitigation methods will be further reviewed. In order to compare their performance, a common measure, *multipath envelop*, is introduced in the following.

4.1.3 Characterizing multipath envelope

As introduced in chapter 2, in a signal tracking process, the input signal is correlated with an early and a late code replica. These are replicas with a delay of plus or minus $dT_c/2$ in comparison with the delay of a prompt code. The parameter d is referred to as the early-late spacing in chips, and T_c is the code chip duration. If the loop

is in lock, the delay of the prompt replica is the delay estimate for the input signal, which has an error $\delta\tau$ compared to the true delay. An extended description on the code delay lock loop (DLL) can be found in chapter 2, section 2.3.3.

By subtracting the early and late correlator outputs, a coherent discriminator $D(\delta\tau)$, as a function of the code error $\delta\tau$, can be yielded

$$D(\delta\tau) = A_0 \cos(\delta\phi) [R(\delta\tau + dT_c/2) - R(\delta\tau - dT_c/2)] \quad (4.1)$$

$$+ \sum_{i=1}^K A_i \cos(\psi_i - \delta\phi) [R(\delta\tau - \tau_i + dT_c/2) - R(\delta\tau - \tau_i - dT_c/2)]$$

where both the line-of-sight (LOS) signal and its reflections (multipath) from surroundings have been taken into account. $R(\tau)$ is the correlation function of the pseudo-random code. For the ideal unfiltered BPSK-R modulated code, $R(\tau)$ has a triangle shape, which is equal to $1 - |\tau|/T_c$ for $|\tau| \leq T_c$ and equal to zero elsewhere. The term $\delta\tau$ and $\delta\phi$ are code and phase tracking errors, A_0 and A_i are signal amplitudes for the LOS and multipath components, respectively, τ_i and ψ_i are the extra time delay and extra phase delay caused by the i th multipath. Note that τ_i is always positive as the multipath always arrives later than the LOS signal, and ψ_i is a function of the time delay by $\psi_i = 2\pi f\tau_i$.

In the absence of multipath, i.e., $K = 0$, the zero point of the code discriminator is at $\delta\tau = 0$, with a linear pull-in region around it where $D(\delta\tau) = -2A_0\delta\tau/T_c$ (van Nee, 1995). Any non-zero value of $D(\delta\tau)$ is caused by code thermal noise and will be captured and reported to the code NCO to speed-up or slow-down the local code replica generator in the DLL.

In the presence of multipath, the tracking loop still continues to maintain the discriminator output zero. However, the value of $\delta\tau$ that fulfils $D(\delta\tau) = 0$ is no longer zero, indicating that the loop is no more tracking the direct LOS signal, but the compound of the direct and reflected signals. To obtain additional insight, the following analysis focuses on the case of just one multipath, so $K = 1$, and the multipath envelope will be used to characterize multipath effects on the code tracking loop.

The discriminator zero point for the compound signal with one multipath is now simply that the value of the tracking error $\delta\tau$ satisfies

$$D_0(\delta\tau) + \frac{A_1 \cos(\psi_1 - \delta\phi)}{A_0 \cos(\delta\phi)} D_0(\delta\tau - \tau_1) = 0 \quad (4.2)$$

where $D_0(\delta\tau)$ is assumed as the multipath-free discriminator. The term $A_1 \cos(\psi_1 - \delta\phi)$ is referred to as the *phase-dependent multipath amplitude*, which has a plus or minus value, corresponding to positive or negative multipath signals. The multipath envelope is considered as the upper and lower bounds of the multipath errors when $\psi_1 - \delta\phi$ is equal to 0° and 180° . The term $A_0 \cos(\delta\phi)$ is referred to as the *phase error-dependent LOS amplitude*, which is always positive since the phase error $\delta\phi$ is no larger than a quarter of carrier cycle. The closed-form solution for Eq.(4.2) is

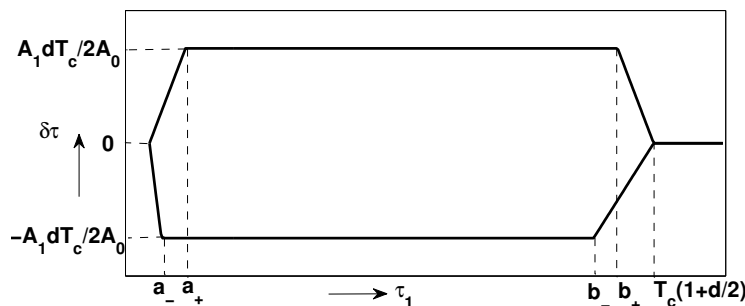


Figure 4.1: BPSK-R code multipath envelope as the function of the chip duration T_c , early-late spacing d and multipath-to-signal amplitude ratio A_1/A_0

(van Nee, 1995)

$$\delta\tau = \begin{cases} \frac{a_1}{a_0 + a_1} \tau_1 & 0 \leq \tau_1 < a \\ \frac{a_1}{2a_0} d T_c & a \leq \tau_1 \leq b \\ \frac{a_1}{2a_0 - a_1} \left[T_c \left(1 + \frac{d}{2}\right) - \tau_1 \right] & b < \tau_1 \leq T_c \left(1 + \frac{d}{2}\right) \\ 0 & \tau_1 > T_c \left(1 + \frac{d}{2}\right) \end{cases} \quad (4.3)$$

where

$$\begin{aligned} a &= \frac{a_0 + a_1}{2a_0} d T_c \\ b &= T_c \left[1 - d \left(1 - \frac{a_0 + a_1}{2a_0} \right) \right] \\ a_0 &= A_0 \cos(\delta\phi) \\ a_1 &= A_1 \cos(\psi_1 - \delta\phi) \end{aligned} \quad (4.4)$$

under the condition that $d \leq 1$ and $-a_0 \leq a_1 \leq a_0$.

The multipath error swings between its upper and lower bounds and has a sinusoidal-like shape since the phase-dependent multipath amplitude a_1 changes along with the multipath phase ψ_1 .

Disregarding the phase dependency driven by $\cos(\psi_1 - \delta\phi)$ and $\cos(\delta\phi)$, the same expression as (4.3) for the multipath envelope can be written by substituting $\delta\phi = 0^\circ$ into a_0 , and substituting $\psi_1 = 0^\circ$ and $\psi_1 = 180^\circ$, respectively, into a_1 for the upper and lower bound.

Figure 4.1 depicts the multipath envelope as a function of the multipath delay τ_1 . The values a_+ , b_+ and a_- , b_- represent the values of a and b in Eq.(4.4) for which a_0 is A_0 , and a_1 is $+A_1$ and $-A_1$, respectively. When the multipath-to-signal amplitude ratio A_1/A_0 approaches one ($A_1 \rightarrow A_0$), maximum errors of up to $dT_c/2$ can be expected. Note also that if the early-late spacing d is smaller than one chip, the multipath envelope has a hexagon shape. When d is one chip long, then $a_+ = b_+$ and $b_- = a_-$, then Figure 4.1 becomes a quadrangle.

From Eq.(4.3) and Figure 4.1, it is clear that an easy way to reduce the multipath amplitude (y -axis in Figure 4.1) is to use a narrow spacing, i.e., $d < 1$. This is called

narrow early-minus-late (nEML) correlator. However, the role of d only asserts itself once the multipath delay increases to a_+ and a_- for positive multipath and negative multipath, respectively. The short-delay multipath is irrelevant to d , and it grows linearly with the delay. The negative error grows faster than the positive error. Their slopes k_+ and k_- can be easily obtained from Eq.(4.3)

$$k_+ = \frac{1}{1 + A_1/A_0}, \quad k_- = -\frac{1}{1 - A_1/A_0}. \quad (4.5)$$

Obviously, a high A_1/A_0 results in a sharp descent on the negative error, and a relatively gradual increase on the positive error, regardless of correlator spacing.

4.1.4 Several receiver-internal multipath mitigation techniques

Double Delta Correlator

The double delta ($\Delta\Delta$) correlator improves the multipath resistant capability as opposed to the nEML correlator. The $\Delta\Delta$ is a general expression for discriminators that are formed by two pairs of correlators (two early and two late correlators).

Examples for the implementation of the $\Delta\Delta$ concept include the high resolution correlator (HRC) (Mcgraw and Braash, 1999), the Astech's strobe correlator (Garin et al., 1996) and the NovAtel's pulse aperture correlator (PAC) (Jones et al., 2004). They have similar performance, but slightly differ in implementations (Irsigler and Eissfeller, 2003).

Taking the HRC technique for example, the $\Delta\Delta$ discriminator linearly combines two early and two late correlators as follows (Mcgraw and Braash, 1999)

$$D_{\Delta\Delta} = (I_{E_1} - I_{L_1}) - \frac{1}{2}(I_{E_2} - I_{L_2}) \quad (4.6)$$

where I indicates the in-phase correlator using a cosine carrier replica for the carrier wipe-off (down-conversion from the IF signal to baseline), as opposed to the quadrature (Q) correlator where a sine carrier wipe-off is used.

The basic concept of $\Delta\Delta$ discriminator is illustrated in Figure 4.2 (a), where the spacing between E_1 and L_1 is d and the spacing between E_2 and L_2 is $2d$. In this way, the linear combination in Eq.(4.6) will cancel out the medium-delay multipath, as shown in Figure 4.2 (b). However, it is ineffective against the short-delay multipath. The short-delay multipath error stays the same as compared to the standard nEML correlator.

Optimum discriminator shaping

The above double delta concept makes use of two pairs of correlators to form a new multipath-resistant discriminator. Based on this idea, Pany et al. (2005) proposed a technique to optimize the discriminator shape by placing a bank of correlators and combining them with distinct positions and weights in the way that the multipath reduction is maximized. The optimum discriminator shape is defined by two requirements: (1) Linearity around the code tracking point; and (2) Zero value outside the linear region, see Figure 4.3.

The linear region around the tracking point determines the allowable range, where the theory of the linearized tracking loop can be applied. Within the linear region, the discriminator has a linear relation to the error, so that the discriminator

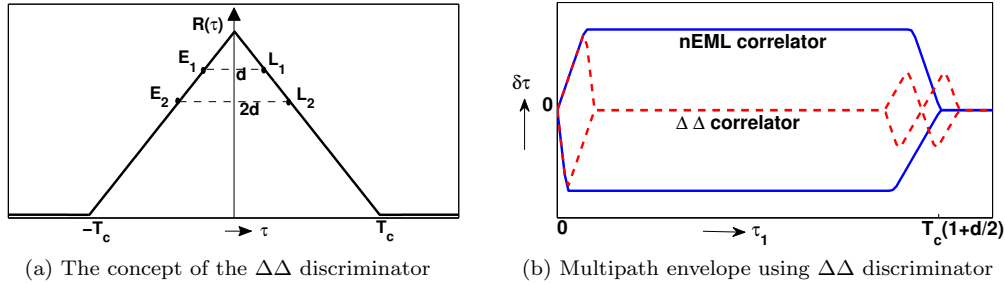


Figure 4.2: Double Delta correlator concept and performance

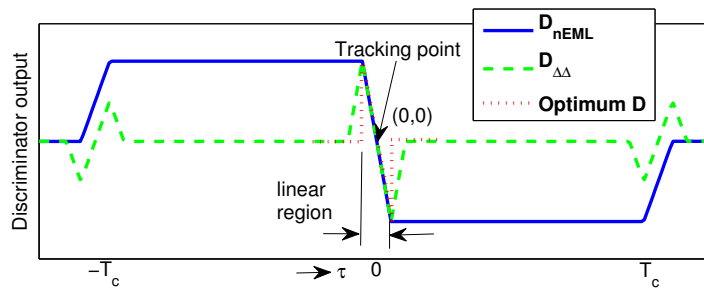


Figure 4.3: Optimum discriminator shaping

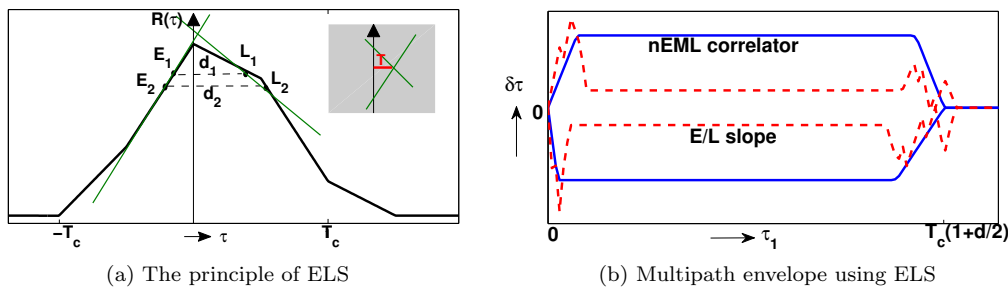


Figure 4.4: Early/late slope (ELS) concept and performance

output can be used to adjust the speed of code generator. Outside this region, if the discriminator output is zero, the best multipath resistance capability can be expected. The optimum discriminator is compared to the nEML discriminator and the $\Delta\Delta$ discriminator in Figure 4.3. As can be seen, the $\Delta\Delta$ discriminator is sub-optimum since the discriminator output for errors at around $\pm T_c$ are not zeros, resulting in a fluctuated multipath envelope when the multipath is delayed by $\pm T_c$, see Figure 4.2 (b). Therefore, the optimum discriminator shaping technique assures better multipath immunity, but at the expense of using more correlators. The smaller the linear region, the better the multipath mitigation performance. On the other hand, the linear region should be sufficiently large in order to cover the range of expected tracking errors caused by thermal noise in high dynamic environments.

From Figure 4.3, we can also see that the optimum discriminator is still inefficient against the short-delay multipath since the correlation distortion is impossible to avoid when the multipath delay is close to the tracking point. In fact, for the short-delay multipath, no matter how good is the discriminator optimization, the linear region around the origin of the discriminator still results in multipath errors. A similar conclusion was also mentioned in Pany et al. (2005) that the short-delay multipath cannot be mitigated by any method that is compatible with discriminator modification.

Early/Late slope technique

Short-delay-multipath may be partially mitigated if multipath errors or multipath parameters (amplitude, delay and phase) are estimated.

The Early/Late slope technique (ELS) (Townsend and Fenton, 1994) is an example of multipath error estimation. It has been implemented in some NovAtel receivers. The general idea is to determine the slopes of both sides of the peak in the auto-correlation function. Once both slopes are known, they can be used to compute a pseudorange correction that can be subtracted from the measured pseudorange.

The principle and performance are illustrated in Figure 4.4. The slope on each side of the auto-correlation is determined by two early (E_1, E_2) or two late (L_1, L_2) correlators with dedicated early-late spacing d_1 and d_2 . As shown in Figure 4.4 (a), since the tracking point that enables $I_{E_1} = I_{L_1}$ is no longer zero, the slopes on both sides of the distorted auto-correlation function are not equal. Making use of the distinct slopes, the pseudorange correction can be interpreted as the τ -coordinate of the intersection of two straight lines with these distinct slopes k_1 and k_2 (Townsend and Fenton, 1994)

$$T = \frac{I_{E_1} - I_{L_1} + \frac{d_1}{2}(k_1 + k_2)}{k_1 - k_2} \quad (4.7)$$

where T is the pseudorange correction, and k_1 and k_2 are equal to

$$k_1 = \frac{I_{E_1} - I_{E_2}}{0.5(d_2 - d_1)}, \quad k_2 = -\frac{I_{L_1} - I_{L_2}}{0.5(d_2 - d_1)}. \quad (4.8)$$

The multipath mitigation performance in Figure 4.4 (b) results from $d_1 = 0.1$ chips and $d_2 = 1/6$ chips. As shown, the performance for the short-delay multipath is even worse than the standard nEML discriminator with $d = 0.1$ chips.

A-Posteriori Multipath Estimation

Neither the slope-based multipath estimation nor the discriminator modification are effective for the short-delay multipath. Sleewaegen and Boon (2001) proposed a signal strength-based multipath estimation method, which is designated for mitigating the short-delay multipath. It is based on the property that the signal strength (reported as the in-phase correlator or C/N_0 in receivers) is highly correlated with the multipath error. This property is attractive for the short-delay multipath as the sensitivity of the signal strength to multipath is maximized for short delays (Sleewaegen and Boon, 2001).

This method is called a-posteriori multipath estimation (APME), as it relies on an independent a-posteriori multipath estimation without any impact to the code tracking loop.

Traditionally, the signal strength reported by a receiver is the reading of the prompt correlator output (I_P). The APME method innovatively proposed other equivalent signal strength expressions on the basis of early or late correlators I_{E_i} or I_{L_i} (i being any positive integer at a delay of $id/2$ from the prompt correlator)

$$S_P = I_P, \quad S_{E_i} = \frac{I_{E_i}}{1 - id/2}, \quad S_{L_i} = \frac{I_{L_i}}{1 - id/2} \quad (4.9)$$

where S_P , S_{E_i} and S_{L_i} denote the signal strength computed from the prompt, early and late correlators, respectively. The scaling factor $1/(1 - id/2)$ has been applied to I_{E_i} and I_{L_i} to compensate for the triangular shape of the correlation peak. All of these estimators for the signal strength yield the same result in case no multipath is present. When multipath enters the receiver, different but highly-correlated signal strengths can be obtained.

The APME method demonstrated that the multipath estimation $\delta\hat{\tau}$ (in chips) using S_P and S_{L_2} has a good agreement to the actual multipath error, especially at short delays (Sleewaegen and Boon, 2001), in chips,

$$\begin{aligned} \delta\hat{\tau} &= 0.42 \frac{S_{L_2} - S_P}{S_P} \\ &= -0.42 \left(1 - \frac{I_{L_2}}{I_P} \frac{1}{1 - d} \right) \end{aligned} \quad (4.10)$$

where 0.42 was an empirical coefficient, and the reason of using S_P and S_{L_2} rather than other correlators was not mentioned in Sleewaegen and Boon (2001).

In the following, this chapter proposes improvements to APME by using more correlators than I_{L_2} and I_P with the attempt of providing a best possible resemblance between the estimation and actual error in the least-squares sense. The amount, the weights and the exact locations of correlators will be discussed. A simple implementation strategy will also be proposed for reducing the computational load of using multiple correlators.

Several other techniques such as the multipath estimation delay lock loop (MEDLL) (van Nee et al., 1994), the multipath mitigation technique (MMT) (Weill, 2002) and the vision correlator (Fention and Jones, 2005) for the estimation of multipath parameters are all generally based on the maximum likelihood theory and computationally intensive. They will not be further discussed in this chapter.

4.2 Theory of the signal strength-based multipath envelope curve fitting

4.2.1 Characterizing the relation between the multipath error and the signal strength

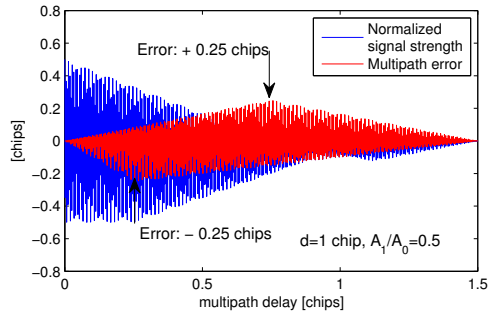
Recalling that the parameters used to describe a multipath include the multipath amplitude A_1 , delay τ_1 and phase ψ_1 with respect to the LOS signal amplitude A_0 , delay τ_0 and phase ψ_0 , then the compound signal strength reported by the prompt correlator S_P changes when any of the above parameters changes.

In order to explore the correlation between the signal strength and the multipath error, Figure 4.5 depicts the normalized signal strength S_P/A_0 and the multipath error $\Delta\tau$ as the multipath delay τ_1 varies from 0 to more than 1 chip and the multipath phase varies according to $\psi_1 = 2\pi f\tau_1$. Instead of looking at only the envelope (upper and lower bounds), the swings between the bounds for both S_P/A_0 and $\Delta\tau$ are also depicted, where we can find that they are both phase-dependent and periodic. Here, the normalized signal strength has been shifted by 1, which is $S_P/A_0 - 1$, in order to eliminate the LOS signal strength and emphasize only on the multipath contributions. Different correlator spacings and different multipath-to-signal amplitude ratios are examined.

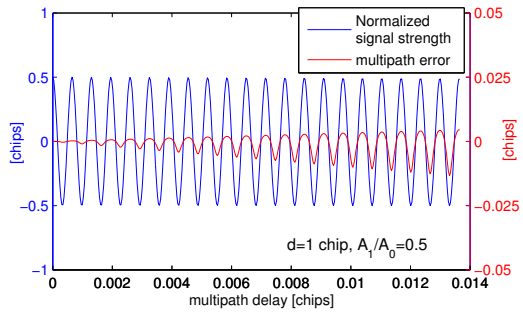
To have a better exploration specifically to the short-delay multipath, illustrations for the multipath delayed by less than 0.014 chips (equivalent to 4 m for the BPSK-R(1) code and 0.4 m for the BPSK-R(10) code) are shown in Figure 4.6.

By analyzing Figure 4.5 and 4.6, following characteristics can be obtained:

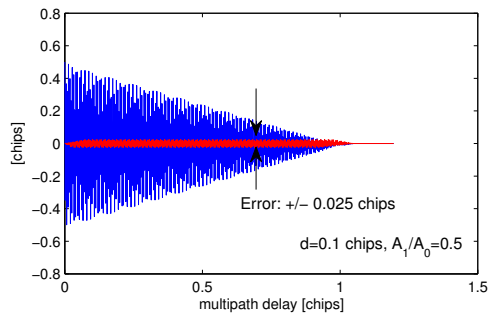
- Comparing Figure 4.5 (a) and (b), it is found that the narrow correlator spacing, $d=0.1$ chips (in (b)), provides a much better multipath resistance than $d=1$ chip (in (a)). However, for the short-delay multipath, as shown in Figure 4.6 (a) (b), the multipath errors for $d=0.1$ chips and $d=1$ chip are exactly the same. This demonstrates the aforementioned statement in section 4.1.3 that a narrow spacing cannot help any more when it comes to short delays.
- By zooming in the swings between the upper and lower bounds, as depicted in Figure 4.6 (a-c), it can be seen that both the multipath error and the signal strength vary sinusoidally as a function of multipath delay, and they have an *in-phase correlation*. This means that one could build up a good multipath estimator by properly scaling the signal strength.
- The in-phase correlation has not been much exploited until the APME method, mainly because the proportionality factor that links the signal strength and the multipath error generally changes along with the change of the multipath delay. As shown in Figure 4.5, when the multipath delay increases from 0 to 1.5 chips, the overall trends for the multipath error and signal strength are different. The envelope of the multipath error is increasing, keeping constant and dropping afterwards, while the signal strength envelope always declines. Therefore, scaling a stand-alone signal strength reported from only the prompt correlator is impossible to adaptively estimate the multipath error with any delay. Other early or late correlators have to be used in order to introduce different scaling factors. By weighting and combining multiple scaled correlators, multipath at different delays are then possible to be estimated. The APME method makes use of two correlators. Better performance is expected once more correlators are involved.
- From Figure 4.6, for the short-delay multipath, the error is not perfectly sinusoidal and in fact contains sharp discontinuities because of the asymmetrical property for positive and negative components. According to Eq.(4.5), the envelopes of positive and negative errors at short delays have distinct slopes of $1/(1 + A_1/A_0)$ and $-1/(1 - A_1/A_0)$, respectively. Obviously, the higher is A_1/A_0 , the larger is the asymmetry and the harder is the multipath estimation. This will also result in a non-zero mean of multipath, implying the fact that multipath can not be completely eliminated by simply averaging/smoothing over time.



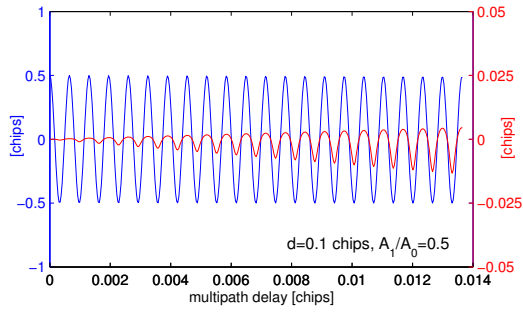
(a)



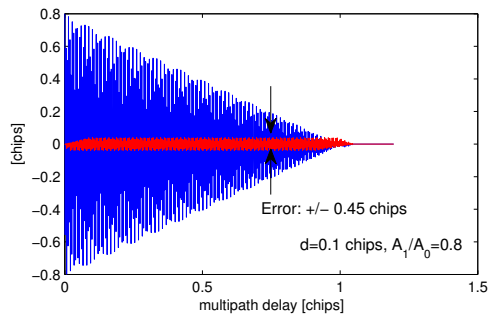
(a)



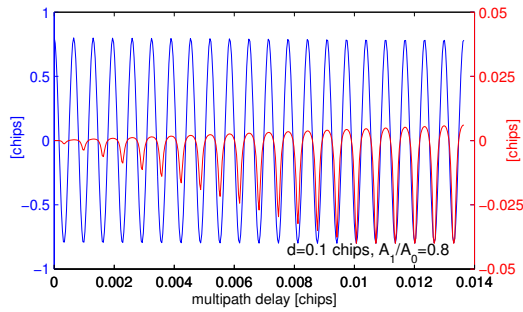
(b)



(b)



(c)



(c)

Figure 4.5: Normalized and shifted signal strength $S_0/A_0 - 1$ and the multipath error $\delta\tau$ as the multipath delay ranges from 0 to 1.5 chips.

Figure 4.6: Detailed view of the normalized and shifted signal strength $S_0/A_0 - 1$ and the multipath error $\delta\tau$ as the multipath delay ranges from 0 to 0.014 chips.

4.2.2 Principle of the multipath envelope curve fitting

Taking advantage of the nature that the code multipath error has an in-phase correlation with the signal strength, scaling and combining signal strengths reported from multiple correlators is reasonable to yield a good estimation of multipath that will then act as a pseudorange correction.

For the sake of simplicity, the early/late correlators used in the following derivations are redefined as I_i , where i is a negative integer for the early correlator, while it is a positive integer for the late correlator. The prompt correlator is then expressed as I_0 . The spacing between adjacent correlators is defined as Δd (in chips), which is not necessary to be equal to $d/2$ where d is regarded as the DLL early-late spacing. The signal strength reported from the i th correlator I_i can be expressed as

$$S_i = \frac{\gamma_i I_i}{1 - |i|\Delta d} \quad (4.11)$$

where an appropriate scaling factor of $\gamma_i/(1 - |i|\Delta d)$ has been applied to early/late correlators I_i . The factor $1/(1 - |i|\Delta d)$ is used to compensate for the triangular shape of the BPSK-R code correlation peak, and the factor γ_i (close to 1) is used to account for the rounding of the correlation peak due to the limited signal bandwidth. A term $d_i = |i|\Delta d$ is defined as the spacing of I_i away from the prompt correlator I_0 .

The signal strength estimators from different correlators yield distinct signal strength values in the presence of multipath. They all show in-phase correlations with the multipath error, and can be weighted and combined for multipath estimation.

For a real implementation in the receiver, the absolute value of the signal strength is given typically in units of Watt or represented by digits (when signal is processed after ADC in a software-defined receiver), which should be translated to non-unit values by normalization. Then, S_i is normalized by A_0 or S_0 . Taking the APME method for example, Figure 4.7 shows the normalized signal strength S_0/A_0 , S_{+2}/A_0 and S_0/S_0 , S_{+2}/S_0 , respectively, which are computed from a prompt I_0 and a late correlator I_{+2} as a function of the multipath delay.

From Figure 4.7 (a), it appears that the difference between S_0/A_0 and S_{+2}/A_0 resembles the multipath error envelope of a nEML discriminator. More specifically, they are zeros for multipath delay $\tau_1 = 0$ and $\tau_1 > 1.2$ chips and do not vary much in the range of $0.1 < \tau_1 < 1$ chips. If the signal strength is normalized by S_0 , as depicted in Figure 4.7 (b), this resemblance is more clear. Only envelopes are illustrated. Since the true phase-dependent multipath error and signal strength have an in-phase correlation, their sinusoidal-like swings between the lower and upper bounds can be synchronized once the envelope is very well resembled.

The APME method claims that both the quantity $0.42(S_{+2} - S_0)/A_0$ and $0.42(S_{+2} - S_0)/S_0$ have a good agreement to the multipath error, as depicted in Figure 4.8. However, the applicability of using $0.42(S_{+2} - S_0)/A_0$ in a real-life situation is limited because A_0 is the LOS signal amplitude which is unknown at the receiver, while S_0 reported by the prompt correlator for the compound signal is much easier to apply. From Figure 4.8, the resemblance between the multipath error and its estimate thereof using $0.42(S_{+2} - S_0)/S_0$ is not very good for the positive short-delay multipath, and it is also overestimated for the negative multipath.

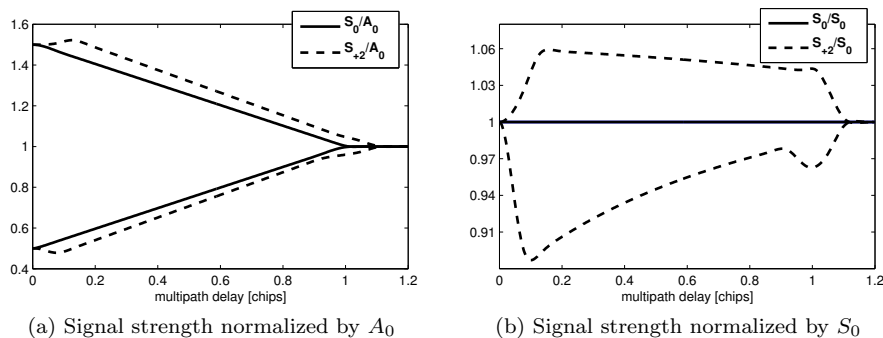


Figure 4.7: Normalized signal strength computed from I_0 and I_{+2} . The spacing between adjacent correlators is $d/2$ with $d = 0.1$ chips as the DLL early-late spacing. The multipath-to-signal amplitude ratio is $A_1/A_0 = 0.5$, and the front-end bandwidth is 24 MHz for BPSK-R(1) code.

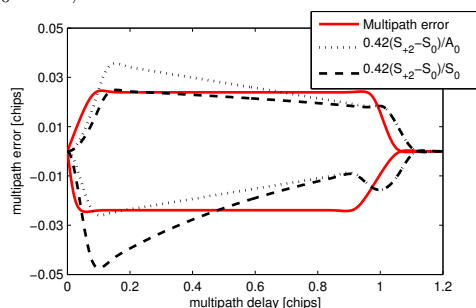


Figure 4.8: Agreement between the multipath error envelope and its estimation using S_0 and S_{+2} , $\Delta d = 1/2d$

Improvements

To improve the model for multipath estimation, two possibilities will be discussed in this section:

(1) Consider more signal strength estimators than S_0 and S_{+2} to enable a best possible resemblance between the true multipath and its estimate in the least-squares sense.

The multipath estimation is now obtained by linearly combining M early, one prompt and N late signal strength estimators

$$\delta\hat{\tau}(\tau_1) = \sum_{i=-M}^N \left(w_i \frac{S_i(\tau_1)}{S_0(\tau_1)} \right) \quad (4.12)$$

where $\delta\hat{\tau}(\tau_1)$ is the estimated multipath as the function of the multipath delay τ_1 , w_i is the weight for the i th normalized signal strength estimator S_i/S_0 , and $S_i = \gamma_i I_i / (1 - |i|\Delta d)$ has a spacing of $|i|\Delta d$ away from I_0 .

The best possible resemblance is the one that enables a least squared sum of residuals between the true multipath $\delta\tau(\tau_1)$ and its estimate $\delta\hat{\tau}(\tau_1)$ for a given set

of multipath delays $\tau_1 = l_1, \dots, l_m$. The cost function J is then written as

$$\begin{aligned} J &= \min \sum_{\tau_1=l_1}^{l_m} (\delta\tau(\tau_1) - \delta\hat{\tau}(\tau_1))^2 \\ &= \min \sum_{\tau_1=l_1}^{l_m} \left(\delta\tau(\tau_1) - \left(\sum_{i=-M}^N w_i \frac{S_i(\tau_1)}{S_0(\tau_1)} \right) \right)^2. \end{aligned} \quad (4.13)$$

This process is named here as *multipath envelope curve fitting*. The curve of the true multipath envelope in the nEML discriminator has a hexagon shape, which is expected to be well fitted by linearly combining multiple signal strength estimators. The amount and exact spacings between the correlators will be discussed in section 4.2.4.

(2) Curve-fit the multipath envelope of the $\Delta\Delta$ discriminator rather than the nEML discriminator. Compared to the hexagon multipath curve in the nEML discriminator, the medium delayed multipath in the $\Delta\Delta$ discriminator has already been largely eliminated, leading to a different target error curve $\delta\tau(\tau_1)$ in the objective function of (4.13), see Figure 4.2 in section 4.1.4. Therefore, curve fitting the $\Delta\Delta$ multipath envelope could assure a better multipath reduction.

Formulating the resemblance between the true multipath error and its estimate as $\mathbf{z} = \mathbf{H}\mathbf{w} + \mathbf{e}$

$$\begin{bmatrix} \delta\tau(l_1) \\ \delta\tau(l_2) \\ \vdots \\ \delta\tau(l_m) \end{bmatrix} = \begin{bmatrix} \frac{S_{-M}(l_1)}{S_0(l_1)} & \frac{S_{-M+1}(l_1)}{S_0(l_1)} & \dots & \frac{S_{+N}(l_1)}{S_0(l_1)} \\ \frac{S_{-M}(l_2)}{S_0(l_2)} & \frac{S_{-M+1}(l_2)}{S_0(l_2)} & \dots & \frac{S_{+N}(l_2)}{S_0(l_2)} \\ \dots & \dots & \dots & \dots \\ \frac{S_{-M}(l_m)}{S_0(l_m)} & \frac{S_{-M+1}(l_m)}{S_0(l_m)} & \dots & \frac{S_{+N}(l_m)}{S_0(l_m)} \end{bmatrix} \begin{bmatrix} w_{-M} \\ w_{-M+1} \\ \vdots \\ w_{+N} \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_m \end{bmatrix} \quad (4.14)$$

the problem now is defined as finding the weights \mathbf{w} for each signal strength estimator that minimize the squared sum of residuals $\sum_{j=1}^{j=m} e_j^2 = \sum_{\tau_1=l_1}^{\tau_1=l_m} (\delta\hat{\tau}(\tau_1) - \delta\tau(\tau_1))^2$ in the least-squares sense.

The cost function of (4.13) can then be rewritten as

$$J = \mathbf{e}^T \mathbf{e} = (\mathbf{z} - \mathbf{H}\mathbf{w})^T (\mathbf{z} - \mathbf{H}\mathbf{w}). \quad (4.15)$$

According to the least-squares adjustment, the cost function J reaches a minimum if the derivatives of J with respect to \mathbf{w} equal to zero

$$\frac{\partial J}{\partial \mathbf{w}} = 0. \quad (4.16)$$

Then, the weights can be obtained as

$$\mathbf{w} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{z}. \quad (4.17)$$

In case that no multipath exists, there is, of course, no multipath error, thus $\delta\hat{\tau} = \delta\tau = 0$, and the signal strengths reported by different correlators are equal,

$S_i = S_0$. Substituting these conditions into Eq.(4.12), it can be found that the sum of the weights yields zero as expected. This a-priori information can be considered as a linear constraint to the least-squares process

$$\mathbf{C}\mathbf{w} = 0 \quad (4.18)$$

where

$$\mathbf{C} = [1 \quad 1 \quad \cdots \quad 1 \quad 1] \quad (4.19)$$

\mathbf{C} has the size of $1 \times (M + N + 1)$.

To solve the linear constrained least-squares adjustment, the cost function is reformulated in (Xu, 2007)

$$F = (\mathbf{z} - \mathbf{H}\mathbf{w})^T(\mathbf{z} - \mathbf{H}\mathbf{w}) + 2\mathbf{K}^T(\mathbf{C}\mathbf{w}) \quad (4.20)$$

where \mathbf{K} is a gain vector to be determined when F reaches its minimum by differentiating F with respect to \mathbf{w}

$$\frac{\partial F}{\partial \mathbf{w}} = 0. \quad (4.21)$$

For simplification, let $\mathbf{Q} = (\mathbf{H}^T\mathbf{H})^{-1}$. One then has

$$\mathbf{K} = (\mathbf{C}\mathbf{Q}\mathbf{C}^T)^{-1}(\mathbf{C}\mathbf{Q}\mathbf{H}^T\mathbf{z}) \quad (4.22)$$

$$\mathbf{w} = \mathbf{Q}\mathbf{H}^T\mathbf{z} - \mathbf{Q}\mathbf{C}^T\mathbf{K}. \quad (4.23)$$

To sum up, the principle of the multipath envelope curve fitting method is to estimate the multipath error using the linear combination of multiple weighted signal strength estimators in order to best fit the true error in the least squares sense.

4.2.3 Variance

The quality of the multipath estimation should not be only investigated from an accuracy point of view. It is also important to evaluate the noise on the multipath estimation since this noise will be added to the measurement noise when subtracting the multipath estimation from the measurement. To this end, the variance of the multipath estimation has to be derived.

The variance approximation using the first order Taylor expansion for an equation $f(X_1, X_2, \cdots, X_p)$ with variables X_1, X_2, \cdots, X_p will be used in the derivation (Seltman, 2012)

$$\begin{aligned} \text{var}(f(X_1, X_2, \cdots, X_p)) &\approx \sum_{i=1}^p f'_{X_i}(\Theta)^2 \text{var}(X_i) \\ &+ 2 \sum_{1 \leq i < j \leq p} f'_{X_i}(\Theta) f'_{X_j}(\Theta) \text{cov}(X_i, X_j). \end{aligned} \quad (4.24)$$

Let $\Theta = (E(X_1), E(X_2), \cdots, E(X_p))$, where $E()$ is the expectation operator for variables.

The multipath estimation $\delta\hat{\tau}$ can be written as a function of multiple correlator outputs $I_{-M}, \dots, I_0, I_1, \dots, I_N$ by substituting Eq.(4.11) into (4.12). Here, an infinite bandwidth ($\gamma_i = 1$) is assumed for simplicity, as, in fact, γ_i is very close to 1 in wide bandwidth and will not impact the variance significantly

$$\begin{aligned}\delta\hat{\tau} &= \sum_{i=-M}^N w_i \frac{S_i}{S_0} \\ &= \sum_{i=-M}^N \left(\frac{w_i}{1 - |i|\Delta d} \right) \frac{I_i}{I_0} \\ &= \sum_{-M \leq i \leq N, i \neq 0} \alpha_i \frac{I_i}{I_0} + w_0\end{aligned}\quad (4.25)$$

where $\alpha_i = w_i/(1 - |i|\Delta d)$ is redefined as the coefficient for correlators. The weight for the prompt correlator w_0 is a deterministic value after estimation, which does not contribute to the variance estimation. Therefore, the variance of $\delta\hat{\tau}$ is equal to the variance of the following function with $I_{-M}, \dots, I_0, I_1, \dots, I_N$ as variables

$$f(I_{-M}, \dots, I_0, I_1, \dots, I_N) = \sum_{i \neq 0} \frac{\alpha_i I_i}{I_0} \quad (4.26)$$

where the first order partial derivatives with respect to each variable is

$$f'_{I_i(i \neq 0)}(\Theta) = \frac{\alpha_i}{E(I_0)}, \quad f'_{I_0}(\Theta) = -\frac{\sum_{i \neq 0} \alpha_i E(I_i)}{E^2(I_0)} \quad (4.27)$$

with Θ now turning to $\Theta = (E(I_{-M}), \dots, E(I_0), E(I_1), \dots, E(I_N))$. By first order Taylor expansion of variance of Eq.(4.26), we then get $\sigma_{\delta\hat{\tau}}^2$ as

$$\begin{aligned}\sigma_{\delta\hat{\tau}}^2 &= \text{var}(f(I_{-M}, \dots, I_0, I_1, \dots, I_N)) \\ &\approx \sum_{i \neq 0} f'_{I_i}(\Theta)^2 \text{var}(I_i) + f'_{I_0}(\Theta)^2 \text{var}(I_0) \\ &\quad + 2 \sum_{i < j, i \neq 0, j \neq 0} f'_{I_i}(\Theta) f'_{I_j}(\Theta) \text{cov}(I_i, I_j) + 2 \sum_{i \neq 0} f'_{I_i}(\Theta) f'_{I_0}(\Theta) \text{cov}(I_i, I_0) \\ &= \sum_{i \neq 0} \left(\frac{\alpha_i}{E(I_0)} \right)^2 \text{var}(I_i) + \left(\frac{\sum_{i \neq 0} \alpha_i E(I_i)}{E^2(I_0)} \right)^2 \text{var}(I_0) \\ &\quad + 2 \sum_{i < j, i \neq 0, j \neq 0} \frac{\alpha_i \alpha_j}{E^2(I_0)} \text{cov}(I_i, I_j) - 2 \sum_{i \neq 0} \frac{\alpha_i}{E(I_0)} \frac{\sum_{i \neq 0} \alpha_i E(I_i)}{E^2(I_0)} \text{cov}(I_i, I_0)\end{aligned}\quad (4.28)$$

where $\text{var}(I_i)$ is the variance of I_i , and $\text{cov}(I_i, I_j)$ is the covariance between I_i and I_j .

Assuming all correlator outputs $I_{-M}, \dots, I_0, I_1, \dots, I_N$ are Gaussian random variables, their mean are related to the signal-to-noise ratio C/N_0 , the integration time T in the correlation, and the spacing $i\Delta d$ apart from the central prompt cor-

relator (Dierendonck et al., 1992)

$$\begin{bmatrix} E(I_{-M}) \\ E(I_{-M+1}) \\ \vdots \\ E(I_0) \\ E(I_1) \\ \vdots \\ E(I_N) \end{bmatrix} = \sqrt{2\frac{C}{N_0}T} \begin{bmatrix} R(-M\Delta d) \\ R((-M+1)\Delta d) \\ \vdots \\ 1 \\ R(\Delta d) \\ \vdots \\ R(N\Delta d) \end{bmatrix} \quad (4.29)$$

where $R(\cdot)$ is the normalized correlation function, $R(\tau) = 1 - |\tau|$, $|\tau| \leq 1$ (in chips), and $R(\tau) = R(-\tau)$.

The noise on correlators is white with respect to time, while the correctors with different delays $I_{-M}, \dots, I_0, I_1, \dots, I_N$ along the delay dimension are highly correlated, resulting in non-zero values in their covariance. The variance-covariance for $I_{-M}, \dots, I_0, I_1, \dots, I_N$ are expressed in the following correlation coefficient matrix $\mathbf{\Omega}$ (derivations can be found in Appendix C)

$$\begin{aligned} \mathbf{\Omega} &= \begin{bmatrix} \text{var}(I_{-M}) & \text{cov}(I_{-M}, I_{-M+1}) & \cdots & \text{cov}(I_{-M}, I_N) \\ \text{cov}(I_{-M+1}, I_{-M}) & \text{var}(I_{-M+1}) & \cdots & \text{cov}(I_{-M+1}, I_N) \\ \text{cov}(I_{-M+2}, I_{-M}) & \text{cov}(I_{-M+2}, I_{-M+1}) & \cdots & \text{cov}(I_{-M+2}, I_N) \\ & \cdots & & \\ \text{cov}(I_N, I_{-M}) & \text{cov}(I_N, I_{-M+1}) & \cdots & \text{var}(I_N) \end{bmatrix} \\ &= \begin{bmatrix} 1 & R(-\Delta d) & \cdots & R(-(M+N)\Delta d) \\ R(\Delta d) & 1 & \cdots & R(-(M+N-1)\Delta d) \\ R(2\Delta d) & R(\Delta d) & \cdots & R(-(M+N-2)\Delta d) \\ & \cdots & & \\ R((M+N)\Delta d) & R((M+N-1)\Delta d) & \cdots & 1 \end{bmatrix}. \end{aligned} \quad (4.30)$$

Substituting the mean, variance and covariance expressions in Eq.(4.29) and (4.30) into (4.28), the variance of the noise on the multipath estimation is (expressed in chips²)

$$\begin{aligned} \sigma_{\delta\hat{\tau}}^2 &= \frac{1}{2TC/N_0} \sum_{i \neq 0} \frac{1}{(1 - |i|\Delta d)^2} w_i^2 + \frac{1}{2TC/N_0} \left(\sum_{i \neq 0} w_i \right)^2 \\ &+ \frac{1}{TC/N_0} \sum_{i < j, i \neq 0, j \neq 0} \frac{1 - |i-j|\Delta d}{(1 - |i|\Delta d)(1 - |j|\Delta d)} w_i w_j \\ &- \frac{1}{TC/N_0} \sum_{i \neq 0} \left(\left(\sum_{i \neq 0} w_i \right) w_i \right). \end{aligned} \quad (4.31)$$

This is the variance before the low pass filter in the code tracking loop. After the low pass filter with an equivalent noise bandwidth of B_L , the term $1/(2TC/N_0)$ shall be substituted to $B_L/(2C/N_0)$. Then, the variance of the multipath estimation noise $\sigma_{\delta\hat{\tau}}^2$ can be compared with the code thermal noise σ_{nEML}^2 or $\sigma_{\Delta\Delta}^2$ using either the nEML or $\Delta\Delta$ discriminator, which also depend on B_L , C/N_0 and d (expressed

in chips²) (Borio, 2012)

$$\sigma_{\text{nEML}}^2 = \frac{B_L}{2C/N_0} d \quad (4.32)$$

$$\sigma_{\Delta\Delta}^2 = \frac{B_L}{2C/N_0} \frac{d_1 d_2}{d_2 - d_1}. \quad (4.33)$$

Note that Eq.(4.31) to (4.33) are simplified as they assume an infinite front-end bandwidth. A complete expression of the DLL thermal noise can be found in section 2.3.3.

The multipath estimation is meaningful only if it does not introduce too much noise. This means the value of $\sigma_{\delta\hat{\tau}}^2$ should be smaller or comparable with the value of σ_{nEML}^2 or $\sigma_{\Delta\Delta}^2$. Such noise, in addition, needs more attention in space applications where the short-delay multipath dominates the contaminated signal. The magnitude of the short-delay multipath error could be as small as the multipath estimation noise. In this case, the multipath estimation is not longer effective if the error introduced by estimation is in the same magnitude level as the error without estimation.

4.2.4 Discussions on the amount and locations of correlators

A compromise between the curve fitting performance and the estimation noise should be made that ensures a precise curve fitting and would not introduce too much noise. Both the estimation performance and noise are dependent on the amount and the exact locations of correlators. Then, two main questions arise: how many correlators would be? And how small the spacing between adjacent correlators should be?

To investigate and answer these questions, Figure 4.9 illustrates the normalized signal strength S_i/S_0 based on largely spaced correlators $|i|\Delta d \geq 1/2d$ (in (a)(b) when $\Delta d = 1/2d$) and small spaced correlators $|i|\Delta d < 1/2d$ (in (c)(d) when $\Delta d = 1/8d$), respectively. The nEML discriminator with the early-late spacing of $d = 0.1$ chips is used in this example.

By analysing Figure 4.9 (a-d), the following characteristics can be obtained:

- Since the nEML discriminator is used, the multipath envelope has a hexagon shape. However, only Figure (b) and (c) for late correlators with large spacings and early correlators with small spacings show resemblance to the hexagon shape. This means they are good candidates for curve fitting. The correlators in Figure (a) and (d) could instead be used to balance the shape change as the multipath delay increases, and it is not necessary to use many of them.
- Multipath with short delays are difficult to be detected. If the spacing between adjacent correlators is $1/2d$, as in cases of Figure (a) and (b), S_i/S_0 from different early/late correlators, i.e., ($i \neq 0$), are not distinguishable when the multipath delay is smaller than 0.02 chips. It is then useless to combine more correlators than I_0 and I_{+2} since other correlators do not provide new information about the distortion. On the contrary, the short-delay multipath can be better detected by smaller spaced correlators, as in cases of Figure (c) and (d). A good curve fitting performance for the short-delay multipath can then be expected by using smaller spaced correlators. Several largely spaced correlators can be used to balance the curve fitting performance for medium/large delayed multipath, and the number of them shall be small.

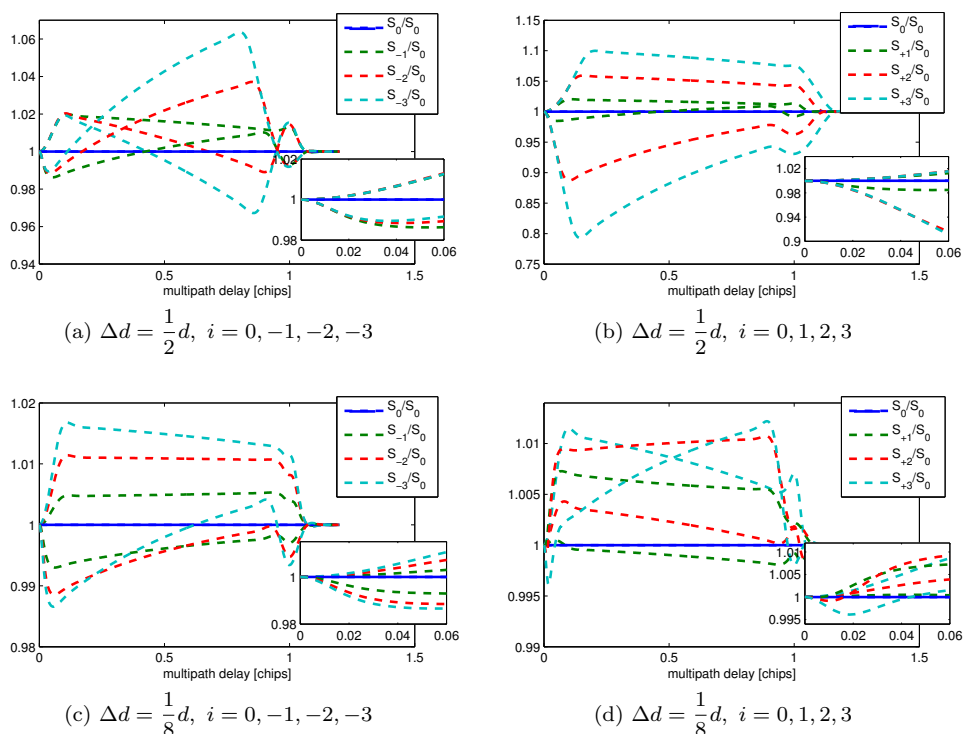


Figure 4.9: Normalized signal strength S_i/S_0 calculated from multiple correlators with spacings of $1/2d$ (a)(b) and $1/8d$ (c)(d), respectively, between adjacent correlators. S_0/S_0 is in solid black, $S_i/S_0 (i \neq 0)$ is in coloured dash, $d = 0.1$ chips and the nEML discriminator is used.

- By looking at the absolute values (y-axis) of S_i/S_0 , it can be found that S_i/S_0 with large spacings (in Figure (a)(b)) have larger absolute values than S_i/S_0 with small spacings (in Figure (c)(d)). This indicates that more weights have to be given to the smaller spaced correlators, which will introduce more unexpected noise since the variance of the multipath estimation becomes larger for bigger weights. From a noise perspective, the number of small spaced correlators should not be large.

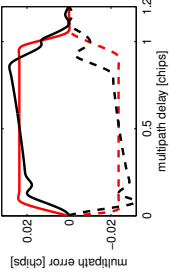
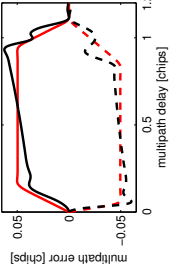
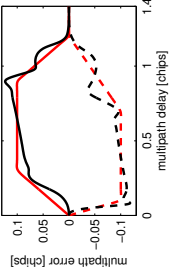
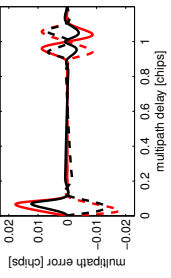
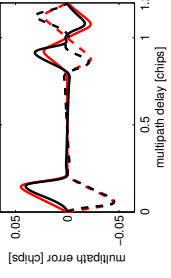
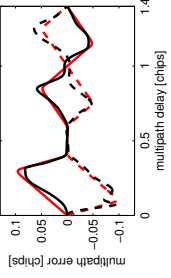
These characteristics shall be carefully considered in order to reach a compromise between the curve fitting performance and the estimation noise. The specific choice of correlators is also dependent on the specifications on the tracking loop, especially the type of discriminator (nEML or $\Delta\Delta$) and the early-late spacing. For the nEML discriminator, it should also be noted that the tracking point satisfies $D = I_E - I_L = 0$. This means if both $I_E = R(-1/2d)$ and $I_L = R(1/2d)$ are used simultaneously in the least-squares, the weights for them can be wrongly estimated because they have exactly the same value and the same contribution to curve fitting. Therefore, once the correlator serial number $\pm i$ satisfies $|i|\Delta d = 1/2d$, only one of them ($+i$ or $-i$) will be chosen in the curve fitting.

4.2.5 Applications on the BPSK-R code

The idea of the multipath envelope curve fitting has been applied to the BPSK-R(1) and BPSK-R(10) codes in Table 4.1 and 4.2, respectively. Different early-late spacings and different discriminators are used, which characterize the settings of the receiver and the true multipath errors. Other parameters, including the spacing between adjacent correlators Δd and the serial number of correlators i are used for indicating the amount and exact locations of correlators. Weights w_i are obtained from the curve fitting. The variance of the estimated multipath $\sigma_{\delta\hat{\tau}}^2$ is also calculated according to Eq.(4.31), which should be smaller or comparable with the variance of the thermal noise σ_{nEML}^2 or $\sigma_{\Delta\Delta}^2$ in either the nEML or $\Delta\Delta$ tracking loop.

In Table 4.1, a bandwidth of 24 MHz is used for the BPSK-R(1) code, which is wide enough to capture its 2 MHz mainlobe as well as several sidelobes. This indicates that the rounding effect at the correlation peak is small so that a narrow spacing discriminator (e.g., $d = 0.1$ or 0.2 chips) can be applied. For various scenarios with d ranging from 0.1, 0.2 to 0.4 chips, the spacing Δd between adjacent correlators for the curve fitting has always been chosen to be 0.05 chips, which is equal to $1/2d$ for $d=0.1$ chips but is much smaller than $1/2d$ for $d=0.2$ and 0.4 chips.

Table 4.1: Multipath envelope curve fitting performance for BPSK-R(1): $A_1/A_0 = 0.5$ chips

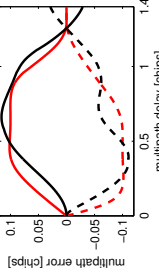
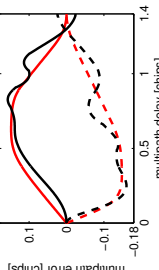
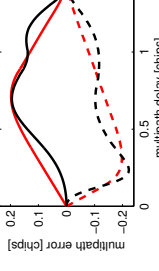
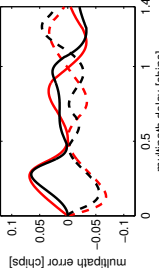
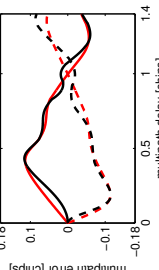
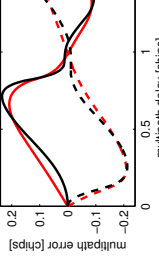
BPSK-R(1), bandwidth 24MHz, nEML discriminator			
	$d = 0.1$	$d = 0.2$	$d = 0.4$
Spacing Δd :	$1/2d$	$1/4d$	$1/8d$
Serial number i :	[-3 -1 0 2 3 4]	[-3 -2 -1 0 1 3 4 5]	[-6 -5 -4 -2 -1 0 1 2 3 5 6]
Weight w_i : ¹	[-0.31 0.50 -0.46 0.40 -0.21 0.08]	[0.07 -0.71 1.92 -0.49 -0.69 -0.15 0.03 0.01]	[-0.07 0.45 -1.49 1.54 0.01 -0.17 -0.34 -0.72 0.64 0.05 0.04]
σ_{nEML} [m]: ²	0.231	0.368	0.521
$\sigma_{\delta\hat{\tau}}$ [m]: ²	0.145	0.456	0.480
Performance: ³ (293 m/chip)			
BPSK-R(1), bandwidth 24MHz, $\Delta\Delta$ discriminator			
	$d_1 = 0.1, d_2 = 0.2$	$d_1 = 0.2, d_2 = 0.4$	$d_1 = 0.4, d_2 = 0.8$
Spacing Δd :	$1/2d_1$	$1/4d_1$	$1/8d_1$
Serial number i :	[-3 0 1 2 3 4]	[-3 -2 -1 0 1 3 4 5]	[-6 -5 -4 -2 -1 0 1 2 3 5 6]
Weight w_i : ¹	[0.04 -0.80 1.34 -0.58 -0.04 0.02]	[-0.27 -0.12 1.45 -0.59 -0.51 -0.23 0.23 0.04]	[0.29 0.08 -0.86 0.77 0.24 -0.24 -0.55 -0.36 0.59 0.45 -0.40]
$\sigma_{\Delta\Delta}$ [m]: ²	0.368	0.521	0.737
$\sigma_{\delta\hat{\tau}}$ [m]: ²	0.252	0.344	0.441
Performance: ³ (293 m/chip)			

¹ Weights are obtained by curve fitting in the least-squares.

² The code tracking loop thermal noise σ_{nEML} or $\sigma_{\Delta\Delta}$ by using nEML or $\Delta\Delta$ discriminator and the multipath estimation noise $\sigma_{\delta\hat{\tau}}$ are all calculated when $C/N_0 = 45$ dB/Hz and $B_L = 0.5$ Hz.

³ The true multipath error is in red, while the estimation thereof is in black.

Table 4.2: Multipath envelope curve fitting performance for BPSK-R(10): $A_1/A_0 = 0.5$ chips

BPSK-R(10), bandwidth 45 MHz, nEML discriminator			
	$d = 0.4$	$d = 0.6$	$d = 0.8$
Spacing Δd :	$1/2d$		
Serial number i :	$[-3 \ 0 \ 1 \ 2 \ 3 \ 4]$		
Weight w_i : ¹	$[0.28 \ -0.69 \ 0.03$ $0.35 \ 0.09 \ -0.05]$	$[-3 \ -2 \ -1 \ 0 \ 1 \ 3 \ 4]$ $[1.03 \ -2.40 \ 2.85 \ -1.49$ $-0.27 \ 0.47 \ -0.20]$	$[-1 \ 0 \ 1 \ 2 \ 3]$ $[0.70 \ -0.29 \ -0.77$ $0.42 \ -0.06]$
σ_{nEML} [m]: ²	0.089	0.118	0.147
$\sigma_{\delta\hat{\tau}}$ [m]: ²	0.045	0.267	0.114
Performance: ³ (29.3 m/chip)			
BPSK-R(10), bandwidth 45 MHz, $\Delta\Delta$ discriminator			
	$d_1 = 0.4, d_2 = 0.8$	$d_1 = 0.6, d_2 = 1.2$	$d_1 = 0.8, d_2 = 1.6$
Spacing Δd :	$1/2d_1$		
Serial number i :	$[-3 \ 0 \ 1 \ 2 \ 3 \ 4]$		
Weight w_i : ¹	$[0.34 \ -1.57 \ 1.60$ $-0.48 \ 0.15 \ -0.05]$	$[-3 \ -2 \ -1 \ 0 \ 1 \ 3 \ 4]$ $[0.02 \ -0.15 \ 1.88 \ -1.28$ $-0.45 \ -0.25 \ 0.24]$	$[-1 \ 0 \ 1 \ 2 \ 3]$ $[0.97 \ -0.61 \ -0.45$ $-0.02 \ 0.10]$
$\sigma_{\Delta\Delta}$ [m]: ²	0.147	0.180	0.208
$\sigma_{\delta\hat{\tau}}$ [m]: ²	0.069	0.184	0.127
Performance: ³ (29.3 m/chip)			

¹ Weights are obtained by curve fitting in the least-squares.² The code tracking loop thermal noise σ_{nEML} or $\sigma_{\Delta\Delta}$ by using nEML or $\Delta\Delta$ discriminator and the multipath estimation noise $\sigma_{\delta\hat{\tau}}$ are all calculated when $C/N_0 = 45$ dB/Hz and $B_L = 2$ Hz.³ The true multipath error is in red, while the estimation thereof is in black.

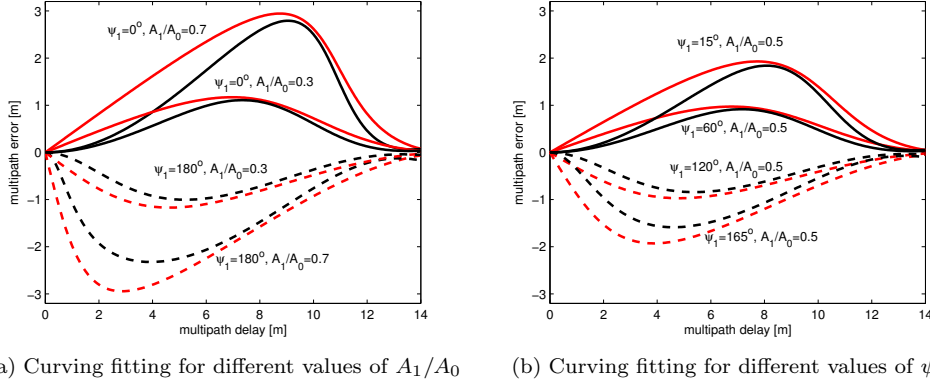


Figure 4.10: Actual multipath error (red) using the $\Delta\Delta$ discriminator and the estimation (black) thereof for different values of multipath to signal amplitude ratio and multipath phase, in the case of the BPSK-R(10) with the front-end bandwidth of 45 MHz and d of 0.4 chips.

Curve fitting performances in Table 4.1 show that the short-delay multipath can be better fitted with $\Delta d < 1/2d$ than the case of $\Delta d = 1/2d$. This result is consistent with the aforementioned characteristics on choosing correlators. Comparing the standard nEML and $\Delta\Delta$ discriminators, the $\Delta\Delta$ outperforms the nEML in that the medium-delay multipath has been greatly eliminated and the short-delay multipath can also be slightly better fitted.

For the BPSK-R(10) code in Table 4.2, the bandwidth of 45 MHz is not very high compared to its 20 MHz mainlobes. Therefore, the bandlimiting effect is more serious that requires the spacing d to be chosen a bit wider. The applications in Table 4.2 take $d = 0.4, 0.6$ and 0.8 chips for instance. It is clear for all scenario that the curving fitting performances for the short-delay multipath in the $\Delta\Delta$ discriminator are better than the nEML discriminator. Comparing the BPSK-R(10) code to the BPSK-R(1) code, lower estimation variance and smaller multipath errors can be obtained. This is mainly because the BPSK-R(10) code has a chip duration of around 29.3 m, which is only one tenth of the 293 m chip duration of the BPSK-R(1) code. This benefit compensates the drawbacks of using larger spacings and experiencing bandlimiting effects. The original maximal short-delay multipath error for the BPSK-R(10) is generally around 3 m. After the multipath estimation and subtraction from the original measurement, the remaining error can be significantly reduced to less than 1 m.

In both Table 4.1 and 4.2, weights are deduced from the envelope curve fitting for a particular case of multipath-to-signal amplitude ratio $A_1/A_0 = 0.5$ and particular multipath phases $\psi_1 = 0^\circ$ and 180° . As the multipath amplitude and phase are unknown before the estimation, it is necessary to demonstrate that the weights obtained for $A_1/A_0 = 0.5, \psi_1 = 0^\circ$ and 180° are equally able to be used in the estimation for other A_1/A_0 and ψ_1 values as well. As depicted in Figure 4.10, the agreements between the actual error and the estimation thereof are quiet good for multiple different A_1/A_0 and ψ_1 combinations thanks to the in-phase correlations between them that makes the estimation adaptable to various situations.

4.2.6 Implementation

Considering the intensive computational load of using multiple correlators, a more efficient implementation is proposed by means of a so-called “curve fitting code”, which is a linear combination of multiple early/late shifted code replicas $c^F(t) = \sum_{i=-M}^N \beta_i c(t + i\Delta d)$. The idea is to build this curve fitting code $c^F(t)$, whose cross-correlation with the incoming code $c(t)$ is equal to the term of $\sum_{i=-M}^N w_i S_i / \gamma_0$ so that the multipath estimation $\delta\hat{\tau} = (\sum_{i=-M}^N w_i S_i / \gamma_0) / I_0$ can be efficiently obtained by dividing only two correlators: the numerator is $\sum_{i=-M}^N w_i S_i / \gamma_0$ via the cross-correlation between $c^F(t)$ and $c(t)$, and the denominator is I_0 via the auto-correlation between $c(t)$ and the replica of itself. Derivations for $c^F(t)$ are as follows.

A prompt correlator I_0 in the coupled DLL-PLL tracking loops can be expressed as

$$\begin{aligned} I_0 &= R(\delta\tau) \cos(\delta\phi) \\ &= \frac{1}{L} \sum_{l=0}^{L-1} c(l)c(l - \delta\tau) \cos(\delta\phi) \end{aligned} \quad (4.34)$$

where $R(\delta\tau)$ is the auto-correlation between the incoming sampled PRN code $c(l)$ and its replica $c(l - \delta\tau)$, $\delta\tau$ and $\delta\phi$ are the code and phase tracking errors in the DLL and PLL, respectively. They are caused by thermal noise and (or) multipath. The length of the sampled PRN code is L . Then, an early or late correlator can be written as

$$\begin{aligned} I_i &= R(\delta\tau - i\Delta d) \cos(\delta\phi) \\ &= \frac{1}{L} \sum_{l=0}^{L-1} c(l)c(l - \delta\tau + i\Delta d) \cos(\delta\phi). \end{aligned} \quad (4.35)$$

Substituting (4.35) into the expression of the multipath estimation, we get

$$\begin{aligned} \delta\hat{\tau} &= \sum_{i=-M}^N w_i \frac{S_i}{S_0} \\ &= \sum_{i=-M}^N w_i \frac{\gamma_i I_i}{(1 - |i|\Delta d)\gamma_0 I_0} \\ &= \sum_{i=-M}^N w_i \frac{\gamma_i R(\delta\tau - i\Delta d) \cos(\delta\phi)}{(1 - |i|\Delta d)\gamma_0 I_0} \\ &= \sum_{i=-M}^N w_i \frac{\gamma_i \frac{1}{L} \sum_{l=0}^{L-1} c(l)c(l - \delta\tau + i\Delta d) \cos(\delta\phi)}{(1 - |i|\Delta d)\gamma_0 I_0} \\ &= \frac{1}{I_0} \frac{1}{L} \sum_{l=0}^{L-1} c(l) \sum_{i=-M}^N \frac{w_i \gamma_i}{(1 - |i|\Delta d)\gamma_0} c(l - \delta\tau + i\Delta d) \cos(\delta\phi) \\ &= \frac{1}{I_0} \underbrace{\frac{1}{L} \sum_{l=0}^{L-1} c(l)c^F(l - \delta\tau) \cos(\delta\phi)}_{\text{cross-correlation between } c^F(t) \text{ and } c(t)} \end{aligned} \quad (4.36)$$

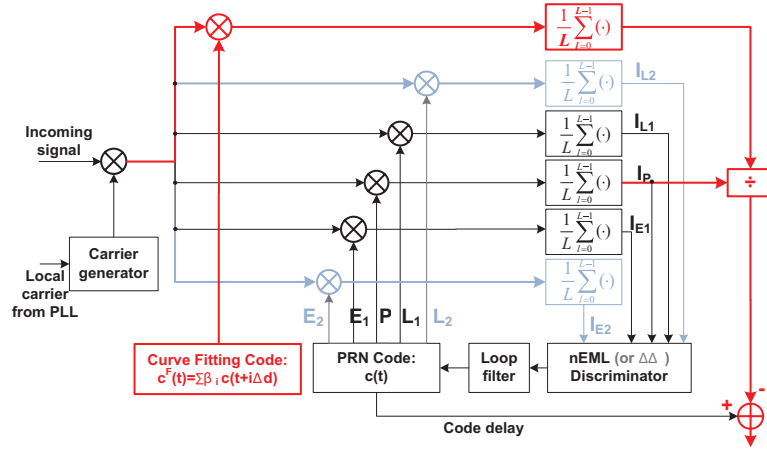


Figure 4.11: Implementation of the multipath curve fitting method with minimized computational effects

where $c^F(t)$ is defined as the *curve fitting code*, and it is equal to the linear combination of $M + N + 1$ early/late shifted code replicas

$$\begin{aligned}
 c^F(t) &= \sum_{i=-M}^N \frac{w_i \gamma_i}{(1 - |i| \Delta d) \gamma_0} c(t + i \Delta d) \\
 &= \sum_{i=-M}^N \beta_i c(t + i \Delta d) .
 \end{aligned} \tag{4.37}$$

It can be seen from Eq.(4.36) that an implementation of the cross-correlation between $c^F(t)$ and $c(t)$ in the tracking loop will equivalently deliver the same result as implementing multiple correlators and linearly combining them. Thus, the computational load in the receiver can be dramatically reduced by pre-calculating and storing $c^F(t)$ in the receiver memory and using $c^F(t)$ as a normal code replica in the implementation.

Figure 4.11 illustrates how to use $c^F(t)$ for the multipath estimation in the standard tracking loop with either a nEML or $\Delta\Delta$ discriminator. It can be seen that the cross-correlation between $c^F(t)$ and $c(t)$ (in red) is calculated in-conjunction with the tracking loop. However, the estimation process lends itself to an independent open loop where the output does not feed back to the discriminator. This means the tracking loop still keeps its own performance without the impact from the multipath estimation, thus can provide the same dynamic tolerance and thermal noise as usual.

In Figure 4.11, the nEML discriminator requires three correlators - early I_{E1} , prompt I_P and late I_{L1} correlators in the tracking loop, while the $\Delta\Delta$ discriminator requires two extra correlators - very early I_{E2} and very late I_{L2} correlators for better multipath rejection capability. The multipath estimation process is implemented by only introducing one more correlator apart from those correlators required by discriminators, assuring a very low computational load. The coefficients β in the $c^F(t)$ code need to be chosen consistently with the type of discriminator before storing it into the receiver memory.

4.2.7 Limitations

In an environment consisting of several reflectors, a large number of multipath signals may exist at the same time. If there is not a dominating component among them, the superposition of different multipath with different delays and phases will break the in-phase correlation between the multipath error and signal strength. In that case, the curve fitting method will lose its effectiveness. However, if a dominating multipath component exists, which has a significant larger magnitude than other components, the in-phase correlation is still present and the overall error can still be estimated.

Most of space applications have a relatively “clean” multipath environment. This can be demonstrated by observing the sinusoidal behavior in the code, carrier phase and (or) signal to noise ratio (SNR) time series (Bilich and Larson, 2007; Reichert and Axelard, 1999; Reichert, 1999). In phase or code observables, residuals should be analysed or differences should be formed in order to eliminate all other error sources (Smyrnaioi et al., 2013) and isolate multipath impact. In contrast, the SNR observable can represent the signal strength change and is a direct way to observe multipath without any sophisticated data post-processing. The SNR can thus be used to check the sinusoidal behavior caused by multipath.

4.3 Verification

4.3.1 Software-defined signal simulator and receiver

A software-defined signal simulator and receiver have been built to demonstrate the multipath mitigation performance. They can be used in the early design phase to foretell hazardous environmental configurations that can cause severe multipath. They can also aid in finding the best antenna type, location and orientation within a given environment, and provide a quantitative estimate of multipath errors on measurements before and after the multipath mitigation. Furthermore, implementing the software-defined simulator and receiver is the most convenient starting point as they are easy and transparent to reconfigure and control. Some unwanted error sources, e.g., the atmospheric error and clock offset, can be avoided so that the isolated multipath effects will be highlighted.

The multipath simulator architecture, depicted in Figure 4.12, consists of a signal & multipath generator, bandpass filter and quantization chain. The multipath, with a certain delay, phase and amplitude, has been added to the LOS signal. These multipath parameters will be determined by the antenna-reflector geometry and the antenna gain pattern. White noise is also added to the compound signal before it goes to the bandpass filter and quantization. Therefore, the simulator emulates the LOS and multipath propagations and also considers the receiver front-end conditioning process so that a digitalized noisy intermediate-frequency (IF) signal can be produced. Several similar multipath simulators in Byun et al. (2002) and Smyrnaioi et al. (2013) demonstrated that the simulated multipath errors have a good representation of the real multipath when the antenna-reflector geometry, antenna gain pattern and signal polarization have been carefully characterized.

The obtained digitalized signal is then processed in a software receiver through the acquisition and tracking process, see Figure 4.13. Acquisition is a global search in a two dimensional space for the approximate values of code delay and Doppler frequency. The acquisition results are fed into the coupled DLL and PLL tracking

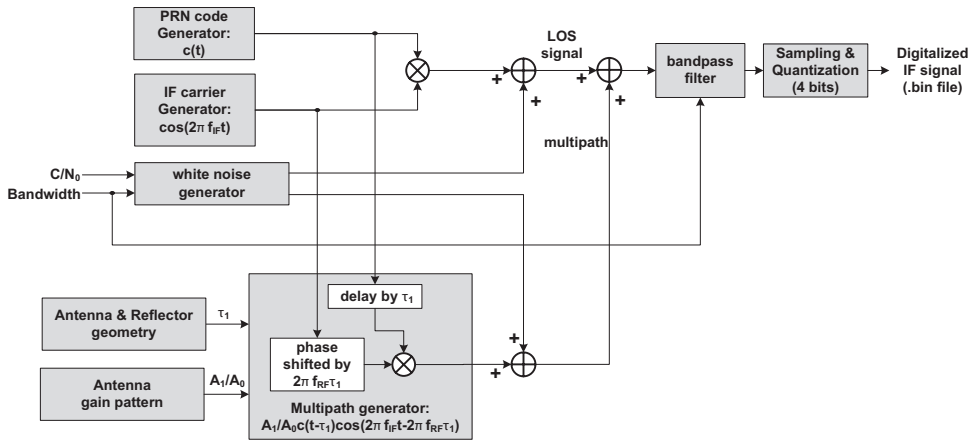


Figure 4.12: Software-defined multipath simulator architecture

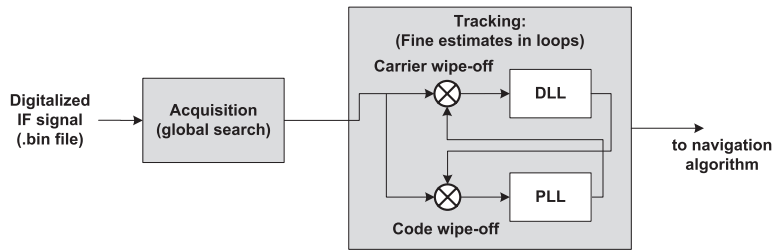


Figure 4.13: Software-defined receiver architecture

loops to refine the estimates and also track the dynamic change. Extensive description on the acquisition and tracking can be found in chapter 2. The implementation of the software receiver in this thesis is based on the work in Borre et al. (2007), where the DLL has been modified to the one in Figure 4.11 to correct the multipath contaminated code delay. Both the nEML and $\Delta\Delta$ discriminators have been implemented and compared.

4.3.2 Simulation settings

A simple scenario of a plane reflector (e.g. the solar panel) on the spacecraft is assumed for simulation. The spacecraft receives the BPSK-R GNSS-like ranging signals from the other spacecraft in a formation flying.

In this scenario, antenna is assumed with gain patterns in Figure 4.14, similar to NovAtel NOV702GG antenna. For each instant in time, the electric field vector is decomposed into two orthogonal circular polarization states: the right-hand and left-hand circular polarizations (RHCP and LHCP). The antenna is designed with different gain patterns to different polarization components, so that the RHCP component is reinforced while the LHCP component is suppressed. A regular GNSS-like LOS signal consists of only a pure RHCP component. Its polarization will be changed to a combination of LHCP and RHCP or pure LHCP after reflection, depending on reflection angles. According to Smyrnaiois et al. (2013), for a reflection

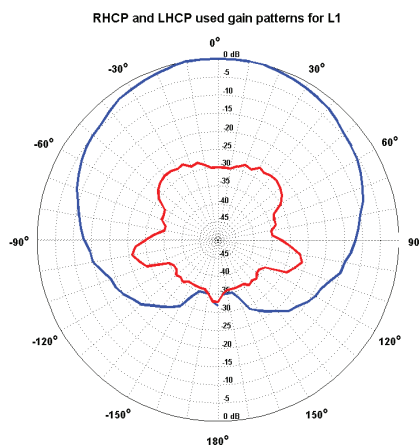


Figure 4.14: The Novatel NOV702GG antenna gain patterns [dB] for both orthogonal polarizations (RHCP in blue and LHCP in red) with respect to boresight angles (NovAtel, 2009)

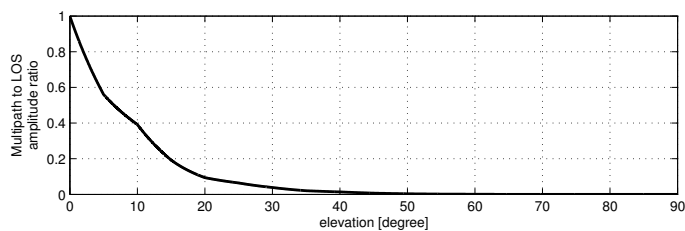


Figure 4.15: The multipath to LOS amplitude ratio as a function of the elevation angle for a perfect plane reflector beneath the antenna (Smyrnaiois et al., 2013)

angle between 0° and 30° (e.g., for the concrete reflector), both RHCP and LHCP exist but still the RHCP is larger in magnitude. The reflected signal is now called right-hand elliptically polarized (RHEP), with the eccentricity of the polarization ellipse getting bigger as the reflection angle increases. In this case, the reflected signal has little loss and both the LOS and multipath will be reinforced with a certain gain. For instance, for the extreme case of the reflection angle of 0° , the reflected signal (multipath) have most of the signal energy in the RHCP component. Since the LOS signal is also RHCP, the antenna thus applies a very similar gain for both the LOS and multipath, leading to the multipath to LOS amplitude ratio approaching to 1. This is illustrated in 4.15.

When the reflection angle is between 30° and 90° , the LHCP component turns to have a higher magnitude, which results in the change of the polarization from RH to LH. As the reflection angle approaches 90° , the reflected signal turns to a pure LHCP. Notice that the reflected signal with a reflection angle of 90° has the angle of arrival of -90° . Therefore, in Figure 4.14, the LOS with RHCP has a minimum gain attenuation of 0 dB (in blue), while the multipath is LHCP with angle of arrival of -90° , thus having amplitude attenuation of -32 dB (in red). The multipath thus has be strongly attenuated compared to the LOS. In Figure 4.15, we can also see that

Table 4.3: Specifications for the software simulator and receiver

Simulator	PRN Code	BPSK-R(10)
	IF frequency [MHz]	30
	Sampling frequency [MHz]	90
	Filtering Bandwidth [MHz]	45
	C/N_0 [dB]	60
	Multipath delay τ_1	Eq. (4.38)
	Multipath phase ψ_1	$\psi_1 = 2\pi f\tau_1$
	Multipath to LOS amplitude ratio	Figure 4.15
	Elevation	$0^\circ - 90^\circ - 0^\circ$
Antenna-reflector distance H [m]	2	
Receiver	DLL discriminator	$\Delta\Delta$ or nEML
	DLL early-late spacing [chips]	$d_1=0.4, (d_2=0.8 \text{ for } \Delta\Delta)$
	DLL noise bandwidth [Hz]	1
	PLL discriminator	arctan
	PLL noise bandwidth [Hz]	40
	Correlator serial number and weights	Table 4.2

the multipath to LOS amplitude ratio in general is big at low elevations, while it approaches zero as the elevation increases.

A simple multipath geometry scenario is assumed with the plane reflector beneath the antenna for simulation. The multipath delay can be expressed as function of the elevation el and the perpendicular distance between the antenna and the reflector H ,

$$\tau_1 = 2H \sin(el). \quad (4.38)$$

Assuming a constant relative orbit angular velocity between two spacecraft, the elevation angle is changing at a constant speed from 0° to 90° and going back to 0° when the spacecraft (where the signal is transmitted) passes throughout the coverage of the antenna on the other spacecraft (where the signal is received). This is a simple antenna-reflector geometry scenario. In fact, the relative geometry and relative attitude between spacecraft determines the change of el , and thus the change of multipath delay.

The settings for the software simulator and receiver are specified in Table 4.3.

4.3.3 Performance

The impact of multipath in the software receiver is illustrated in Figure 4.16 for the BPSK-R(10) with the $\Delta\Delta$ discriminator. As can be seen, correlators, DLL and PLL discriminator outputs are all impacted by multipath oscillations. As the elevation angle changes from 0° to 90° , the multipath delay changes from 0 to maximum 4 m according to Eq.(4.38) with an antenna-reflector distance H of 2 m. The multipath to LOS amplitude ratio also changes from approximately 1 to 0 according to Figure 4.15. To this end, the resulted multipath error is highly oscillated at low elevations mainly due to the high multipath to LOS amplitude ratio. The multipath error has been significantly attenuated at high elevations as the amplitude ratio approaches zero. The oscillation frequency is dependent on the multipath delay, while the multipath magnitude is determined by both the amplitude ratio and multipath delay.

The in-phase correlation between the multipath error (bottom figure) and the signal strength (or correlator outputs, top figure) is clearly visible. The bottom figure

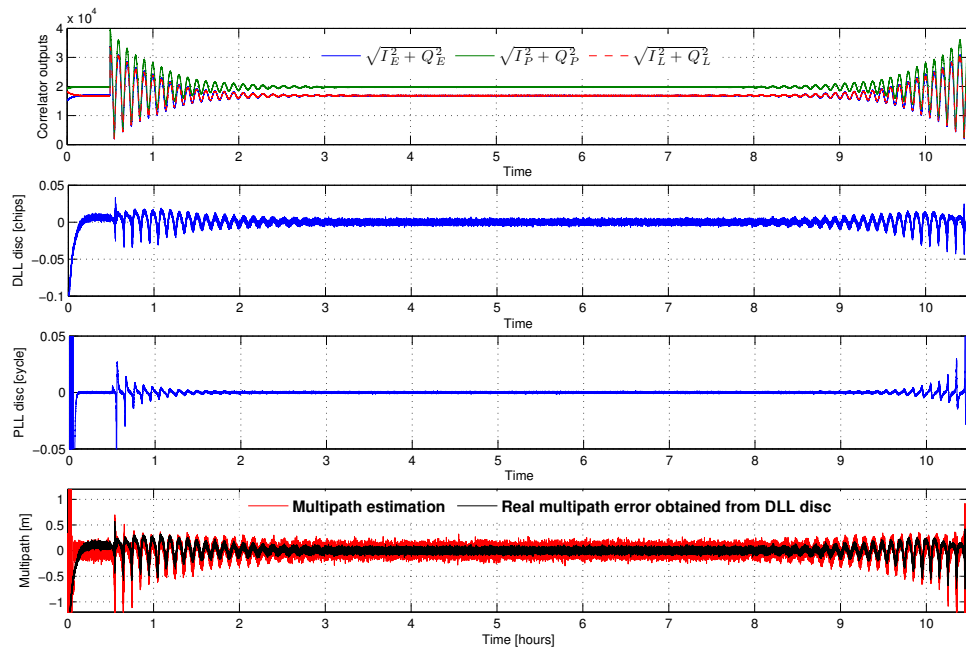


Figure 4.16: Multipath estimation in the software receiver. The implemented code is BPSK-R(10) and the $\Delta\Delta$ DLL discriminator is used. The first 0.5 h are without multipath, the rest of time is multipath contaminated when the elevation angle changes from 0° to 90° and goes back to 0° in the end. The multipath delay and relative amplitude with respect to the LOS signal changes accordingly.

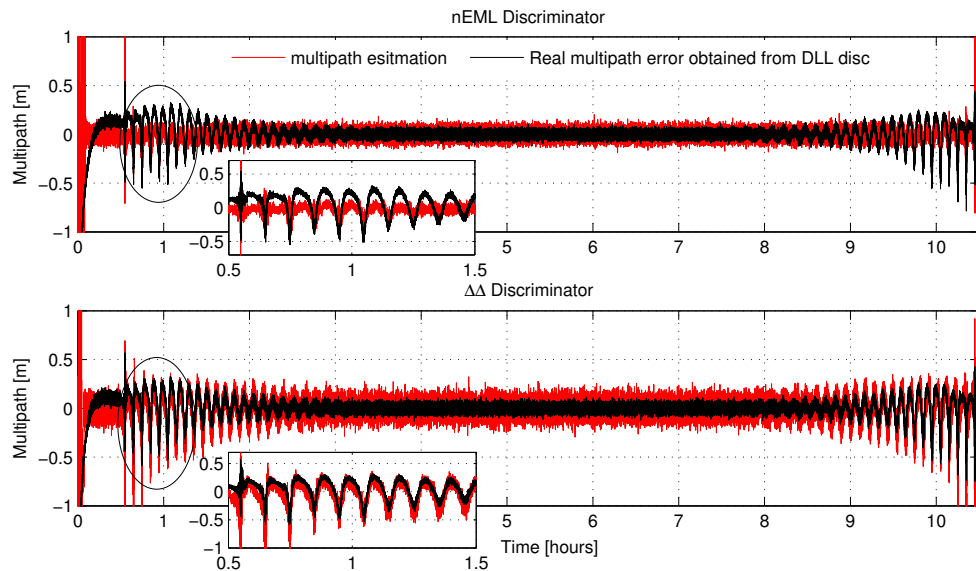


Figure 4.17: Multipath estimation in the nEML vs. $\Delta\Delta$ discriminator. Receiver specifications are in Table 4.3 and the parameters used in the multipath estimation are in Table 4.2

shows that the multipath estimation result (in red) has a very good agreement to the actual error (in black). During the period between 0.5 h to 1 h, the multipath is overestimated compared to its true value. In this period, the multipath has a very short delay of less than 1.3 m and more difficult to estimate.

Figure 4.17 illustrates the comparison between the nEML and $\Delta\Delta$ discriminators using estimation parameters in Table 4.2. As can be seen, the multipath is hardly able to estimate in the nEML discriminator. This is due to the fact that the correlator weights obtained in the curve fitting process are optimized for a complete range of multipath delay from 0 to 1.4 chips (around 0 to 40 m for the BPSK-R(10) code). It is not optimized for short delays, but a compromise for all short, medium and large delayed multipath (see the performance in Table 4.2. In comparison, using the $\Delta\Delta$ discriminator, the short-delay multipath can be better estimated. Therefore, the $\Delta\Delta$ discriminator is suggested to use, especially for space applications where short-delay and very-short-delay multipath are dominating.

4.4 Chapter summary

This chapter proposed a promising multipath mitigation method particularly for the short-delay multipath. The method was developed based on the fact that the signal strength (reported by early or late correlators) has an in-phase correlation with the multipath error. By linearly combining multiple signal strength estimators, the multipath error can be accurately estimated. The weights for the linear combination were obtained by curve fitting based on the least-squares adjustment. A reference multipath error curve is required for curve fitting. This reference curve can be generated by using a standard narrow correlator or a double delta correlator. It was found that the double delta correlator can result in a better estimation of the multipath error than the standard narrow correlator in the least-squares sense, especially for the short-delay multipath.

Apart from the performance evaluation of this new multipath estimation method, this chapter also derived the closed-form expression for the estimation noise. Furthermore, a simple implementation strategy was proposed that enables the multipath estimation operated in conjunction with the tracking loop with a minimum computational effect. Software simulator and receiver were also built, in which the effectiveness of the proposed method in mitigating short delay multipath were very well demonstrated.

Chapter 5

Carrier Phase Multipath Effects and Mitigation Methods

This chapter proposes a promising multiple antenna-based carrier phase multipath estimation method using an extended real-valued or complex-valued Kalman filter.

The chapter starts with the categorisation to existing carrier phase multipath estimation methods in section 5.1. SNR-based, multiple antenna-based, and mapping-based methods are three main categories in the phase multipath mitigation domain. In section 5.2, the theory of multiple antenna-based method is extended and improved for real-time applications by using an extended real-valued or complex-valued Kalman filter. Cascaded procedures are also proposed in order to split the multipath correction process into cascaded filters before and after fixing integer ambiguities. The filter performance is evaluated in section 5.3 in three aspects: the sensitivity to initial conditions, the tolerance to large noise on observations and the robustness in multi-reflection environments. Finally, in section 5.4, multipath effects on the integer ambiguity resolution are examined.

This chapter and chapter 4 both focus on the discussion on multipath and the associated mitigation methods. Chapter 4 deals with the code multipath, while this chapter addresses the carrier phase multipath. The phase multipath has more significant effects in impeding a reliable ambiguity resolution. However, the constrained ambiguity resolution, introduced in chapter 3, will be demonstrated to be more robust to multipath in this chapter.

5.1 Problem statement and existing Methods

The topic of multipath mitigation in phase measurements has received considerable attention in the literature. However, most of the existing research on the phase multipath mitigation is based on the assumption that ambiguities are fixed such that the phase residual, dominated only by multipath, can be used to construct the multipath (Reichert, 1999; Ray, 2000). On the other hand, it is well known that the integer ambiguities are difficult to resolve in the presence of phase multipath. Therefore, the phase multipath mitigation and the integer ambiguity resolution are

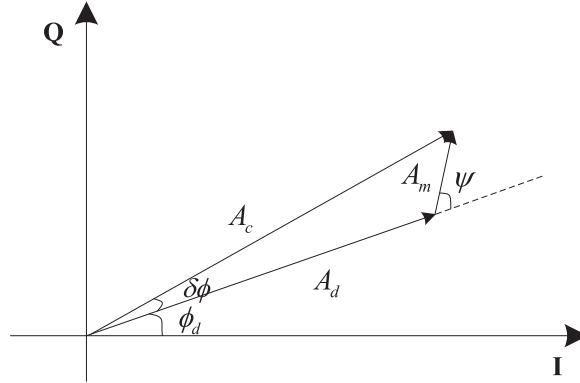


Figure 5.1: Phasor diagram

coupled obstacles in high precision applications. Several other multipath estimation methods, i.e. the signal-to-noise ratio (SNR) based methods, do not require ambiguities to be fixed. However, they usually work only in post-processing applications. It is of great importance to design a real-time method to facilitate the ambiguity resolution as well as to increase the positioning accuracy on the fly.

Before the new estimation method is proposed, the phase multipath properties are characterized and several existing methods are collated in the following.

5.1.1 Characterizing phase multipath and SNR

The carrier tracking loop can be represented in terms of a phasor diagram that shows the phase relationship between I and Q channels (Figure 5.1).

When no multipath is present, the phasor diagram would contain a single phasor for the direct signal of amplitude A_d . Any misalignment of the local replica and the incoming carrier results in a nonzero phase angle ϕ_d , which is measured and tracked by the carrier tracking loop.

In the presence of multipath, one or more additional phasors are introduced to the phasor diagram. The carrier tracking loop attempts to track a composite signal which is the vector sum of all phasors (direct plus multipath). By tracking the composite phasor, the carrier tracking loop reports an incorrect phase measurement with phase error $\delta\phi$ due to multipath. This phase error can easily be derived from geometric relationships expressed in the phasor diagram and can be described in terms of the multipath amplitude A_m and phase ψ

$$\begin{aligned}\delta\phi &= \arctan\left(\frac{A_m \sin \psi}{A_d + A_m \cos \psi}\right) \\ &= \arctan\left(\frac{\alpha \sin \psi}{1 + \alpha \cos \psi}\right)\end{aligned}\quad (5.1)$$

$$\begin{aligned}A_c &= \sqrt{(A_d + A_m \cos \psi)^2 + (A_m \sin \psi)^2} \\ &= A_d \sqrt{1 + \alpha^2 + 2\alpha \cos \psi}\end{aligned}\quad (5.2)$$

where $\alpha = A_m/A_d$ denotes the relative amplitude coefficient of the multipath with respect to the direct signal.

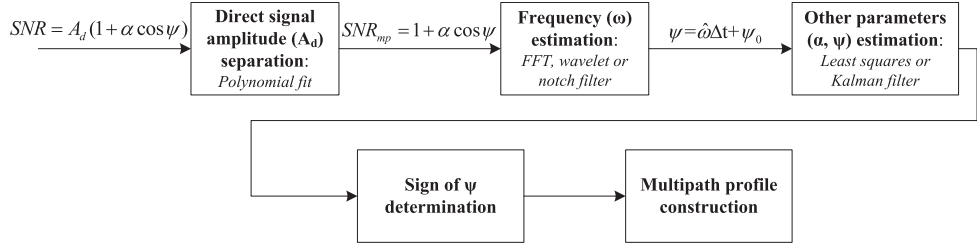


Figure 5.2: Basic steps in the existing SNR-based carrier phase multipath estimation methods

In terms of the phasor diagram, changes in ψ cause the multipath phasor to spin around the end of the direct phasor. The phase error $\delta\phi$ then oscillates between an absolute maximum when $\psi = 90^\circ$ or 270° and a minimum (no phase error) when $\psi = 0^\circ$ or 180° . Likewise, the composite signal amplitude A_c can also be expressed in terms of α and ψ , which however shows a maximum at $\psi = 0^\circ$ or 180° and a minimum at $\psi = 90^\circ$ or 270° . The change in ψ determines the time-variability of the multipath.

Approximates to $\delta\phi$ and A_c are given below under the assumption of a small value of α ,

$$\delta\phi \approx \frac{\alpha \sin \psi}{1 + \alpha \cos \psi} \approx \alpha \sin \psi \quad (5.3)$$

$$A_c \approx A_d(1 + \alpha \cos \psi) \quad (5.4)$$

where an obvious *out-of-phase* (or *quadratic*) relationship between A_c and $\delta\phi$ can be observed since $\alpha \sin \psi$ and $\alpha \cos \psi$ have 90° phase shift. This relationship holds true for a single dominating multipath environment but will break in the presence of n multipathed signals when the superposition of n sine or cosine waveforms, i.e., $\sum_{i=1}^n \alpha_i \sin \psi_i$ or $\sum_{i=1}^n \alpha_i \cos \psi_i$, plays a role. The composite signal amplitude A_c is obtained from the prompt correlator in the code tracking loop. The signal-to-noise (SNR) measurement, provided by most of the existing receivers, is equivalent to the value of A_c . To be consistent with the literature, the notation SNR is used instead of A_c to denote the composite signal amplitude in following sections.

5.1.2 SNR based multipath estimation

The SNR data is an additional observable quantity that can be used to assess and potentially correct the carrier phase multipath. Previous studies have incorporated SNR measurements in correcting the phase multipath in aerospace environments (Axelrad et al., 1996; Comp and Axelrad, 1998; Reichert and Axelard, 1999) and also for geodetic applications (Scappuzzo, 1997; Bilich, 2006; Bilich et al., 2008; Rost and Wanninger, 2009). Although their methods differ in specific implementations, basic steps as depicted in Figure 5.2 are more or less involved in their estimation procedure.

As shown in Eq. (5.2), the direct signal amplitude A_d is the dominant trend in a SNR time series. The multipath with a smaller amplitude is modulated on top of A_d . To best reveal the time-variability of the multipath, it is often necessary to separate the contribution of the direct signal from the multipath-contaminated SNR in the first step. Since A_d is mainly determined by the antenna gain changes with respect to the incident signal direction, the estimation of A_d in Axelrad et al.

(1996) thus requires the knowledge of antenna gain patterns, the satellite elevation and an approximate *a priori* vehicle attitude. Alternately, Bilich (2006) and Rost and Wanninger (2009) make use of a low-order polynomial fit in the SNR time series in order to estimate the dominant amplitude trend contributed by the direct signal. Fitting a polynomial may result in sections of the remaining SNR time series with non-zero mean. This changing mean could result from either a slowly varying multipath component (i.e. a close-in reflector) or a direct amplitude residual due to the non-perfectly modelled polynomial fit (Bilich, 2006).

After the direct signal amplitude removal, the next step is to reveal multipath parameters from the time-varying oscillatory behavior of SNR. In theory, it is possible to estimate the multipath frequency ω in addition to α and ψ using a least-squares (LS) method or a Kalman filter (KF). However, as demonstrated in Scappuzzo (1997); Axelrad et al. (1996); Comp and Axelrad (1998) and Bilich et al. (2008), a LS or KF with a single SNR as input is insufficiently robust to estimate a non-stationary ω in the presence of noise; initializing the LS or KF with the pre-estimated frequency $\hat{\omega}$ increases robustness and assists in the convergence of the multipath parameter estimation. Thus, the frequency has been primarily extracted from a batch of measurements by Fourier or wavelet transforms Scappuzzo (1997); Bilich et al. (2008) or by an adaptive notch filter (Comp and Axelrad, 1998) before the LS or KF is applied. The implementation of the KF afterwards, for instance, takes $\mathbf{x} = [\alpha \cos(\psi), \alpha \sin(\psi)]^T$ as the state vector, which has been propagated by utilizing the phase propagation of $\psi_{k+1} = \psi_k + \hat{\omega}_k \Delta t$ with a pre-estimated frequency $\hat{\omega}_k$ from one time t_k to the next t_{k+1}

$$\begin{bmatrix} \alpha_{k+1} \cos(\psi_{k+1}) \\ \alpha_{k+1} \sin(\psi_{k+1}) \end{bmatrix} = \begin{bmatrix} \cos(\hat{\omega}_k \Delta t) & -\sin(\hat{\omega}_k \Delta t) \\ \sin(\hat{\omega}_k \Delta t) & \cos(\hat{\omega}_k \Delta t) \end{bmatrix} \begin{bmatrix} \alpha_k \cos(\psi_k) \\ \alpha_k \sin(\psi_k) \end{bmatrix} + \mathbf{Q} \quad (5.5)$$

where $\Delta t = t_{k+1} - t_k$ is the time interval for the state update, and the process noise \mathbf{Q} for the state propagation needs to account for the unmodelled amplitude update $\alpha_k \rightarrow \alpha_{k+1}$ and the imperfect estimation of $\hat{\omega}_k$. The multipath amplitude α and phase ψ can then be extracted from the orthogonal sine and cosine state pair.

Since the oscillation in the SNR measurement is driven by the cosine of ψ , it is insensitive to the sign of the change in the multipath phase. This is problematic because the carrier phase multipath error $\delta\phi$ in Eq.(5.1) is sensitive to the $\sin\psi$ term. An incorrect determination of the sign of ψ will yield an inverted phase correction profile, essentially doubling the potential multipath error instead of removing it. Therefore, in Comp and Axelrad (1998), before the actual multipath profile is constructed, the proper sign of ψ is determined by checking all possible multipath profiles with different signs against phase residuals. The correct sign produces the lowest root mean squares error. Simpler solutions are suggested in Bilich (2006), e.g., checking whether the satellite is ascending or descending as a function of time to determine the sign of ψ , or obtaining aids from the pseudorange multipath oscillations.

To sum up the SNR-based methods, they are effective to correct the phase multipath, however, should only be used in post-processing applications as the direct signal amplitude removal process and the frequency estimation process generally require a batch of sufficiently long measurements.

5.1.3 Multi-antenna based multipath estimation

Taking advantage of the fact that multipath errors have spatial correlations between multiple closely spaced antennas, Ray (1999, 2000) introduced an extended Kalman filter (EKF) for the single differenced phase multipath removal by taking single differenced phase residuals as observations.

Recall that the phase multipath error in Eq.(5.1) is a function of α and ψ . For several closely spaced antennas in an array, an identical gain pattern (in all directions) for each antenna can be assumed. This introduces identical amplification to the direct signal and identical attenuation to the multipath reflected from the same reflector. Thus, the relative amplitude coefficient α_i for each antenna ($i = 1, 2, \dots, n$) can be assumed to be the same, $\alpha = \alpha_i$. Furthermore, the multipath phase ψ_i of antenna i ($i \neq 1$) with respect to the phase ψ_1 of antenna 1 (reference) has a spatial correlation (Ray, 2000)

$$\psi_i = \psi_1 + d_{i1}f(el, az) \quad (5.6)$$

where d_{i1} is the *a-priori* baseline length between antenna 1 and i , and $f(el, az)$ is a function of the signal elevation (el) and azimuth (az). Thus, an extended Kalman filter was built by Ray (2000) with the state \mathbf{x}

$$\mathbf{x} = [\psi_1, \alpha, el, az]^T . \quad (5.7)$$

Single-differenced (SD) carrier phase residuals between antennas are used to update the state variables after removing the true range difference and integer ambiguities, which leave only the difference of multipath and phase noise between antennas (Ray, 1999)

$$\begin{aligned} Res_{\Delta\phi_{i1}} &= \delta\phi_i - \delta\phi_1 + \Delta\varepsilon \\ &= \arctan\left(\frac{\alpha \sin(\psi_i)}{1 + \alpha \cos(\psi_i)}\right) - \arctan\left(\frac{\alpha \sin(\psi_1)}{1 + \alpha \cos(\psi_1)}\right) + \Delta\varepsilon \\ &= \arctan\left(\frac{\alpha \sin(\psi_1 + d_{i1}f(el, az))}{1 + \alpha \cos(\psi_1 + d_{i1}f(el, az))}\right) - \arctan\left(\frac{\alpha \sin(\psi_1)}{1 + \alpha \cos(\psi_1)}\right) + \Delta\varepsilon \\ &= g(\psi_1, \alpha, el, az) + \Delta\varepsilon \end{aligned} \quad (5.8)$$

where $Res_{\Delta\phi_{i1}}$ denotes the SD phase residual, $\Delta\varepsilon$ is the SD phase noise, and $g(\psi_1, \alpha, el, az)$ is a nonlinear function of state variables which should be linearized in the EKF.

It was reported in Ray (1999) that up to 60% improvement in terms of the positioning accuracy can be achieved after removing the multipath. However, it is not clear what the dynamic model is (how state variables propagate in time) in the EKF. It was only stated in Ray (2000) that the EKF performance is very sensitive to the process noise on each state variable and an empirical process noise was used in his algorithm. Another limitation of this method is that it is only effective in mitigating phase multipath after integer ambiguities are fixed and the true range difference is removed from the SD phase measurement, in which way the phase residual can be extracted and treated as observations in the EKF. However, it is difficult in most applications to obtain the phase residual in a way of keeping the embedded multipath undistorted as formulated as in Eq.(5.8).

5.1.4 Multipath mapping

Exploiting the repeatability of the multipath error in a specific environment, several researchers also recommended various means of devising maps of the multipath environment surrounding an antenna. These maps provide multipath corrections for each satellite signal as a function of the azimuth and elevation. They can be stored in memory as look-up tables on the fly. Each table is made up of bins with certain azimuth and elevation intervals. The bin size depends on the velocity of variations of multipath errors with respect to the signal direction of arrival. Small bins give high accuracy but also require more memory to store (Pasetti and Giulicchi, 1999). Other alternative approaches such as compiling multipath corrections by a spherical harmonic approximation model or a linear polynomial model only require to store the coefficients of the model in the memory and are thus less memory consuming (Reichert, 1999). The look-up table or the coefficients of the model shall be obtained in advance on ground calibration, i.e., in an anechoic chamber, before they can be stored in the memory of the on-board computer on the fly.

The limitation of these multipath mapping approaches is that they only work well if the antenna environment remains constant. Studies have shown that the phase multipath is sensitive to even small environmental changes (Axelrad et al., 1996). For example, on the PRISMA mission, the multipath map construction was performed both on ground and on the fly. Two constructed maps show an overall consistency but differences in some areas reach a magnitude as big as the error itself. Discrepancies can be attributed to the imperfect fidelity of the radio electric mock-ups in the anechoic chamber but may also result from the potential changes between the ground and flight calibration environments. It is also mentioned in Delpech et al. (2011) that the amount of effort needed to improve the calibration fidelity at radiated level appears very heavy and will never represent a full warranty of performance enhancement on the fly.

5.2 Theory of multi-antenna based multipath estimation on the fly

In either the SNR based or multi-antenna based multipath estimation procedure, a fundamental element in measurements is a sine or cosine waveform with the time-varying amplitude, frequency and phase. The amplitude and phase are relatively easy to estimate with the *a priori* pre-estimated frequency after a FFT/wavelet transform or a notch filter. However, this leads to a post-processing approach. It is of great importance to have a reliable and robust estimator for the simultaneous amplitude, frequency and phase estimation on the fly. Inspired by the method in Ray (2000), both real-valued and complex-valued extended Kalman filters are designed in the following to cope with this task.

5.2.1 Kalman filter and extended Kalman filter

Kalman filter (KF)

A Kalman filter (KF) is a recursive state estimation method. It utilizes a time series of observations $\mathbf{y}_1, \dots, \mathbf{y}_{k+1}$ up to and including the one made at time t_{k+1} as well as the dynamics of the state propagation from one time to another, to determine the state vector at time t_{k+1} . More formally, the KF operates the *time update (predictor)*

and *measurement update (corrector)* recursively to produce a statistically optimal estimate of the underlying system state vector.

The KF applies to the case that both the dynamic model and the observation model are linear with respect to the state vector

$$\dot{\mathbf{x}}(t) = \mathbf{F}(t)\mathbf{x}(t) + \mathbf{G}(t)\mathbf{w}(t) \quad (5.9)$$

$$\mathbf{y}(t) = \mathbf{H}(t)\mathbf{x}(t) + \mathbf{v}(t) \quad (5.10)$$

where $\mathbf{x}(t)$ is the state vector, $\mathbf{y}(t)$ is the measurement vector, $\mathbf{v}(t)$ is the noise on the measurement, $\mathbf{w}(t)$ is the system noise in the dynamic model due to unaccounted system perturbations, $\mathbf{F}(t)$ is the dynamic matrix, $\mathbf{G}(t)$ is the coefficient to shape the system noise, and $\mathbf{H}(t)$ is called *design matrix* to link the observation with the state variables.

We assume that measurements are only available at specific values of time, at $t = t_k$, $k = 1, 2, \dots$; thus, the measurement equation can be treated as a discrete-time equation, whereas the state equation is a continuous-time equation and shall be discretized. The approach to discretizing the state equation begins with the solution of Eq.(5.9) (Mendel, 1995)

$$\mathbf{x}(t) = \Phi(t, t_0)\mathbf{x}(t_0) + \int_{t_0}^t \Phi(t, \tau)\mathbf{G}(\tau)\mathbf{w}(\tau)d\tau \quad (5.11)$$

where $\Phi(t, \tau)$ is called *state transition matrix*. For time $t \in [t_k, t_{k+1}]$, set $t_0 = t_k$ and $t = t_{k+1}$, the state propagation model is obtained

$$\mathbf{x}(t_{k+1}) = \Phi(t_{k+1}, t_k)\mathbf{x}(t_k) + \underbrace{\int_{t_k}^{t_{k+1}} \Phi(t_{k+1}, \tau)\mathbf{G}(\tau)\mathbf{w}(\tau)d\tau}_{\mathbf{w}_k} \quad (5.12)$$

where \mathbf{w}_k is the discrete-time noise sequence. It has a covariance of

$$\mathbf{Q}_{k+1|k} = \int_{t_k}^{t_{k+1}} \int_{t_k}^{t_{k+1}} \Phi(t_{k+1}, \tau)\mathbf{G}(\tau)\mathbf{Q}(\tau)\mathbf{G}^T(\tau)\Phi^T(t_{k+1}, \tau)d\tau d\tau \quad (5.13)$$

where $\mathbf{Q}(\tau)$ is the covariance of $\mathbf{w}(t)$. $\mathbf{Q}_{k+1|k}$ is called *process noise* of the underlying system. In general, we shall compute $\Phi(t_{k+1}, t_k)$ and $\mathbf{Q}_{k+1|k}$ using numerical integration, and they change from one time interval to the next when $\mathbf{F}(t)$, $\mathbf{G}(t)$, and $\mathbf{Q}(t)$ change from one time interval to the next. Great simplifications of calculating $\Phi(t_{k+1}, t_k)$ and $\mathbf{Q}_{k+1|k}$ can be made when $\mathbf{F}(t)$, $\mathbf{G}(t)$, and $\mathbf{Q}(t)$ are approximately constant during the time interval $[t_k, t_{k+1}]$, i.e., if

$$\mathbf{F}(t) = \mathbf{F}_k, \mathbf{G}(t) = \mathbf{G}_k, \mathbf{Q}(t) = \mathbf{Q}_k, \text{ for } t \in [t_k, t_{k+1}], \quad (5.14)$$

the transition matrix $\Phi_{k+1|k}$ is then equal to (Mendel, 1995)

$$\begin{aligned} \Phi_{k+1|k} &= e^{\mathbf{F}\Delta t} \\ &= \mathbf{I} + \mathbf{F}\Delta t + \mathbf{F}^2\frac{\Delta t^2}{2} + \mathbf{F}^3\frac{\Delta t^3}{3!} + \dots \end{aligned} \quad (5.15)$$

where $\Delta t = t_{k+1} - t_k$.

Given the above calculations, the time update (a predictor step) in a KF includes the state propagation $\mathbf{x}_{k|k} \rightarrow \mathbf{x}_{k+1|k}$ and the covariance propagation $\mathbf{P}_{k|k} \rightarrow \mathbf{P}_{k+1|k}$, which can be written as

$$\hat{\mathbf{x}}_{k+1|k} = \Phi_{k+1|k} \hat{\mathbf{x}}_{k|k} \quad (5.16)$$

$$\hat{\mathbf{P}}_{k+1|k} = \Phi_{k+1|k} \hat{\mathbf{P}}_{k|k} \Phi_{k+1|k}^T + \mathbf{Q}_{k+1|k} \quad (5.17)$$

where \mathbf{P} denotes *the covariance of the state vector*.

The measurement update (a corrector step) in the KF is given as

$$\hat{\mathbf{x}}_{k+1|k+1} = \hat{\mathbf{x}}_{k+1|k} + \mathbf{K}_{k+1} \mathbf{v}_{k+1} \quad (5.18)$$

$$\hat{\mathbf{P}}_{k+1|k+1} = (\mathbf{I} - \mathbf{K}_{k+1} \mathbf{H}_{k+1}) \hat{\mathbf{P}}_{k+1|k} \quad (5.19)$$

with the *gain matrix* \mathbf{K}_{k+1} and the *innovation residual* \mathbf{v}_{k+1}

$$\mathbf{K}_{k+1} = \hat{\mathbf{P}}_{k+1|k} \mathbf{H}_{k+1}^T (\mathbf{H}_{k+1} \hat{\mathbf{P}}_{k+1|k} \mathbf{H}_{k+1}^T + \mathbf{R}_{k+1})^{-1} \quad (5.20)$$

$$\mathbf{v}_{k+1} = \mathbf{y}_{k+1} - \mathbf{H}_{k+1} \hat{\mathbf{x}}_{k+1|k} \quad (5.21)$$

where \mathbf{R} is the *measurement noise covariance matrix*. The innovation residual \mathbf{v} can be interpreted as the amount of new information being introduced into the system from the measurements. The importance of \mathbf{v} is that it permits us to replace the measurements by their informationally equivalent innovations. The gain matrix \mathbf{K} , on the other hand, is representative of a weighting factor indicating how much of the new information should be accepted by the system. Roughly speaking, the gain matrix weighs the innovations from the measurements to the current knowledge of the state. Derivations for Eq.(5.18) to (5.21) can be found in Mendel (1995) and Teunissen et al. (2009).

Provided the time update and the measurement update epoch by epoch, the KF works as a dynamical feedback system. The gain matrix \mathbf{K}_{k+1} and predicted- and filtering-error covariance matrices $\mathbf{P}_{k+1|k}$ and $\mathbf{P}_{k+1|k+1}$ comprise a matrix feedback system operating within the KF. After a certain value of updates over time, $\mathbf{P}_{k+1|k+1}$ and $\mathbf{K}_{k+1|k+1}$ reach limiting values. These limiting values are typically independent on $\mathbf{P}_{0|0}$, but influenced by model parameters and the measurements.

A divergence phenomenon may occur in the KF when either the process noise \mathbf{Q} or the measurement noise \mathbf{R} or both are too small. In essence, the KF locks onto the wrong values for the state, but believes them to be the true values (Mendel, 1995). For instance, in an extreme case with $\mathbf{Q} = 0$, as $k \rightarrow \infty$, the KF is rejecting new measurements because it believes the dynamic model to be the true precise model; but, of course, it is not the true model in most applications. On the other hand, a too-large \mathbf{Q} increases the uncertainty of the state estimate. The penalty for this is that the state estimate may fluctuate widely around its true value. Speaking of \mathbf{R} , its value is inversely proportional to the value of the gain matrix \mathbf{K} , meaning that a preciser measurement (small \mathbf{R}) enables a faster respond of the system. However, a too-small \mathbf{R} drives the system responding too fast based on only the current measurement, resulting in a possibility of locking onto the wrong values of the state. On the other hand, a too-large value of \mathbf{R} reduces the sensitivity of the filter to the noise and risks a long convergence time. The choice of \mathbf{Q} and \mathbf{R} is related to measures to achieve guarantees of the degree of stability of a KF.

Extended Kalman filter (EKF)

Consider that the dynamic or the measurement model or both are nonlinear functions with respect to the state \mathbf{x} ,

$$\text{Model 1: } \dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), t) + \mathbf{G}(t)\mathbf{w}(t) \quad (5.22)$$

$$\text{or Model 2: } \mathbf{x}(t) = \mathbf{f}(\mathbf{x}(t_0), t_0) + \mathbf{G}(t)\mathbf{w}(t) \quad (5.23)$$

$$\mathbf{y}(t) = \mathbf{h}(\mathbf{x}(t), t) + \mathbf{v}(t) \quad (5.24)$$

where \mathbf{f} and \mathbf{h} are nonlinear functions, which may depend both implicitly and explicitly on t , and it is assumed that both \mathbf{f} and \mathbf{h} are continuous and continuously differentiable with respect to all the elements of \mathbf{x} . Note that the dynamic model can be expressed in two forms, Eq.(5.22) and (5.23). The former is the nominal nonlinear differential equation and its discretized solution involves numerical integrations, whereas the latter represents a nonlinear state propagation directly from one time to the next and is thus simpler to implement. The availability of model (5.22) or (5.23) depends on applications.

An extended Kalman filter (EKF) is applied to “extend” the Kalman filter to nonlinear systems by linearizing \mathbf{f} and \mathbf{h} functions at the nominal state vectors $\hat{\mathbf{x}}_{k|k}$ and $\hat{\mathbf{x}}_{k+1|k}$, respectively, in the prediction and correction steps

$$\mathbf{f}(\mathbf{x}(t), t) = \mathbf{f}(\hat{\mathbf{x}}_{k|k}, t) + \mathbf{F}_{\mathbf{x}_k} \delta \mathbf{x} + \text{higher order terms} \quad (5.25)$$

$$\mathbf{h}(\mathbf{x}(t), t) = \mathbf{h}(\hat{\mathbf{x}}_{k+1|k}, t) + \mathbf{H}_{\mathbf{x}_{k+1}} \delta \mathbf{x} + \text{higher order terms} \quad (5.26)$$

where $\mathbf{F}_{\mathbf{x}_k}$ and $\mathbf{H}_{\mathbf{x}_{k+1}}$ are Jacobian matrices for p elements in the state vector and q measurements

$$\mathbf{F}_{\mathbf{x}_k} = \left. \frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}} \right|_{\mathbf{x}=\hat{\mathbf{x}}_{k|k}} = \left[\begin{array}{ccc} \frac{\partial f_1(\mathbf{x})}{\partial \mathbf{x}(1)} & \cdots & \frac{\partial f_1(\mathbf{x})}{\partial \mathbf{x}(p)} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_p(\mathbf{x})}{\partial \mathbf{x}(1)} & \cdots & \frac{\partial f_p(\mathbf{x})}{\partial \mathbf{x}(p)} \end{array} \right]_{\mathbf{x}=\hat{\mathbf{x}}_{k|k}} \quad (5.27)$$

$$\mathbf{H}_{\mathbf{x}_{k+1}} = \left. \frac{\partial \mathbf{h}(\mathbf{x})}{\partial \mathbf{x}} \right|_{\mathbf{x}=\hat{\mathbf{x}}_{k+1|k}} = \left[\begin{array}{ccc} \frac{\partial h_1(\mathbf{x})}{\partial \mathbf{x}(1)} & \cdots & \frac{\partial h_1(\mathbf{x})}{\partial \mathbf{x}(p)} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_q(\mathbf{x})}{\partial \mathbf{x}(1)} & \cdots & \frac{\partial h_q(\mathbf{x})}{\partial \mathbf{x}(p)} \end{array} \right]_{\mathbf{x}=\hat{\mathbf{x}}_{k+1|k}} \quad (5.28)$$

The EKF linearizes the nonlinear functions \mathbf{f} and \mathbf{h} about each new estimate as it becomes available at each prediction and correction step. The purpose of relinearizing about the filter’s output is to use a better reference trajectory for $\hat{\mathbf{x}}_k$. Doing this, $\delta \mathbf{x}_k = \mathbf{x}_k - \hat{\mathbf{x}}_k$ will be held as small as possible, so that the linearization assumptions are less likely to be violated.

The EKF is designed to work well as long as $\delta \mathbf{x}_k$ is “small”. The *iterated* EKF (Jazwinski, 1970; Mendel, 1995), depicted in Figure 5.3, is designed to keep $\delta \mathbf{x}_k$ as small as possible. The iterated EKF differs from the EKF in that it iterates the correction equations L times until $\|\hat{\mathbf{x}}_{k|k,L} - \hat{\mathbf{x}}_{k|k,L-1}\| \leq \epsilon$. Corrector 1 computes \mathbf{v}_k , \mathbf{K}_k and $\hat{\mathbf{P}}_{k|k}$ using $\mathbf{x} = \hat{\mathbf{x}}_{k|k-1}$; corrector 2 computes these quantities using $\mathbf{x} = \hat{\mathbf{x}}_{k|k,1}$; corrector 3 computes these quantities using $\mathbf{x} = \hat{\mathbf{x}}_{k|k,2}$; etc. This improves

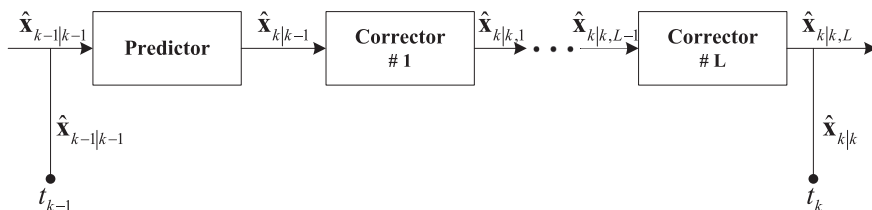


Figure 5.3: Iterated extended Kalman filter to provide a refined estimate of $\hat{\mathbf{x}}_{k|k}$ starting from $\hat{\mathbf{x}}_{k-1|k-1}$ (Mendel, 1995)

the convergence properties of the EKF since convergence is related to how close the nominal value of the state vector is to its actual value. Often, just adding one additional corrector (i.e., $L = 2$) leads to substantially better results for $\hat{\mathbf{x}}_{k|k}$ than are obtained using the non-iterated EKF (Mendel, 1995).

After initializations for $\hat{\mathbf{x}}_{0|0}$ and $\hat{\mathbf{P}}_{0|0}$, the implementation of the iterated EKF includes cycles of time updates and iterated measurement updates following each other epoch by epoch, as represented below

$$\begin{aligned}
 & \textbf{Time update (Predictor):} \\
 \text{Model 1: } \hat{\mathbf{x}}_{k+1|k} &= \hat{\mathbf{x}}_{k|k} + \int_{t_k}^{t_{k+1}} \mathbf{f}(\hat{\mathbf{x}}(t|t_k), t) dt \\
 \text{or Model 2: } \hat{\mathbf{x}}_{k+1|k} &= \mathbf{f}(\hat{\mathbf{x}}_{k|k}) \\
 \hat{\mathbf{P}}_{k+1|k} &= \mathbf{F}_{\mathbf{x}_k} \hat{\mathbf{P}}_{k|k} \mathbf{F}_{\mathbf{x}_k}^T + \mathbf{Q}_{k+1|k} \\
 & \textbf{Measurement update (Corrector):} \\
 \hat{\mathbf{x}}_{k+1|k+1,0} &= \hat{\mathbf{x}}_{k+1|k} \\
 \hat{\mathbf{P}}_{k+1|k+1,0} &= \hat{\mathbf{P}}_{k+1|k} \\
 \text{for } s = 1 : L & \\
 \mathbf{v}_{k+1} &= \mathbf{y}_{k+1} - \mathbf{h}(\hat{\mathbf{x}}_{k+1|k+1,s-1}) \\
 \mathbf{H}_{\mathbf{x}_{k+1}} &= \left. \frac{\partial \mathbf{h}(\mathbf{x})}{\partial \mathbf{x}} \right|_{\mathbf{x}=\hat{\mathbf{x}}_{k+1|k+1,s-1}} \\
 \mathbf{K}_{k+1} &= \hat{\mathbf{P}}_{k+1|k+1,s-1} \mathbf{H}_{\mathbf{x}_{k+1}}^T (\mathbf{H}_{\mathbf{x}_{k+1}} \hat{\mathbf{P}}_{k+1|k+1,s-1} \mathbf{H}_{\mathbf{x}_{k+1}}^T + \mathbf{R}_{k+1})^{-1} \\
 \hat{\mathbf{x}}_{k+1|k+1,s} &= \hat{\mathbf{x}}_{k+1|k+1,s-1} + \mathbf{K}_{k+1} \mathbf{v}_{k+1} \\
 \hat{\mathbf{P}}_{k+1|k+1,s} &= (\mathbf{I} - \mathbf{K}_{k+1} \mathbf{H}_{\mathbf{x}_{k+1}}) \hat{\mathbf{P}}_{k+1|k+1,s-1} \\
 \text{end} & \\
 \hat{\mathbf{x}}_{k+1|k+1} &= \hat{\mathbf{x}}_{k+1|k+1,L} \\
 \hat{\mathbf{P}}_{k+1|k+1} &= \hat{\mathbf{P}}_{k+1|k+1,L} .
 \end{aligned} \tag{5.29}$$

5.2.2 Model of the satellite-antenna-reflector geometry

For the multipath estimation problem, it is clear that both the multipath error and the SNR measurement are functions of the sine or cosine waveforms of the multipath phase ψ . It is of great importance to understand the time-dependency of ψ in a specific satellite-antenna-reflector geometry.

Figure 5.4 illustrates a generalized geometry. An extra path length $2l \sin(\beta)$ is travelled by the multipath signal with respect to the direct signal, where l denotes the perpendicular antenna-reflector distance and β is the angle of reflection. Since the reflector is tilted at angle γ , $\beta = el - \gamma$ with el being the satellite elevation angle. Translating the extra path length in meters to the carrier phase in rads, ψ is

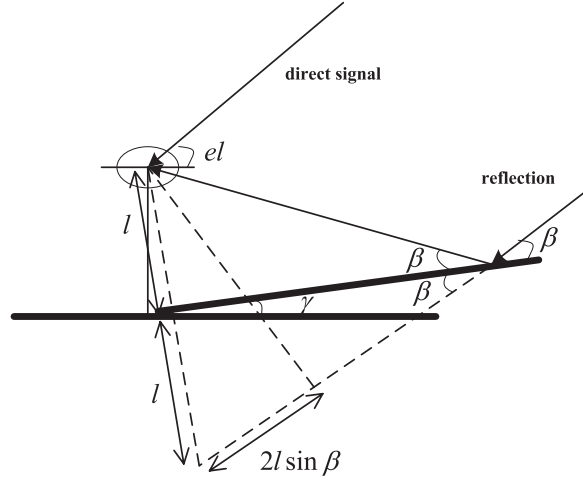


Figure 5.4: Geometry of a multipath reflection from a planar surface tilted at angle γ and located at a perpendicular distance l from the antenna phase center (Bilich et al., 2008)

obtained

$$\begin{aligned}\psi &= \frac{2\pi}{\lambda} 2l \sin(\beta) \\ &= \frac{2\pi}{\lambda} 2l \sin(el - \gamma)\end{aligned}\quad (5.30)$$

where λ is the carrier signal wavelength. By assuming the reflector persists so that γ and l are constant, the only time-dependent factor remaining in Eq.(5.30) is the satellite elevation angle, which changes according to the relative motions between the satellite and the receiver. By taking the time-derivative of ψ , the multipath frequency ω is equal to

$$\omega = \frac{d\psi}{dt} = \frac{2\pi}{\lambda} 2l \cos(el - \gamma) \frac{del}{dt}\quad (5.31)$$

If we regard $\beta = el - \gamma$ from 0° to 90° as a complete satellite pass, the multipath frequency declines from its maximum to 0 in this pass. The ability to estimate the multipath frequency via the FFT/wavelet in the literature is significantly determined by the number of multipath cycles which can be contained in the range of a complete satellite pass (Bilich, 2006). For an increasing antenna-reflector distance l , an increasing number of complete cycles can be observed in the SNR data.

In a closely-spaced multi-antenna scenario, since signals arrived at different antennas are parallel, the angles of reflection β of different multipath signals on different antennas are the same, provided they are reflected from the same reflector. Multipath phases on different antennas as well as of different frequencies will then be strongly correlated

$$\begin{aligned}\psi_{i,f_1} &= \psi_{1,f_1} + 2\pi \left(\frac{2l_i}{\lambda_1} - \frac{2l_1}{\lambda_1} \right) \sin(\beta), \quad i = 2, \dots, n \\ \psi_{i,f_2} &= \psi_{1,f_1} + 2\pi \left(\frac{2l_i}{\lambda_2} - \frac{2l_1}{\lambda_1} \right) \sin(\beta), \quad i = 1, 2, \dots, n\end{aligned}\quad (5.32)$$

where ψ_{i,f_j} denotes the multipath phase of the carrier j on antenna i , λ_1 and λ_2 denote the carrier wavelengths of frequency 1 and 2. Regarding ψ_{1,f_1} as the multipath phase on the reference antenna of frequency 1, the multipath phases on other auxiliary antennas and other frequencies can all be expressed in functions of ψ_{1,f_1} and β given *a-priori* knowledge of the antenna-reflector distance l_i .

Recall that both the multipath error of Eq.(5.1) and the SNR profile of Eq.(5.2) are dependent on the multipath phase ψ as well as the relative amplitude coefficient α . Provided all the closely-spaced antennas have similar gain patterns, the direct signals are equally amplified and the reflected signals are equally attenuated by each antenna during the signal reception. Therefore, α can be assumed to be the same on each antenna. It should be noted that antennas have different gain patterns for different frequencies, thus α_{f_1} and α_{f_2} are used to denote two distinct amplitude coefficients of frequency 1 and 2, while A_{d,f_1} and A_{d,f_2} represent two distinct direct signal amplitudes on these two frequencies. Substituting Eq.(5.32) to (5.2), SNRs are rewritten as functions of ψ_{1,f_1} , β , α_{f_1} (or α_{f_2}) and A_{d,f_1} (or A_{d,f_2})

$$\begin{aligned} SNR_{1,f_1} &= A_{d,f_1} \sqrt{1 + \alpha_{f_1}^2 + 2\alpha_{f_1} \cos(\psi_{1,f_1})} & (5.33) \\ SNR_{i,f_1} &= A_{d,f_1} \sqrt{1 + \alpha_{f_1}^2 + 2\alpha_{f_1} \cos(\psi_{1,f_1} + 2\pi(2l_i/\lambda_1 - 2l_1/\lambda_1) \sin(\beta))} \\ SNR_{i,f_2} &= A_{d,f_2} \sqrt{1 + \alpha_{f_2}^2 + 2\alpha_{f_2} \cos(\psi_{1,f_1} + 2\pi(2l_i/\lambda_2 - 2l_1/\lambda_1) \sin(\beta))}. \end{aligned}$$

To reveal the SNR oscillations due to multipath, A_{d,f_1} and A_{d,f_2} shall be removed by taking the ratio of SNRs between pairs of antennas

$$\frac{SNR_{i,f_1}}{SNR_{1,f_1}}, \frac{SNR_{i,f_2}}{SNR_{1,f_2}}, i = 2, \dots, n. \quad (5.34)$$

By treating ratios of SNRs as measurements, extended Kalman filters (real and complex) are designed in the following to estimate multipath parameters of ψ_{1,f_1} , β , α_{f_1} and α_{f_2} . Before elaborating this, we divide a complete multipath correction procedure into several steps in the next section to clarify the preparation and refinement before and after the estimation.

5.2.3 Multipath correction procedure

A complete multipath correction procedure is shown in Figure 5.5. It has been split into three cascaded EKFs: the first EKF is used to filter out the noise on ratios of SNRs before they can be treated as observations; the second successive EKF is used to estimate multipath parameters, which are then reformulated to construct the multipath errors in order to remove these errors from the phase measurements; the integer ambiguity resolution is accelerated due to the multipath removal; after ambiguities are fixed, a combined LOS, LOS rate and multipath estimator will be applied in the third cascaded EKF to guarantee the achievement of mm-order LOS accuracy in the end. Here, it should be noted that the EKFs mentioned in this chapter all utilize real-valued numbers. The chapter will also propose an extended complex-valued Kalman filter (ECKF) as the substitute in the second and third cascaded real-valued EKFs. The performance of EKF and ECKF will also be discussed.

More specifically, we start the estimation procedure from the noise filtering on ratios of SNRs. Consider Gaussian random variables R and S , the ratio R/S is not

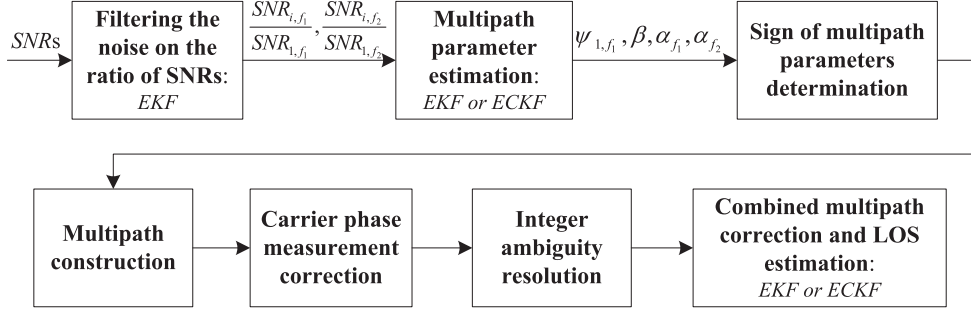


Figure 5.5: Multipath estimation procedure

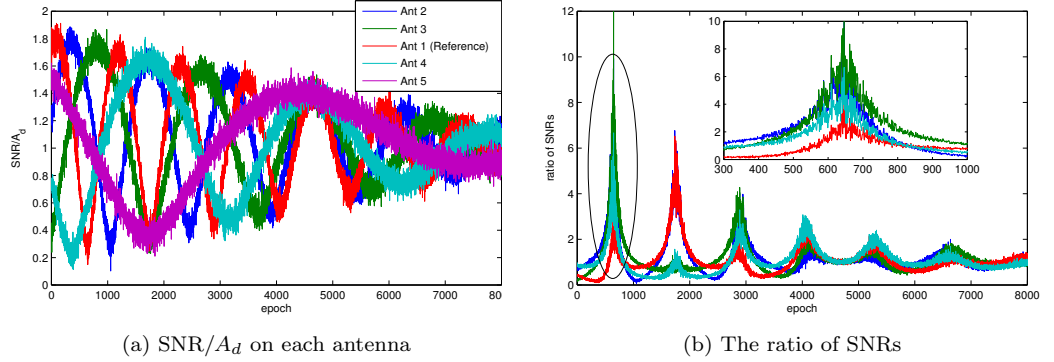


Figure 5.6: The ratio of SNRs. The simulation condition is: the multipath amplitude coefficient α is changing from 0.8 to 0.1 as the satellite elevation increasing from 5° to 45° , the noise on the SNR/A_d has a standard derivation of 0.05, and the perpendicular distance between antennas and the reflector is 1 m (reference), 0.8 m, 0.6 m, 0.4 m and 0.2 m, respectively.

Gaussian random any more. The approximation for the variance of R/S is equal to (Seltman, 2012)

$$\text{Var}\left(\frac{R}{S}\right) \approx \frac{E^2(R)}{E^2(S)} \left[\frac{\text{Var}(R)}{E^2(R)} - 2 \frac{\text{Cov}(R, S)}{E(R)E(S)} + \frac{\text{Var}(S)}{E^2(S)} \right] \quad (5.35)$$

where $E(R)$, $E(S)$ and $\text{Cov}(R, S)$ are expectations and covariance of R and S . This equation clearly shows that the variance of a ratio is not only dependent on the variance of two variables, but also strongly correlated to their mean values. For an increasing ratio, an extremely increased noise can be expected.

Figure 5.6 simulates a single-frequency SNR data on five antennas when the perpendicular distance between antennas and the reflector is assumed 1 m (reference), 0.8 m, 0.6 m, 0.4 m and 0.2 m, respectively. As the satellite elevation increases from 5° to 45° , we assume the multipath relative amplitude coefficient α decreases from 0.8 to 0.1, resulting in a magnitude decreased SNR oscillation, as depicted in Figure 5.6(a). The number of oscillation cycles in the observation period depends on the perpendicular distance between the antenna and the reflector. For an increasing antenna-reflector distance, an increasing number of complete cycles can be observed in the SNR data. The noise on each SNR data is assumed Gaussian distributed random noise with a standard derivation of 0.05. However, according to Eq.(5.35),

Table 5.1: Possible combinations of the signs of α_{f_j} , $\cos(\psi_{1,f_j})$, $\cos(\vartheta_i)$, $\sin(\psi_{1,f_j})$ and $\sin(\vartheta_i)$. Compared to the true sign, “+” indicates a correct match, while “-” indicates an opposite match (180° shifted).

Possible combinations	α_{f_j}	$\cos(\psi_{1,f_j})$	$\cos(\vartheta_i)$	$\sin(\psi_{1,f_j})$	$\sin(\vartheta_i)$
1	+	+	+	+	+
2	+	+	+	-	-
3	-	-	+	+	-
4	-	-	+	-	+

the noise on the ratios of SNRs between antennas is not Gaussian random any more, as demonstrated in Figure 5.6(b). For the moment when the ratio has a large value, the noise on the ratio is also extremely large. This enlarged noise should be filtered out before the ratios of SNRs can be treated as measurements in the subsequent EKF. The method of filtering these noises is proposed in section 5.2.4.

After the noise filtering for the ratios of SNRs, a cascaded EKF or ECKF will be built for the estimation of multipath parameters ψ_{i,f_1} , β , α_{f_1} and α_{f_2} . This process will be elaborated in section 5.2.5.

Since the oscillation in SNRs is driven by the multiplication of α_{f_j} and $\cos(\psi_{i,f_j})$ ($i = 1, \dots, n$, $j = 1, 2$), signs of them are ambiguous. In addition, the estimation for β can also result in an ambiguous sign for the extra multipath phase $\vartheta_{ij} = 2\pi(2l_i/\lambda_{f_j} - 2l_1/\lambda_{f_j}) \sin(\beta)$ of the i th antenna with respect to the reference antenna. This can be explained by formulating the equation for the ratio of SNRs

$$\psi_{i,f_j} = \psi_{1,f_j} + \underbrace{2\pi(2l_i/\lambda_j - 2l_1/\lambda_j) \sin(\beta)}_{\vartheta_{ij}} \quad (5.36)$$

$$\begin{aligned} \frac{SNR_{i,f_j}}{SNR_{1,f_j}} &= \sqrt{\frac{1 + \alpha_{f_j}^2 + 2\alpha_{f_j} \cos(\psi_{1,f_1} + \vartheta_{ij})}{1 + \alpha_{f_j}^2 + 2\alpha_{f_j} \cos(\psi_{1,f_j})}} \\ &= \sqrt{\frac{1 + \alpha_{f_j}^2 + 2\alpha_{f_j} [\cos(\psi_{1,f_j}) \cos(\vartheta_{ij}) - \sin(\psi_{1,f_j}) \sin(\vartheta_{ij})]}{1 + \alpha_{f_j}^2 + 2\alpha_{f_j} \cos(\psi_{1,f_j})}}. \end{aligned} \quad (5.37)$$

From Eq.(5.38), the signs of α_{f_j} , $\cos(\psi_{1,f_j})$, $\cos(\vartheta_{ij})$, $\sin(\psi_{1,f_j})$ and $\sin(\vartheta_{ij})$ have several possible combinations in Table 5.1, which all result in the same value of $SNR_{i,f_j}/SNR_{1,f_j}$. All these combinations have been observed in the demonstration in section 5.3. An incorrect determination of the sign will yield an inverted phase correction profile, essentially doubling the potential multipath error instead of removing it. A simple solution is used here by checking whether the estimated ψ_{1,f_j} and β is correctly ascending or descending as a function of time. The signs of $\sin(\psi_{1,f_j})$ and $\sin(\vartheta_{ij})$ persist for ascending ψ_{1,f_j} and β , while they should be oppositely signed in the descending case.

With the correctly determined signs of α_{f_j} , $\cos(\psi_{1,f_j})$, $\cos(\vartheta_i)$, $\sin(\psi_{1,f_j})$ and $\sin(\vartheta_i)$, the multipath error at each antenna can be constructed and removed from the phase measurement. Once the phase measurement is corrected, the integer ambiguity embedded in the phase measurement will be easier to resolve. This will be verified in section 5.4.

The next and final procedure for the multipath correction combines with the estimation of the relative positioning vectors, i.e, the LOS vector, in the EKF or ECKF.

In addition to state variables needed for the multipath correction, the LOS vector $\mathbf{x}_{LOS} = [x, y, z]^T$ and the LOS rate vector $\dot{\mathbf{x}}_{LOS} = [\dot{x}, \dot{y}, \dot{z}]^T$ will also be included in the state. This is possible because after the ambiguities are fixed, the single differenced (SD) phase measurement between two antennas contains only the true range difference (as a function of \mathbf{x}_{LOS}) and the difference of multipath and random phase noise. Multipath parameters will be better estimated using additional SD phase measurements rather than the stand-alone SNR measurements. The steps for determining signs and constructing multipath can also be avoided. More explanations and derivations will be given in section 5.2.6.

5.2.4 Noise filtering for ratios of SNRs

Taking the ratios of SNRs and the SNR on the reference antenna as state variables in an EKF

$$\begin{aligned} \mathbf{x} &= \left[\mathbf{x}(1) \quad \mathbf{x}(2) \quad \cdots \quad \mathbf{x}(n) \quad \mathbf{x}(n+1) \quad \mathbf{x}(n+2) \quad \cdots \quad \mathbf{x}(2n) \right]^T \\ &= \left[\begin{array}{c|ccc} SNR_{1,f_1} & \frac{SNR_{2,f_1}}{SNR_{1,f_1}} & \cdots & \frac{SNR_{n,f_1}}{SNR_{1,f_1}} \\ \hline SNR_{1,f_2} & \frac{SNR_{2,f_2}}{SNR_{1,f_2}} & \cdots & \frac{SNR_{n,f_2}}{SNR_{1,f_2}} \end{array} \right]^T \end{aligned} \quad (5.38)$$

where n is the number of antennas. The discretized state propagation in the time update from t_k to t_{k+1} can be simply written as

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{s}_k \quad (5.39)$$

$$\mathbf{Q}_{k+1|k} = \text{diag}(Q_{s_1}, Q_{s_2}, \cdots, Q_{s_{2n}}) \quad (5.40)$$

where $\mathbf{s}_k = [s_1, \cdots, s_{2n}]^T$ accounts for the allowance on the speed of the state propagation, which values depend on how fast of the change for each state variable is expected in an update interval, $\mathbf{Q}_{k+1|k}$ denotes the discretized process noise with $Q_{s_1}, \cdots, Q_{s_{2n}}$ as the variance of s_1, \cdots, s_{2n} .

The observations include all SNRs with Gaussian random noise

$$\begin{bmatrix} SNR_{1,f_1} \\ SNR_{2,f_1} \\ \vdots \\ SNR_{n,f_1} \\ SNR_{1,f_2} \\ SNR_{2,f_2} \\ \vdots \\ SNR_{n,f_2} \end{bmatrix} = \begin{bmatrix} \mathbf{x}(1) \\ \mathbf{x}(1)\mathbf{x}(2) \\ \vdots \\ \mathbf{x}(1)\mathbf{x}(n) \\ \mathbf{x}(n+1) \\ \mathbf{x}(n+1)\mathbf{x}(n+2) \\ \vdots \\ \mathbf{x}(n+1)\mathbf{x}(2n) \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_{SNR_{f_1}} \\ \boldsymbol{\varepsilon}_{SNR_{f_2}} \end{bmatrix}. \quad (5.41)$$

As can be seen, this measurement model is nonlinear and shall be linearized with respect to each state variable by partial derivations once a new estimate is available after the time update.

The implementation of this filter follows the steps in Eq.(5.29). Several iterations of relinearization in the measurement update (iterated EKF) are used to enable fast convergence. This is important since the outputs of this filter will be treated as observations and cascaded to another filter for the multipath parameter estimation. A longer convergence time means a waste of observations.

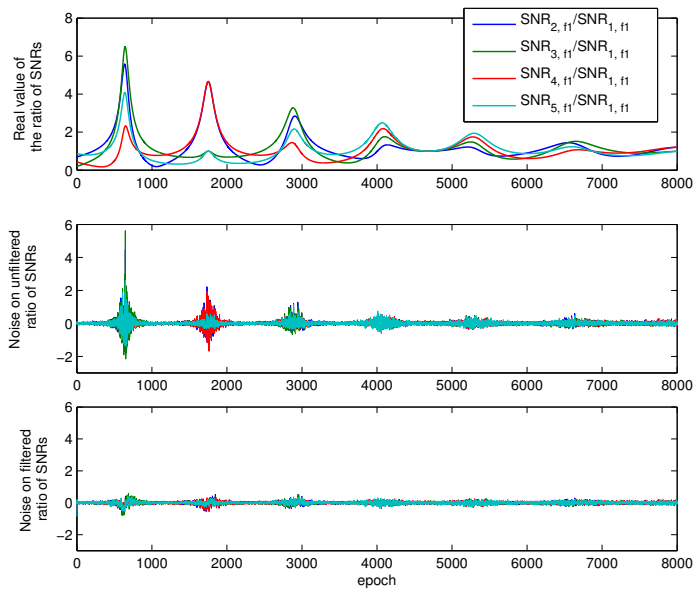


Figure 5.7: The noise on the ratio of SNRs before and after EKF. The SNR measurement has a standard deviation of 0.05, and each state variable has identical process noise of 0.01.

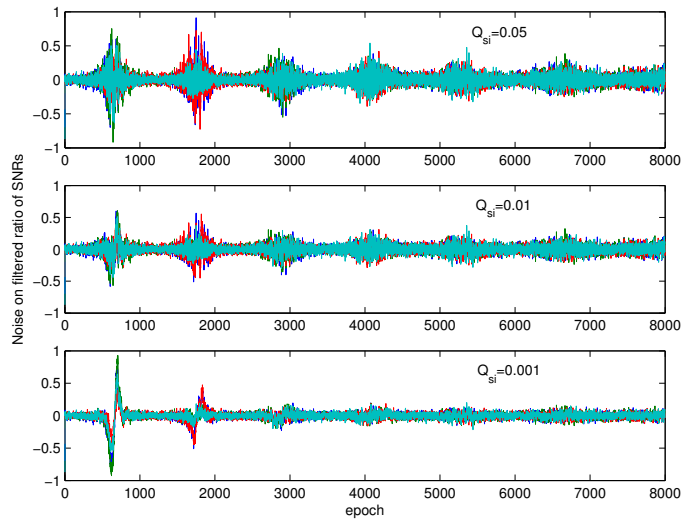


Figure 5.8: The noise on the ratio of SNRs after the EKF given different levels of process noise

The noise on the ratio of SNRs before and after filtered in the EKF is examined and illustrated in Figure 5.7. Single frequency SNR data on five antennas (see Figure 5.6) have been used for this simulation. It is clear that the noise on the ratio of SNRs has been significantly reduced after implementing the EKF, especially for those epochs where the ratio has a large real value. Here, each state variable is assumed having an identical and uncorrected process noise of 0.01. The performance

of using different levels of process noise in the filter is also illustrated in Figure 5.8. We can see that the noise on large ratios is larger than the noise on small ratios. Decreasing the process noise helps to reduce the variance of the noise. However, over-decreasing it will enlarge the absolute value of the noise around the large ratio points, which is due to the fact that the given process noise is too small to account for the large change at the large ratio points.

5.2.5 Multipath correction before fixing integer ambiguities

Revealing the non-stationary amplitude, frequency and phase simultaneously from the sinusoid-driven oscillatory data, like the ratio of SNRs, is a non-trivial task. Once both the amplitude α and phase ψ in a sinusoidal element $\alpha \sin \psi$ are non-stationary (having a non-zero process noise), it is ambiguous for the KF to distinguish whether α or $\sin \psi$ or both causes the oscillation of the observation. The KF can easily lock onto the wrong state and is unaware that the true error variance is diverging. This is why the frequency is usually pre-estimated by FFT and fed into the KF as the *a-priori* parameter in the literature (Axelrad et al., 1996; Bilich et al., 2008). Alternately, having distinct observations with few overlaps will aid in distinguishing the real origin of the oscillation in observations, i.e, using more antennas in this multipath estimation problem. It is also beneficial to constrain specific state variables after the time update to clarify the range of their variations, i.e, using the equality constraint on the norm of $\exp(j\psi)$ and $\exp(-j\psi)$ if they are regarded as state variables instead of ψ , and/or using the inequality constraint on α such as $0 \leq \alpha < 1$, which is valid in this application as the multipath amplitude is always smaller than the direct signal amplitude.

In the following, both extended real-valued and complex-valued Kalman filters are proposed for the multipath parameter estimation. Their performance comparison is presented afterwards.

Using extended real-valued Kalman filter

As ratios of SNRs on dual frequencies are functions of ψ_{1,f_1} , β , α_{f_1} and α_{f_2} , let the state in the real-valued EKF be

$$\mathbf{x} = [\psi_{1,f_1} \quad \omega_1 \quad \beta \quad \omega_2 \quad \alpha_{f_1} \quad \alpha_{f_2}]^T \quad (5.42)$$

where ω_1 and ω_2 donate the associated angular frequencies for the phase ψ_{1,f_1} and β , respectively.

Since $\psi_1 = \omega_1 t$ and $\beta = \omega_2 t$, the continuous-time dynamic model of the state can be written as

$$\begin{bmatrix} \dot{\psi}_{1,f_1} \\ \dot{\omega}_1 \\ -\dot{\beta} \\ \dot{\omega}_2 \\ -\dot{\alpha}_{f_1} \\ \dot{\alpha}_{f_2} \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}}_{\mathbf{F}(t)} \begin{bmatrix} \psi_{1,f_1} \\ \omega_1 \\ \beta \\ \omega_2 \\ \alpha_{f_1} \\ \alpha_{f_2} \end{bmatrix} + \begin{bmatrix} 0 \\ -s\dot{\omega}_1 \\ 0 \\ -s\dot{\omega}_2 \\ s\dot{\alpha}_{f_1} \\ s\dot{\alpha}_{f_2} \end{bmatrix} \quad (5.43)$$

with the continuous-time process noise

$$\mathbf{Q}(t) = \text{diag}(0, Q_{s\dot{\omega}_1}, 0, Q_{s\dot{\omega}_2}, Q_{s\dot{\alpha}_{f_1}}, Q_{s\dot{\alpha}_{f_2}}) \quad (5.44)$$

where $diag()$ means diagonal matrix, $Q_{s\dot{\omega}_1}$, $Q_{s\dot{\omega}_2}$, $Q_{s\dot{\alpha}_{f_1}}$ and $Q_{s\dot{\alpha}_{f_2}}$ account for the variance of the system noise on $\dot{\omega}_1$, $\dot{\omega}_2$, $\dot{\alpha}_{f_1}$ and $\dot{\alpha}_{f_2}$, respectively. The system noises on $\dot{\psi}_{1,f_1}$ and $\dot{\beta}$ are zeros since they have determinate dynamic relations with respect to state variables ω_1 and ω_2 , that is, $\dot{\psi} = \dot{\omega}_1$ and $\dot{\beta} = \dot{\omega}_2$.

According to the derivation in section 5.2.1, when $\mathbf{F}(t)$ and $\mathbf{Q}(t)$ are constant during the time interval $[t_k, t_{k+1}]$, the state transition matrix $\Phi_{k+1|k}$ can be simply calculated as

$$\begin{aligned} \Phi_{k+1|k} &= e^{\mathbf{F}\Delta t} = \sum_{i=0}^{\infty} \frac{\mathbf{F}^i \Delta t^i}{i!} \\ &= \begin{bmatrix} 1 & \Delta t & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & \Delta t & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{aligned} \quad (5.45)$$

and the discrete-time process noise $\mathbf{Q}_{k+1|k}$ is calculated as

$$\begin{aligned} \mathbf{Q}_{k+1|k} &= \int_{t_k}^{t_{k+1}} \int_{t_k}^{t_{k+1}} \Phi(t_{k+1}, \tau) \mathbf{Q}(\tau) \Phi^T(t_{k+1}, \tau) d\tau d\tau \\ &= \begin{bmatrix} \Delta t^4 Q_{s\dot{\omega}_1}/12 & \Delta t^3 Q_{s\dot{\omega}_1}/6 & 0 & 0 & 0 & 0 \\ \Delta t^3 Q_{s\dot{\omega}_1}/6 & \Delta t^2 Q_{s\dot{\omega}_1}/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \Delta t^4 Q_{s\dot{\omega}_2}/12 & \Delta t^3 Q_{s\dot{\omega}_2}/6 & 0 & 0 \\ 0 & 0 & \Delta t^3 Q_{s\dot{\omega}_2}/6 & \Delta t^2 Q_{s\dot{\omega}_2}/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \Delta t^2 Q_{s\dot{\alpha}_{f_1}}/2 & 0 \\ 0 & 0 & 0 & 0 & 0 & \Delta t^2 Q_{s\dot{\alpha}_{f_2}}/2 \end{bmatrix} \end{aligned} \quad (5.46)$$

Regarding the measurement model, it can be easily written by substituting Eq.(5.33) into (5.34). By partial derivations to each measurement equation with respect to each state variable, the measurement model can be linearized to the first-order and substituted to the routine of the iterated EKF of Eq.(5.29) for implementation.

Using extended complex-valued Kalman filter

In the above real valued EKF, the system dynamic is linear while the observation of the state is nonlinear. Moreover, this nonlinearity is strong as the sinusoidal element in the observation function has infinite Taylor series. This could cause a difficulty for the EKF to converge with poorly initial conditions (Nishiyama, 1997). A new solution using the extended complex Kalman filter (ECKF) is reported to be more attractive than the real Kalman filter in terms of the stability and reliability (Nishiyama, 1997; Dash et al., 2000). A unit norm constraint to the complex variable can also be added in this ECKF to assist in the convergence of the filter.

In an ECKF, a sine or cosine waveform is expressed by the sum of two conjugate complex signals

$$\sin(\psi) = -0.5j \exp(j\psi) + 0.5j \exp(-j\psi) \quad (5.47)$$

$$\cos(\psi) = 0.5 \exp(j\psi) + 0.5 \exp(-j\psi) \quad (5.48)$$

where $\psi = \omega t$. Since the propagation of ψ from one time to the next satisfies $\psi_k = \psi_{k-1} + w\Delta t$, this allows $\exp(j\psi_k)$ and $\exp(-j\psi_k)$ being modelled as an *autoregressive* (AR) process which yields the output variable based linearly on its own previous values

$$\exp(j\psi_k) = \exp(jw\Delta t) \exp(j\psi_{k-1}) \quad (5.49)$$

$$\exp(-j\psi_k) = \exp(-jw\Delta t) \exp(-j\psi_{k-1}) \quad (5.50)$$

where $\exp(jw\Delta t)$ is called an AR coefficient. By taking $\exp(jw\Delta t)$, $\exp(j\psi)$ and $\exp(-j\psi)$ as state variables, the dynamics for propagating the phase are easy to build.

Now we look at the multi-antenna based multipath estimation problem. Let $\vartheta = 2\pi/\lambda_1(2l_2 - 2l_1) \sin(\beta)$, multipath phases in Eq.(5.32) on different antennas and different frequencies can then be rewritten as

$$\begin{aligned} \psi_{2,f_1} &= \psi_{1,f_1} + \underbrace{2\pi/\lambda_1(2l_2 - 2l_1) \sin(\beta)}_{\vartheta} \\ \psi_{i,f_1} &= \psi_{1,f_1} + 2\pi/\lambda_1(2l_i - 2l_1) \sin(\beta) \\ &= \psi_{1,f_1} + \left(\frac{l_i - l_1}{l_2 - l_1}\right) \vartheta, \quad i = 3, \dots, n \\ \psi_{i,f_2} &= \psi_{1,f_1} + 2\pi(2l_i/\lambda_2 - 2l_1/\lambda_1) \sin(\beta) \\ &= \psi_{1,f_1} + \left(\frac{l_i/\lambda_2 - l_1/\lambda_1}{l_2/\lambda_1 - l_1/\lambda_1}\right) \vartheta, \quad i = 1, \dots, n. \end{aligned} \quad (5.51)$$

Let the state variables in the ECKF be

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}(1) \\ \mathbf{x}(2) \\ \mathbf{x}(3) \\ \mathbf{x}(4) \\ \mathbf{x}(5) \\ \mathbf{x}(6) \\ \mathbf{x}(7) \\ \mathbf{x}(8) \end{bmatrix} = \begin{bmatrix} \exp(jw_1\Delta t) \\ \exp(j\psi_{1,f_1}) \\ \exp(-j\psi_{1,f_1}) \\ \exp(jw_2\Delta t) \\ \exp(j\vartheta) \\ \exp(-j\vartheta) \\ \alpha_{f_1} \\ \alpha_{f_2} \end{bmatrix} \quad (5.52)$$

where w_1 and w_2 now denote angular frequencies with respect to phases of ψ_{1,f_1} and ϑ , respectively.

The discrete-time state propagation can be written according to the AR process in Eq.(5.49) and (5.50)

$$\mathbf{x}_{k+1|k} = \mathbf{f}(\mathbf{x}_k) = \begin{bmatrix} \mathbf{x}_k(1) \\ \mathbf{x}_k(1)\mathbf{x}_k(2) \\ \mathbf{x}_k(3)/\mathbf{x}_k(1) \\ \mathbf{x}_k(4) \\ \mathbf{x}_k(4)\mathbf{x}_k(5) \\ \mathbf{x}_k(6)/\mathbf{x}_k(4) \\ \mathbf{x}_k(7) \\ \mathbf{x}_k(8) \end{bmatrix} + \begin{bmatrix} s_{x_1} \\ 0 \\ 0 \\ s_{x_4} \\ 0 \\ 0 \\ s_{x_7} \\ s_{x_8} \end{bmatrix} \quad (5.53)$$

$$\mathbf{Q}_{k+1|k} = \text{diag}(Q_{s_{x_1}}, \mathbf{0}_{2 \times 2}, Q_{s_{x_4}}, \mathbf{0}_{2 \times 2}, Q_{s_{x_7}}, Q_{s_{x_8}}) \quad (5.54)$$

where $Q_{s_{x_1}}$ and $Q_{s_{x_4}}$ are accounted for the variations of AR coefficients $\exp(jw_1\Delta t)$ and $\exp(jw_2\Delta t)$ caused by the variations on frequencies w_1 and w_2 , and $Q_{s_{x_7}}$ and $Q_{s_{x_8}}$ represent amplitude variations. The process noise on discrete state variables $\mathbf{x}_{k+1|k}(2)$, $\mathbf{x}_{k+1|k}(3)$, $\mathbf{x}_{k+1|k}(5)$ and $\mathbf{x}_{k+1|k}(6)$ is zero as the AR expressions in Eq.(5.49) and (5.50) are determinate.

Linearizing the dynamic model to the first order, we will get

$$\mathbf{F}_{\mathbf{x}_k} = \left. \frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}} \right|_{\mathbf{x}=\hat{\mathbf{x}}_{k|k}} \quad (5.55)$$

$$= \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hat{\mathbf{x}}_{k|k}(2) & \hat{\mathbf{x}}_{k|k}(1) & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{\hat{\mathbf{x}}_{k|k}(3)}{(\hat{\mathbf{x}}_{k|k}(1))^2} & 0 & \frac{1}{\hat{\mathbf{x}}_{k|k}(1)} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \hat{\mathbf{x}}_{k|k}(5) & \hat{\mathbf{x}}_{k|k}(4) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{\hat{\mathbf{x}}_{k|k}(6)}{(\hat{\mathbf{x}}_{k|k}(4))^2} & 0 & \frac{1}{\hat{\mathbf{x}}_{k|k}(4)} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

In order to keep the norm of the complex state variables equal to 1, normalization after a complete prediction is made

$$\hat{\mathbf{x}}_{k+1|k}(i) = \frac{\hat{\mathbf{x}}_{k+1|k}(i)}{\|\hat{\mathbf{x}}_{k+1|k}(i)\|}, \quad i = 1, \dots, 6. \quad (5.56)$$

In the measurement model, the SNR can be reformulated as a function of \mathbf{x}

$$\begin{aligned} (SNR_{1,f_1}/A_{d,f_1})^2 &= 1 + \alpha_{f_1}^2 + 2\alpha_{f_1} \cos(\psi_{1,f_1}) \\ &= 1 + \alpha_{f_1}^2 + \alpha_{f_1} [\exp(j\psi_{1,f_1}) + \exp(-j\psi_{1,f_1})] \\ &= 1 + \mathbf{x}(7)^2 + \mathbf{x}(7)(\mathbf{x}(1) + \mathbf{x}(2)) \\ (SNR_{2,f_1}/A_{d,f_1})^2 &= 1 + \alpha_{f_1}^2 + 2\alpha_{f_1} \cos(\psi_{1,f_1} + \vartheta) \\ &= 1 + \alpha_{f_1}^2 + \alpha_{f_1} [\exp(j(\psi_{1,f_1} + \vartheta)) + \exp(-j(\psi_{1,f_1} + \vartheta))] \\ &= 1 + \alpha_{f_1}^2 + \alpha_{f_1} [\exp(j\psi_{1,f_1}) \exp(j\vartheta) + \exp(-j\psi_{1,f_1}) \exp(-j\vartheta)] \\ &= 1 + \mathbf{x}(7)^2 + \mathbf{x}(7)[\mathbf{x}(2)\mathbf{x}(5) + \mathbf{x}(3)\mathbf{x}(6)] \\ (SNR_{i,f_1}/A_{d,f_1})^2 &= 1 + \alpha_{f_1}^2 + 2\alpha_{f_1} \cos(\psi_{1,f_1} + \frac{l_i - l_1}{l_2 - l_1} \vartheta) \\ &= 1 + \alpha_{f_1}^2 + \alpha_{f_1} [\exp(j(\psi_{1,f_1} + \frac{l_i - l_1}{l_2 - l_1} \vartheta)) \\ &\quad + \exp(-j(\psi_{1,f_1} + \frac{l_i - l_1}{l_2 - l_1} \vartheta))] \\ &= 1 + \mathbf{x}(7)^2 + \mathbf{x}(7)[\mathbf{x}(2)\mathbf{x}(5)^{\frac{l_i - l_1}{l_2 - l_1}} + \mathbf{x}(3)\mathbf{x}(6)^{\frac{l_i - l_1}{l_2 - l_1}}], \quad i = 3, \dots, n \\ (SNR_{i,f_2}/A_{d,f_2})^2 &= 1 + \alpha_{f_2}^2 + 2\alpha_{f_2} \cos(\psi_{1,f_1} + \frac{l_i/\lambda_2 - l_1/\lambda_1}{l_2/\lambda_1 - l_1/\lambda_1} \vartheta) \\ &= 1 + \mathbf{x}(8)^2 + \mathbf{x}(8)[\mathbf{x}(2)\mathbf{x}(5)^{\frac{l_i/\lambda_2 - l_1/\lambda_1}{l_2/\lambda_1 - l_1/\lambda_1}} + \mathbf{x}(3)\mathbf{x}(6)^{\frac{l_i/\lambda_2 - l_1/\lambda_1}{l_2/\lambda_1 - l_1/\lambda_1}}], \\ &\quad i = 1, \dots, n. \end{aligned} \quad (5.57)$$

Often, the state variables in a Kalman filter are chosen uncorrelated. However, $\mathbf{x}(2) = \exp(j\psi_{1,f_1})$ and $\mathbf{x}(3) = \exp(-j\psi_{1,f_1})$ are conjugated with full correlations,

so as the state variables of $\mathbf{x}(5)$ and $\mathbf{x}(6)$. Thus, apart from treating the ratio of SNRs as the observation, two extra pseudo-observations that satisfy $\mathbf{x}(2)\mathbf{x}(3) = 1$ and $\mathbf{x}(5)\mathbf{x}(6) = 1$ are also added in order to link their correlations

$$\mathbf{y} = \mathbf{h}(\mathbf{x}) = \begin{bmatrix} h_1(\mathbf{x}) \\ \vdots \\ \frac{h_{2n-1}(\mathbf{x})}{\bar{h}_n(\mathbf{x})} \\ \vdots \\ \frac{h_{2n-2}(\mathbf{x})}{\bar{h}_{2n-1}(\mathbf{x})} \\ h_{2n}(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} SNR_{2,f_1}/SNR_{1,f_1} \\ \vdots \\ \frac{SNR_{n,f_1}/SNR_{1,f_1}}{SNR_{2,f_2}/SNR_{1,f_2}} \\ \vdots \\ \frac{SNR_{n,f_2}/SNR_{1,f_2}}{1} \\ 1 \end{bmatrix} + \begin{bmatrix} \mathbf{v}_{f_1} \\ \text{---} \\ \mathbf{v}_{f_2} \\ \text{---} \\ \mathbf{v}^p \end{bmatrix} \quad (5.58)$$

where h_{2n-1} and h_{2n} are those two pseudo-observations with the noise of $\mathbf{v}^p = [v^{p1}, v^{p2}]^T$

$$\begin{aligned} h_{2n-1} &= \mathbf{x}(2)\mathbf{x}(3) + v^{p1} \\ h_{2n} &= \mathbf{x}(5)\mathbf{x}(6) + v^{p2}. \end{aligned} \quad (5.59)$$

If introducing two perfect pseudo-observations without noise, a singular estimation problem or the filter divergence may occur due to these noise-free measurements. A small noise can be artificially added to address this singularity and avoid divergence.

After multipath parameters are estimated using the EKF or ECKF, the sign of $\sin(\psi_{1,f_1})$ and $\sin(\vartheta)$ will be corrected by detecting the estimated ψ_{1,f_1} and ϑ in an ascending or descending trend. Then, the multipath error profile on each antenna can be constructed to correct the carrier phase measurement and facilitate the integer ambiguity resolution.

5.2.6 Combined multipath correction and LOS estimation after fixing integer ambiguities

The integer ambiguity resolution is extensively introduced in chapter 3. Here, we assure these ambiguities have been correctly fixed. Then, the single differenced (SD) phase measurement between a pair of antennas $\Delta\phi_{i1}$ contains only the true range difference (as a function of the antenna baseline and the line of sight (LOS) vector), and the difference of multipath and phase noise between two antennas

$$\Delta\phi_{i1} - \lambda\Delta N = \mathbf{g}_{i1}^T \mathbf{x}_{\text{LOS}} + \delta\phi_{mp,i} - \delta\phi_{mp,1} + \Delta\varepsilon_{\phi_{i1}} \quad (5.60)$$

where $\mathbf{g}_{i1} = (g_{xi1}, g_{yi1}, g_{zi1})^T$ and $\mathbf{x}_{\text{LOS}} = (x, y, z)^T$ are the baseline vector and the LOS vector in the body fixed frame, ΔN is the SD integer ambiguity, which is assumed correctly fixed, and $\Delta\varepsilon_{\phi_{i1}}$ is the SD phase noise. The difference of multipath between two antennas can be formulated as a function of multipath parameters of

ψ_{1,f_1} , β , α_{f_1} and α_{f_2}

$$\begin{aligned}
\Delta\delta\phi_{mp,i1,f_1} &= \delta\phi_{mp,i,f_1} - \delta\phi_{mp,1,f_1} \\
&= \arctan\left(\frac{\alpha_{f_1}\sin(\psi_{i,f_1})}{1 + \alpha_{f_1}\cos(\psi_{i,f_1})}\right) - \arctan\left(\frac{\alpha_{f_1}\sin(\psi_{1,f_1})}{1 + \alpha_{f_1}\cos(\psi_{1,f_1})}\right) \\
&= \arctan\left(\frac{\alpha_{f_1}\sin(\psi_{i,f_1}) - \alpha_{f_1}\sin(\psi_{1,f_1}) + \alpha_{f_1}^2\sin(\psi_{i,f_1} - \psi_{1,f_1})}{1 + \alpha_{f_1}\cos(\psi_{1,f_1}) + \alpha_{f_1}\cos(\psi_{i,f_1}) + \alpha_{f_1}^2\cos(\psi_{i,f_1} - \psi_{1,f_1})}\right) \\
&= \arctan\left(\frac{\begin{aligned} &\alpha_{f_1}\sin(\psi_{1,f_1} + 4\pi/\lambda_1(l_i - l_1)\sin(\beta)) \\ &\quad - \alpha_{f_1}\sin(\psi_{1,f_1}) \\ &\quad + \alpha_{f_1}^2\sin(4\pi/\lambda_1(l_i - l_1)\sin(\beta)) \end{aligned}}{1 + \alpha_{f_1}\cos(\psi_{1,f_1}) + \alpha_{f_1}\cos(\psi_{1,f_1} + 4\pi/\lambda_1(l_i - l_1)\sin(\beta)) + \alpha_{f_1}^2\cos(4\pi/\lambda_1(l_i - l_1)\sin(\beta))}\right) \text{ [rad]} \\
\Delta\delta\phi_{mp,i1,f_2} &= \delta\phi_{mp,i,f_2} - \delta\phi_{mp,1,f_2} \\
&= \arctan\left(\frac{\alpha_{f_2}\sin(\psi_{i,f_2})}{1 + \alpha_{f_2}\cos(\psi_{i,f_2})}\right) - \arctan\left(\frac{\alpha_{f_2}\sin(\psi_{1,f_2})}{1 + \alpha_{f_2}\cos(\psi_{1,f_2})}\right) \\
&= \arctan\left(\frac{\begin{aligned} &\alpha_{f_2}\sin(\psi_{1,f_1} + 4\pi(l_i/\lambda_2 - l_1/\lambda_1)\sin(\beta)) \\ &\quad - \alpha_{f_2}\sin(\psi_{1,f_1} + 4\pi(l_1/\lambda_2 - l_1/\lambda_1)\sin(\beta)) \\ &\quad + \alpha_{f_2}^2\sin(4\pi/\lambda_2(l_i - l_1)\sin(\beta)) \end{aligned}}{1 + \alpha_{f_2}\cos(\psi_{1,f_1} + 4\pi(l_1/\lambda_2 - l_1/\lambda_1)\sin(\beta)) + \alpha_{f_2}\cos(\psi_{1,f_1} + 4\pi(l_i/\lambda_2 - l_1/\lambda_1)\sin(\beta)) + \alpha_{f_2}^2\cos(4\pi/\lambda_2(l_i - l_1)\sin(\beta))}\right) \text{ [rad]}
\end{aligned} \tag{5.61}$$

for $i = 2, \dots, n$. To substitute $\Delta\delta\phi_{mp,i1,f_1}$ or $\Delta\delta\phi_{mp,i1,f_2}$ to the SD measurement model, their unit shall be transformed from rads to meters by multiplying $\lambda_1/2\pi$ or $\lambda_2/2\pi$.

To estimate the LOS vector and correct multipath simultaneously, an extended Kalman filter can be built with the state including the LOS vector $\mathbf{x}_{LOS} = [x \ y \ z]^T$, the LOS rate $\dot{\mathbf{x}}_{LOS} = [\dot{x} \ \dot{y} \ \dot{z}]^T$ and the multipath parameters $\mathbf{x}_{mp} = [\psi_{1,f_1} \ \omega_1 \ \beta \ \omega_2 \ \alpha_{f_1} \ \alpha_{f_2}]^T$,

$$\begin{aligned}
\mathbf{x} &= [\mathbf{x}_{LOS}; \dot{\mathbf{x}}_{LOS}; \mathbf{x}_{mp}] \\
&= [x \ y \ z \ \dot{x} \ \dot{y} \ \dot{z} \ \psi_{1,f_1} \ \omega_1 \ \beta \ \omega_2 \ \alpha_{f_1} \ \alpha_{f_2}]^T \tag{5.62}
\end{aligned}$$

where ω_1 and ω_2 denote angular frequencies with respect to phases ψ_{1,f_1} and β .

The observations now include the SD code measurements $\Delta\rho_{i1}$, the SD phase measurements after removing ambiguities $\Delta\phi_{i1} - \lambda\Delta N$, and the ratios of SNRs of n antennas and dual frequencies

$$\mathbf{y}_k = \begin{bmatrix} \Delta\mathbf{P}(t_k) \\ \Delta\Phi_{f_1}(t_k) - \lambda_1\Delta\mathbf{N}_{f_1} \\ \Delta\Phi_{f_2}(t_k) - \lambda_2\Delta\mathbf{N}_{f_2} \\ \Upsilon_{f_1}(t_k) \\ \Upsilon_{f_2}(t_k) \end{bmatrix} = \begin{bmatrix} \mathbf{G}\mathbf{x}_{LOS,k} \\ \mathbf{G}\mathbf{x}_{LOS,k} + \mathbf{h}_{\delta\phi_{mp,f_1}}(\mathbf{x}_{mp,k}) \\ \mathbf{G}\mathbf{x}_{LOS,k} + \mathbf{h}_{\delta\phi_{mp,f_2}}(\mathbf{x}_{mp,k}) \\ \mathbf{h}_{SNR_{f_1}}(\mathbf{x}_{mp,k}) \\ \mathbf{h}_{SNR_{f_2}}(\mathbf{x}_{mp,k}) \end{bmatrix}_{\mathbf{x}_k = \hat{\mathbf{x}}_{k|k-1}} + \boldsymbol{\varepsilon} \tag{5.63}$$

where

$$\begin{aligned}
\Delta \mathbf{P} &= [\Delta \rho_{21} \quad \Delta \rho_{31} \quad \cdots \quad \Delta \rho_{n1}]^T \\
\Delta \Phi_{f_*} &= [\Delta \phi_{21,f_*} \quad \Delta \phi_{31,f_*} \quad \cdots \quad \Delta \phi_{n1,f_*}]^T \\
\Upsilon_{f_*} &= \left[\frac{SNR_{2,f_*}}{SNR_{1,f_*}} \quad \frac{SNR_{3,f_*}}{SNR_{1,f_*}} \quad \cdots \quad \frac{SNR_{n,f_*}}{SNR_{1,f_*}} \right]^T \\
\mathbf{G} &= [\mathbf{g}_{21} \quad \mathbf{g}_{31} \quad \cdots \quad \mathbf{g}_{n1}]^T \\
\mathbf{h}_{\delta\phi_{mp,f_*}}(\mathbf{x}_{mp,k}) &= [\Delta \delta\phi_{mp,21,f_*}(\mathbf{x}_{mp,k}) \quad \cdots \quad \Delta \delta\phi_{mp,n1,f_*}(\mathbf{x}_{mp,k})]_{\mathbf{x}_{mp,k}=\hat{\mathbf{x}}_{mp,k|k-1}}^T \\
\mathbf{h}_{SNR_{f_*}}(\mathbf{x}_{mp,k}) &= \left[\frac{SNR_{2,f_*}}{SNR_{1,f_*}}(\mathbf{x}_{mp,k}) \quad \cdots \quad \frac{SNR_{n,f_*}}{SNR_{1,f_*}}(\mathbf{x}_{mp,k}) \right]_{\mathbf{x}_{mp,k}=\hat{\mathbf{x}}_{mp,k|k-1}}^T.
\end{aligned}$$

The design matrix \mathbf{H}_k is now equal to

$$\mathbf{H}_{\mathbf{x}_k} = \begin{bmatrix} \mathbf{G} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{G} & \mathbf{0} & \mathbf{H}_{\mathbf{x}_k, \delta\phi_{mp,f_1}} & \mathbf{0} \\ \mathbf{G} & \mathbf{0} & \mathbf{0} & \mathbf{H}_{\mathbf{x}_k, \delta\phi_{mp,f_2}} \\ \mathbf{0} & \mathbf{0} & \mathbf{H}_{\mathbf{x}_k, SNR_{f_1}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{H}_{\mathbf{x}_k, SNR_{f_2}} \end{bmatrix} \quad (5.64)$$

where

$$\begin{aligned}
\mathbf{H}_{\mathbf{x}_k, \delta\phi_{mp,f_*}} &= \left. \frac{\partial \mathbf{h}_{\delta\phi_{mp,f_*}}(\mathbf{x}_{mp,k})}{\partial \mathbf{x}_{mp,k}} \right|_{\mathbf{x}_{mp,k}=\hat{\mathbf{x}}_{mp,k|k-1}} \\
\mathbf{H}_{\mathbf{x}_k, SNR_{f_*}} &= \left. \frac{\partial \mathbf{h}_{SNR_{f_*}}(\mathbf{x}_{mp,k})}{\partial \mathbf{x}_{mp,k}} \right|_{\mathbf{x}_{mp,k}=\hat{\mathbf{x}}_{mp,k|k-1}}.
\end{aligned} \quad (5.65)$$

The dynamic model for \mathbf{x}_{mp} can be found in section 5.2.5, while the dynamic model for \mathbf{x}_{LOS} and $\dot{\mathbf{x}}_{LOS}$ is

$$\begin{bmatrix} \dot{\mathbf{x}}_{LOS}(t) \\ \ddot{\mathbf{x}}_{LOS}(t) \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{0} & \mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} \end{bmatrix}}_{\mathbf{F}(t)} \begin{bmatrix} \mathbf{x}_{LOS}(t) \\ \dot{\mathbf{x}}_{LOS}(t) \end{bmatrix} + \underbrace{\begin{bmatrix} \mathbf{0} \\ \mathbf{I}_3 \end{bmatrix}}_{\mathbf{G}_{\ddot{\mathbf{x}}}(t)} \ddot{\mathbf{x}}_{LOS}(t) \quad (5.66)$$

where $\ddot{\mathbf{x}}(t)$ is the acceleration in orbit or attitude dynamics, which can be written as $\ddot{\mathbf{x}}(t) = f(\mathbf{x}(t), \dot{\mathbf{x}}(t))$. However, it is assumed as random noise here for simplicity. This is valid under the assumption that the time interval Δt is sufficiently short. In this situation, the prediction of the relative LOS and LOS rate $[\mathbf{x}(t), \dot{\mathbf{x}}(t)]^T$ between two spacecraft is linear and described by a purely kinematic model. The estimation may not be as accurate as using the specific orbit dynamic model, but makes it generally usable to more applications in addition to the applications in space.

The transition matrix $\Phi_{LOS,k+1|k}$ can be obtained as

$$\Phi_{LOS,k+1|k} = e^{\mathbf{F}\Delta t} = \sum_{i=0}^{\infty} \frac{\mathbf{F}^i(t)\Delta t^i}{i!} = \begin{bmatrix} \mathbf{I}_3 & \Delta t \mathbf{I}_3 \\ \mathbf{0} & \mathbf{I}_3 \end{bmatrix} \quad (5.67)$$

and its discrete-time process noise is calculated as

$$\begin{aligned}
\mathbf{Q}_{LOS,k+1|k} &= \int_{t_k}^{t_{k+1}} \int_{t_k}^{t_{k+1}} \Phi_{LOS}(t_{k+1}, \tau) \mathbf{G}_{\ddot{\mathbf{x}}}(\tau) \sigma_{\ddot{\mathbf{x}}_{LOS}}^2 \mathbf{G}_{\ddot{\mathbf{x}}}^T(\tau) \Phi_{LOS}^T(t_{k+1}, \tau) d\tau d\tau \\
&= \sigma_{\ddot{\mathbf{x}}_{LOS}}^2 \begin{bmatrix} \Delta t^4/12 & \Delta t^3/12 \\ \Delta t^3/6 & \Delta t^2/2 \end{bmatrix} \otimes \mathbf{I}_3
\end{aligned} \quad (5.68)$$

Table 5.2: Antenna coordinates in the body fixed frame in the simulation

	g_x [m]	g_y [m]	g_z [m]
Ant 1 (Ref)	1.37	0.56	1.00
Ant 2	1.08	1.20	0.80
Ant 3	0.45	0.62	0.60
Ant 4	0.89	0.78	0.40
Ant 5	0.51	0.35	0.20

where \otimes is the Kronecker product.

Denoting the transition matrix and the process noise for the multipath parameter estimation in Eq.(5.45) and (5.46) as $\Phi_{\mathbf{x}_{mp},k+1|k}$ and $\mathbf{Q}_{\mathbf{x}_{mp},k+1|k}$, the final $\Phi_{k+1|k}$ and $\mathbf{Q}_{k+1|k}$ for the state and covariance propagation are

$$\Phi_{k+1|k} = \begin{bmatrix} \Phi_{LOS,k+1|k} & \mathbf{0} \\ \mathbf{0} & \Phi_{\mathbf{x}_{mp},k+1|k} \end{bmatrix}, \mathbf{Q}_{k+1|k} = \begin{bmatrix} \mathbf{Q}_{LOS,k+1|k} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_{\mathbf{x}_{mp},k+1|k} \end{bmatrix}. \quad (5.69)$$

This EKF shall be implemented following the time update and iterated measurement update equations in Eq. (5.29) epoch by epoch. Note that it is also possible to use ECKF for this estimation.

Compared to the solely SNR-based multipath estimation, the more precise phase measurements are added as observations in addition to the ratios of SNRs. It is also found that even if the ratios of SNRs are not included in the measurement model, the filter is still able to precisely estimate the parameters of interest.

5.3 Performance evaluation

The performance of the multipath estimation in an EKF or ECKF is evaluated based on a simple geometry scenario in Figure 5.9, where five antennas are assumed installed on the top surface of a spacecraft and multipath signals are reflected from a big solar panel in the xy -plane in the body fixed frame. The antenna coordinates are given in Table 5.2. The z -coordinate of each antenna represents the perpendicular distance between the antenna and the reflector, while the x - and y - coordinates do not influence the multipath error as the reflector in this simulation is not tilted. However, the x -, y - and z -coordinates all contribute to the determination of the antenna baseline matrix \mathbf{G} . This matrix \mathbf{G} is the design matrix in the IAR and is a part of the design matrix in the EKF for the combined LOS estimation and multipath correction. The matrix \mathbf{G} in the body fixed frame can be easily calculated by subtracting xyz -coordinates of the reference antennas from ancillary antennas.

In this simulation, signals are sent by a satellite with an ascending elevation from 5° to 45° . Measurements (including pseudorange, carrier phase and SNR measurements on five antennas and dual frequencies) of 8000 epochs have been received. The measurement update rate is 1 second. Antenna gain patterns are assumed identical on each antenna in all directions. The relative multipath amplitude coefficient α_{f_1} is assumed decreasing from 0.8 to 0.1 during the satellite's relative motion from low to high elevation, while α_{f_2} is declining from 0.7 to 0.09.

The effectiveness of the EKF and ECKF as the second cascaded filter for the multipath parameter estimation will be discussed in the following based on a proper

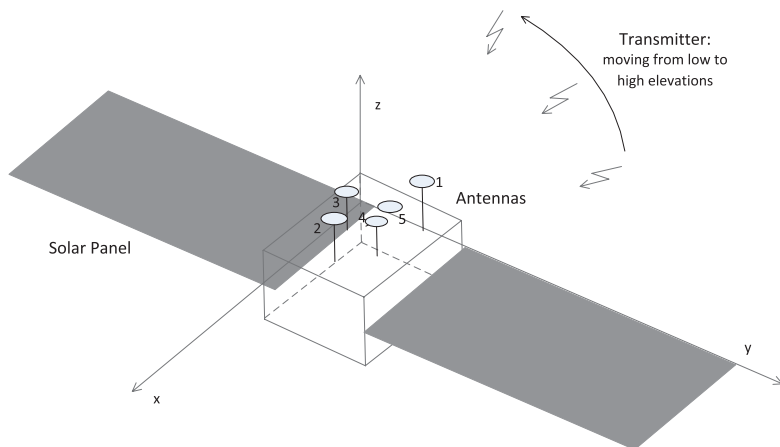


Figure 5.9: A geometry scenario for simulating multipath corrections and integer ambiguity resolution performances

initial conditions. Their response with respect to poor initial conditions, large noise of observations or multi-reflection scenarios will also be explicitly evaluated and compared. Single-frequency SNR measurements will be used in the simulation. The dual-frequency performance will be demonstrated in conjunction with the LOS estimation in the third cascaded filter in section 5.4.

5.3.1 Sensitivity to initial conditions

Due to the nonlinearity of the measurement model, a poorly initial guess of the state may result in divergence of the filter. It is of great importance to examine the filter sensitivity to initial conditions. Figure 5.10 depicts the EKF and ECKF performances in two initial conditions of Table 5.3.

Other important parameters used in the filter include the process noise \mathbf{Q} and the measurement noise \mathbf{R} . We assume here the noise of the original SNR measurement has a standard deviation σ_{SNR} of 0.03. After the noise filtering in an EKF (section 5.2.4), the filtered ratio of SNRs has a noise of around two times larger than σ_{SNR} , which essentially determines the value of \mathbf{R} in the subsequent EKF or ECKF for the multipath parameter estimation. In the ECKF, not only the ratios of SNRs are treated as measurements, two pseudo-observations are also added with a small artificial noise. The specific values of \mathbf{R} and \mathbf{Q} used in this simulation are listed in Table 5.3.

For *Condition I* in Table 5.3, the initial guess of the state is close to its true value. The performance of the EKF and ECKF in this condition is depicted in Figure 5.10 (b) and (c). As can be seen, opposite signs (or 180° phase shift) between the estimated state (solid line) and its true value (dashed line) may show up on α_{f_1} , $\cos(\psi_{1,f_1})$, $\cos(\vartheta)$, $\sin(\psi_{1,f_1})$ and $\sin(\vartheta)$. The ambiguities on signs can easily be corrected by checking whether ψ_{1,f_1} and ϑ are ascending or descending as expected in time. Details can be found in section 5.2.3. Apart from the ambiguous signs, a nearly immediate convergence at the beginning of the time series can be observed due to a good initial guess of the state.

As updated in time, the convergence of both the EKF and ECKF has been broken at the time batch of [4500, 5500] epoch, which is called here “observation

Table 5.3: Specifications of comparison of the EKF and ECKF in terms of the sensitivity to initial conditions

<i>Condition I:</i> the initial guess of the state is close to the true value	
EKF:	$\mathbf{x}_{0 0} = [\omega_1 \ \psi_{1,f_1} \ \omega_2 \ \beta \ \alpha_{f_1}]^T$ $= [0.01 \ 4 \ 0.01 \ 0.1 \ 0.5]^T$, in Figure 5.10 (b)
ECKF:	$\mathbf{x}_{0 0} = [e^{j\omega_1\Delta t} \ e^{j\psi_{1,f_1}} \ e^{-j\psi_{1,f_1}} \ e^{j\omega_2\Delta t} \ e^{j\vartheta} \ e^{-j\vartheta} \ \alpha_{f_1}]^T$ $= [e^{0.01j} \ e^{4j} \ e^{-4j} \ e^{0.01j} \ e^{j\frac{4\pi}{\lambda_1}(l_2-l_1)\sin(0.1)} \ e^{-j\frac{4\pi}{\lambda_1}(l_2-l_1)\sin(0.1)} \ 0.5]^T$ $= [1.00 + 0.01j \ -0.65 - 0.76j \ -0.65 + 0.76j$ $1.00 + 0.01j \ 0.24 - 0.97j \ 0.24 + 0.97j \ 0.5]^T$, in Figure 5.10 (c)
<i>Condition II:</i> the initial guess of the state is far from the true value	
EKF:	$\mathbf{x}_{0 0} = [\omega_1 \ \psi_{1,f_1} \ \omega_2 \ \beta \ \alpha_{f_1}]^T$ $= [-0.5 \ 1 \ 0.5 \ 0.5 \ 0.05]^T$, in Figure 5.10 (d)
ECKF:	$\mathbf{x}_{0 0} = [e^{j\omega_1\Delta t} \ e^{j\psi_{1,f_1}} \ e^{-j\psi_{1,f_1}} \ e^{j\omega_2\Delta t} \ e^{j\vartheta} \ e^{-j\vartheta} \ \alpha_{f_1}]^T$ $= [e^{-0.5j} \ e^{1j} \ e^{-1j} \ e^{0.5j} \ e^{j\frac{4\pi}{\lambda_1}(l_2-l_1)\sin(0.5)} \ e^{-j\frac{4\pi}{\lambda_1}(l_2-l_1)\sin(0.5)} \ 0.05]^T$ $= [0.88 - 0.48j \ 0.54 + 0.84j \ 0.54 - 0.84j$ $0.88 + 0.48j \ 0.99 - 0.04j \ 0.99 + 0.04j \ 0.05]^T$, in Figure 5.10 (e)
Other parameters used in the filter	
EKF:	$\mathbf{P}_{0 0} = \mathbf{I}_5$, $\mathbf{R} = 4\sigma_{\text{SNR}}^2 \mathbf{I}_{n-1}$, $Q_{\omega_1} = 10^{-4}$, $Q_{\omega_2} = 10^{-6}$, $Q_{\alpha} = 10^{-4}$
ECKF:	$\mathbf{P}_{0 0} = \mathbf{I}_7$, $\mathbf{R} = \text{diag}(4\sigma_{\text{SNR}}^2 \mathbf{I}_{n-1}, (10^{-3} + 10^{-4}j)\mathbf{I}_2)$, $Q_{s_{x_1}} = 10^{-3} + 10^{-3}j$, $Q_{s_{x_4}} = 10^{-3} + 10^{-3}j$, $Q_{\alpha} = 10^{-4}$

overlap”. This anomaly occurs at the common multiple points of different SNR oscillatory cycles. The least common multiple of the oscillatory cycles of all observations occur at around 4700 epoch, when all measurements shrink (overlap) into a single measurement and thus could not provide sufficient information to the filter for the estimation of multiple non-stationary state variables. Therefore, in the time batch of [4500, 5500] epoch, both the EKF and ECKF have locked onto a wrong state and are unaware that the true error variance is diverging, as neither of them can distinguish whether the amplitude or the phase or both cause the oscillation of a single (overlapped) measurement. From Figure 5.10 (a), except for the time batch of [4500, 5500] epoch, another observation overlap phenomena occurs at around 1750 epoch when four measurements (ratios of SNRs) collapse to two measurements. This does not result in divergence. However, it may disturb the established convergence of the filter, leading to the changes of signs of α_{f_1} , $\cos(\psi_{1,f_1})$, $\cos(\vartheta)$, $\sin(\psi_{1,f_1})$ or $\sin(\vartheta)$, as illustrated in Figure 5.10 (c) and (e), and Figure 5.11 (b) and (c).

For *Condition II* in Table 5.3, the initial guess of the state is far from its true value, especially the initial guess for frequencies ω_1 and ω_2 , which are more than one order of magnitude larger. Under this initial condition, the EKF in Figure 5.10 (d) takes much longer time to converge while the ECKF in Figure 5.10 (f) is still able to fast track the correct value. Although both the observation models in the EKF and ECKF are nonlinear, the nonlinearity in the EKF is much stronger as the observations are sine or cosine functions of the state and thus have infinite Taylor series. Therefore, the linearization assumption $\mathbf{h}(\mathbf{x}(t)) \approx \mathbf{h}(\mathbf{x}_{k+1|k}) + \mathbf{H}_{\mathbf{x}_{k+1}} \delta \mathbf{x}$ in the EKF is likely to be violated, especially when $\delta \mathbf{x}$ is not small, i.e., for poorly initial conditions. On the contrary, the nonlinearity in the ECKF observation model is relatively weak as it is in the form of \mathbf{x}^r with r being determined by the perpendicular distance between the antenna and reflector. By properly choosing this distance, the nonlinearity of the ECKF can be largely reduced. This makes the ECKF less sensitive to initial conditions.

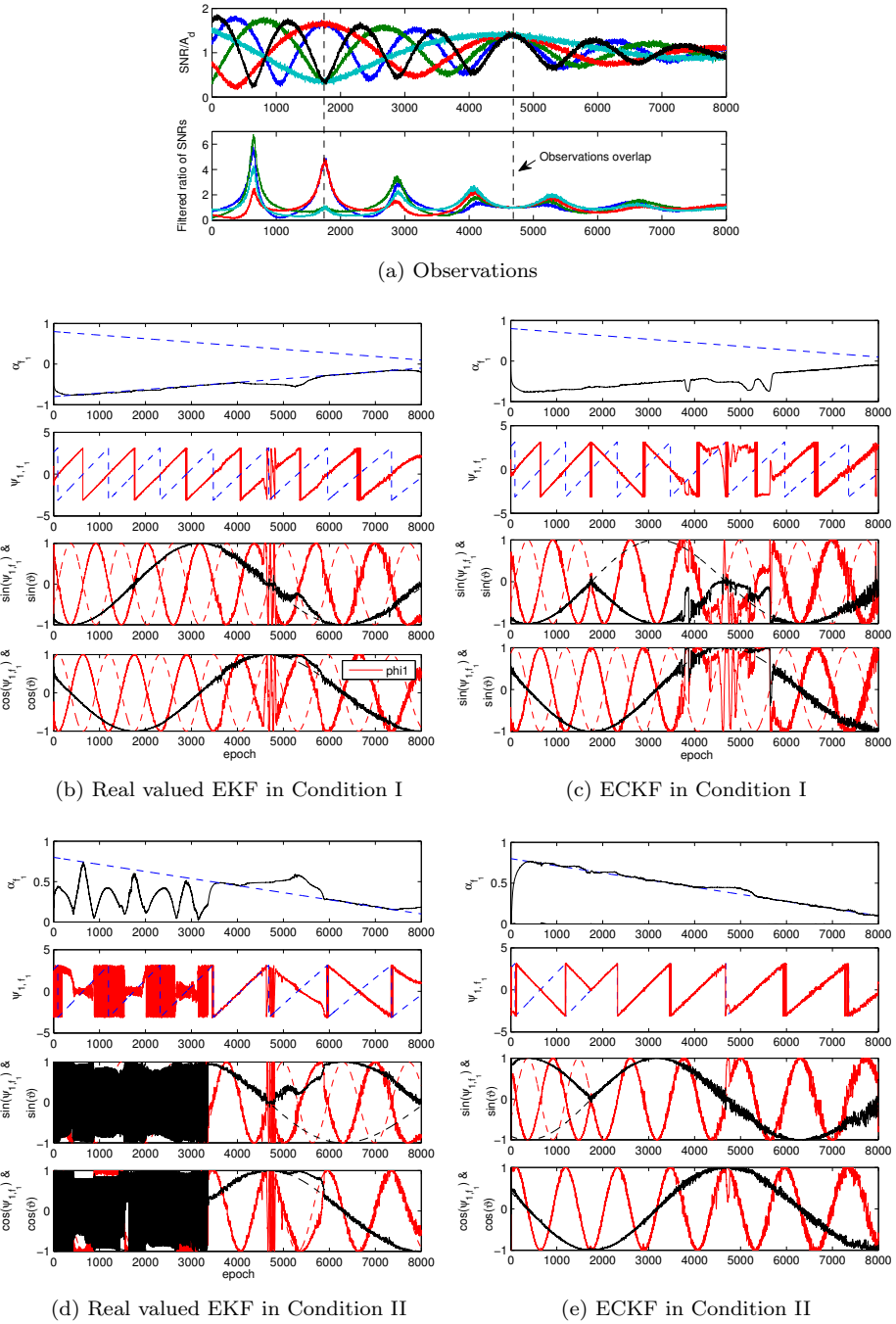


Figure 5.10: Comparison of the real valued EKF and ECKF performance in different initial conditions. The dashed line indicates the true multipath parameters, including the amplitude, phase, and the sine and cosine of the phase, while the solid line denotes the estimation thereof in the EKF or ECKF.

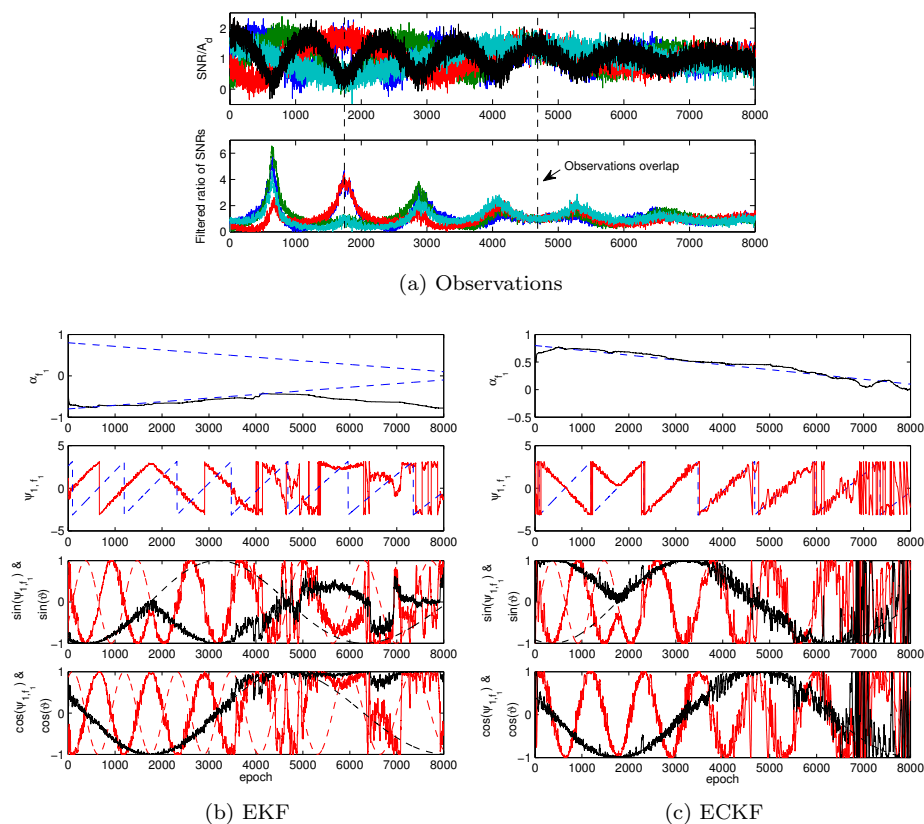


Figure 5.11: Comparison of EKF and ECKF performance given large noise observations. The dashed line indicates the true multipath parameters, including the amplitude, phase, and the sine and cosine of the phase, while the solid line denotes the estimation thereof in the EKF or ECKF.

Moreover, as opposed to *Condition I*, the ECKF in *Condition II* even better responds to the observation overlap phenomena in the time batch of [4500, 5500] epoch. It is found that as long as both the real and imaginary parts of the complex initial state are non-zeros (better to be away from zeros), the ECKF will have a good convergence property.

Fast convergence is of great importance for the multipath estimation as one of the purposes of multipath removal is to accelerate the integer ambiguity resolution using the corrected multipath-free phase measurement.

5.3.2 Tolerance to large noise observations

Since not all receivers provide precise SNR data, it is better to evaluate and compare the EKF and ECKF performances using large noise observations. Figure 5.11 depicts the comparison results when the standard deviation of the SNR measurements is increased from 0.03 to 0.2.

Both the EKF and ECKF in Figure 5.11 (b) and (c) are able to track the correct state at the beginning of the time series. As the multipath amplitude α decreases, the oscillatory magnitude of the SNR observations become smaller and gradually

reach the same magnitude level as the random noise. This makes the observations on different antennas difficult to distinguish. If we define the region that the filter is unable to lock onto the correct state as an undistinguishable region, this undistinguishable region starts from epoch 4000 for the EKF but from epoch 6500 for the ECKF. It is obvious that the ECKF is superior in estimating the state in the large noise environment. In practise, when the multipath is close to the same magnitude level as the random noise, e.g., after epoch 6500 in this simulation, it is not necessary any more to estimate multipath in order to facilitate the subsequent integer ambiguity resolution. As will demonstrated in section 5.4, the integer ambiguity resolution has a certain tolerance to small multipath.

It is suggested to include an inequality constraint on α in the EKF and ECKF in the future work, such as $0 \leq \alpha < 1$, which is valid in this application as the multipath amplitude is always smaller than the direct signal amplitude. This could constrain the propagation of α and help to distinguish the real origin (α or ψ or both) that causes the oscillation in the observation. The performance in large noise environment as well as in observation overlap phenomena are expected to be improved.

5.3.3 Robustness in multi-reflection conditions

The effectiveness of the EKF and ECKF has also been examined in a multiple multipath environment. The multipath signal in space may not only come from the surroundings of the antenna (e.g. solar panel), it could also be reflected from a nearby space vehicle, which introduces high frequency multipath due to its relatively distant distance off the antenna. Therefore, except for the assumed multipath geometry in Figure 5.9, a space vehicle 50 m away in the x -direction is assumed as another multipath source. Due to the relative dynamics, only a short-duration signal reflection off of the space vehicle may be received by the antenna. This reflected signal is assumed with a relatively small amplitude, α ranging from 0.2 to 0.1, while the reflected signal off of the solar panel in the surrounding of the antenna is the dominating multipath source.

Therefore, two oscillatory multipath waves can be observed in the SNR data in Figure 5.12 (a). The slowly varying (lower frequency) multipath trend is caused by the nearby solar panel, while the fast frequency multipath modulated on top of the slowly varying trend is introduced by the distant space vehicle across a short-duration range of [2000, 4000] epoch.

The EKF and ECKF performances under such a multi-reflection scenario are displayed in Figure 5.12 (b) and (c). It can be observed that both the EKF and ECKF are able to sense the correct value in the multi-reflection duration, and there is no apparent difference of their responses to this multi-reflection condition.

5.4 Multipath effects on the integer ambiguity resolution

The carrier phase multipath error easily leads to an incorrect ambiguity resolution and is currently one of few remaining obstacles for high precision real time positioning. In the following, the multipath effects on the integer ambiguity resolution will be evaluated. The performance of the combined multipath and LOS estimation in the third cascaded EKF will also be demonstrated.

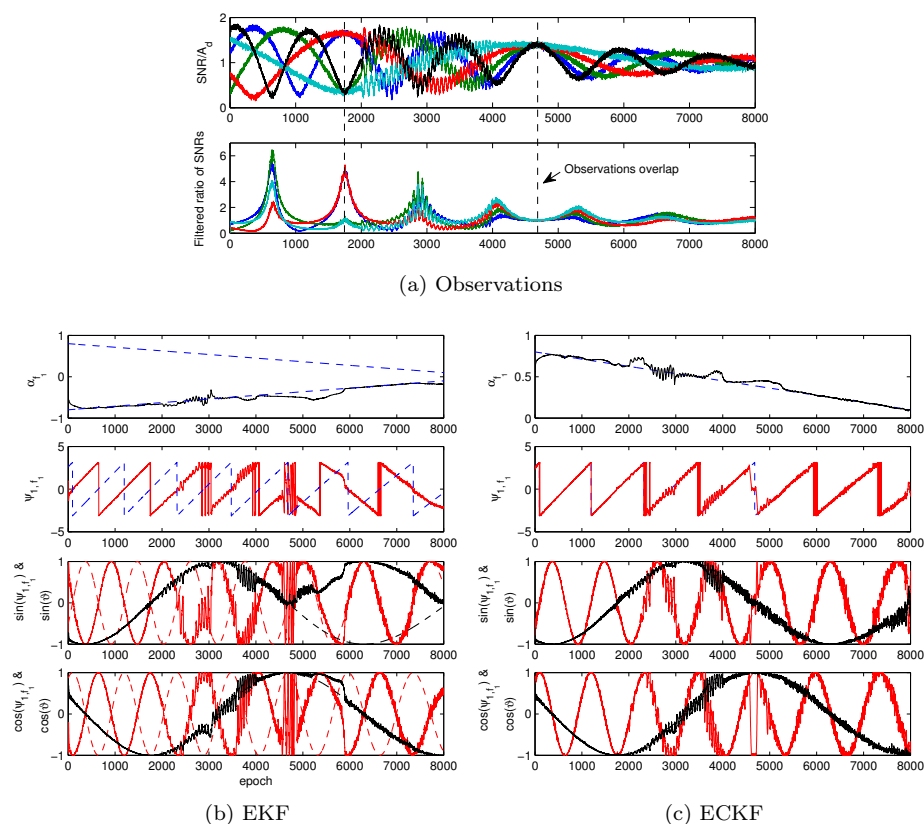


Figure 5.12: Comparison of EKF and ECKF performances under multi-reflection conditions. The dashed line indicates the true multipath parameters, including the amplitude, phase, and the sine and cosine of the phase, while the solid line denotes the estimation thereof in the EKF or ECKF.

5.4.1 IAR acceleration

Multipath causes a fluctuation of the pseudorange and carrier phase measurements, leading to a biased float ambiguity solution and an enlarged search space. As a result, a longer time is often needed to resolve the ambiguity, otherwise the probability of successful ambiguity estimation decreases. This has been demonstrated by simulations using the multipath scenario in Figure 5.9. Dual frequency measurements are now used in the Kalman filter to estimate the LOS, LOS rate and ambiguities without or with multipath corrections. The standard deviation of the pseudorange and carrier phase measurements are 0.5 m and 0.001 m, respectively. Results are displayed in Figure 5.13 and 5.14.

At each time update and measurement update in the Kalman filter, a group of new float ambiguities $\hat{\mathbf{a}}$ and the associated ambiguity covariance matrix $\mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{a}}}$ are obtained epoch by epoch. The LAMBDA method is used to fix the float ambiguities into integers as each new group of $\hat{\mathbf{a}}$ and $\mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{a}}}$ becomes available. If the fixed ambiguities are the same as the true ambiguities in this simulation, it means that the ambiguities are correctly fixed; otherwise, the ambiguities have been wrongly fixed.

In Figure 5.13, uncorrected multipath errors are embedded in the measurements

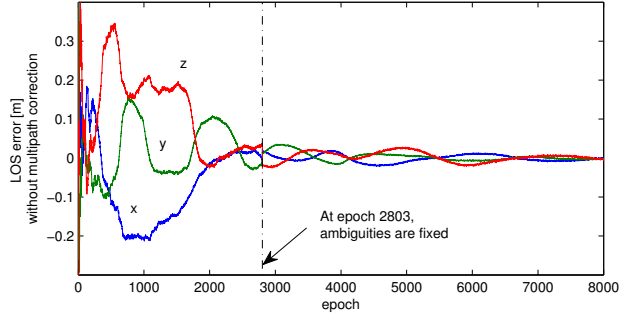


Figure 5.13: LOS error without multipath correction

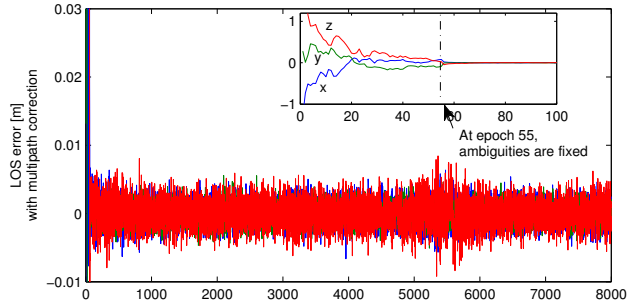


Figure 5.14: LOS error with multipath correction (Note the different scale as compared to Figure 5.13)

and treated as noise. Ambiguities could not be correctly fixed until epoch 2803 when $\mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{a}}}$ is precise enough and $\hat{\mathbf{a}}$ falls in the correct integer least-squares pull-in region. The fluctuations of the multipath not only bias the float ambiguity and make it difficult to resolve, it also bias the float LOS solution and the fixed LOS solution even after the ambiguity is correctly fixed. As shown in Figure 5.13, only a cm-order accuracy can be achieved ultimately due to the uncorrected multipath fluctuation.

As opposed to the time of first fix at epoch 2803 in Figure 5.13, ambiguities have been correctly fixed at epoch 55 in Figure 5.14, demonstrating that a tremendously faster ambiguity resolution can be guaranteed by removing multipath from the phase measurements. After ambiguities are fixed, the combined LOS, LOS rate and multipath estimation in a third cascaded EKF has been used. The result in Figure 5.14 indicates a mm-order accuracy of the LOS estimation.

5.4.2 Multipath robustness in single-epoch IAR

The nearby reflector introduces slowly varying bias-like multipath errors. For the integer ambiguity resolution, this bias in the observation will propagate to the float ambiguity solution, leading to higher possibility of the wrong integer estimation. Specifically, if the SD multipath vector is denoted as $\boldsymbol{\delta}_{mp}$ as the bias, the SD model for the integer ambiguity resolution will be in the form

$$\mathbf{y} = \mathbf{B}\mathbf{x}_{LOS} + \mathbf{A}\mathbf{a} + \mathbf{C}\boldsymbol{\delta}_{mp} + \boldsymbol{\varepsilon} \quad (5.70)$$

where \mathbf{C} consists of 1's and 0's to indicate whether the bias exists in a specific measurement, \mathbf{a} contains all the SD ambiguities, \mathbf{B} includes the antenna baselines matrix, \mathbf{A} contains the wavelength of carrier frequencies. Specific expressions for \mathbf{B} and \mathbf{A} can be found in chapter 3.

If the bias is ignored, the float solutions of $\hat{\mathbf{a}}$ and $\hat{\mathbf{x}}_{LOS}$ are

$$\begin{bmatrix} \hat{\mathbf{a}} \\ \hat{\mathbf{x}}_{LOS} \end{bmatrix} = \begin{bmatrix} \mathbf{Q}_{\hat{\mathbf{x}}\hat{\mathbf{x}}} & \mathbf{Q}_{\hat{\mathbf{x}}\hat{\mathbf{a}}} \\ \mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{x}}} & \mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{a}}} \end{bmatrix} \begin{bmatrix} \mathbf{A}^T \\ \mathbf{B}^T \end{bmatrix} \mathbf{Q}_{\mathbf{y}\mathbf{y}}^{-1} \quad (5.71)$$

where $\mathbf{Q}_{\mathbf{y}\mathbf{y}}$, $\mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{a}}}$, $\mathbf{Q}_{\hat{\mathbf{x}}\hat{\mathbf{x}}}$, $\mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{x}}}$ and $\mathbf{Q}_{\hat{\mathbf{x}}\hat{\mathbf{a}}}$ are variance and covariance matrices for \mathbf{y} , $\hat{\mathbf{a}}$ and $\hat{\mathbf{x}}_{LOS}$. Derivations for their close-form expressions can be found in Appendix A. When the bias is included as the model (5.70), the resulting bias $\delta\hat{\mathbf{a}}$ and $\delta\hat{\mathbf{x}}_{LOS}$ has to be added to the float solutions (Verhagen, 2012)

$$\begin{bmatrix} \delta\hat{\mathbf{a}} \\ \delta\hat{\mathbf{x}}_{LOS} \end{bmatrix} = \begin{bmatrix} \mathbf{Q}_{\hat{\mathbf{x}}\hat{\mathbf{x}}} & \mathbf{Q}_{\hat{\mathbf{x}}\hat{\mathbf{a}}} \\ \mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{x}}} & \mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{a}}} \end{bmatrix} \begin{bmatrix} \mathbf{A}^T \\ \mathbf{B}^T \end{bmatrix} \mathbf{Q}_{\mathbf{y}\mathbf{y}}^{-1} \mathbf{C} \delta_{mp} . \quad (5.72)$$

Figure 5.15 illustrates the impact of a bias in the float ambiguity solution on the ambiguity resolution success rate (Verhagen, 2012). The left figure shows the distribution of float ambiguity estimates in the absence of a bias. All green dots are float estimates for which the corresponding integer least-squares (ILS) solution by the LAMBDA method is correct. Those ambiguities are successfully fixed. The red dots are float estimates that reside in the wrong ILS pull-in region and are thus incorrectly fixed. The success rate in this case is equal to 99.9%. It can be observed that the point cloud is centered at the correct integer vector. The right figure shows how this point cloud is shifted due to a bias $\delta\hat{\mathbf{a}} = [0.25 \ 0.25]^T$. As a consequence, the number of green dots is much smaller, and the bias-affected success rate has decreased to 75%.

If the unbiased float ambiguity $\hat{\mathbf{a}}$ is much more precise than the one shown in Figure 5.15 (a), i.e., having a much smaller elliptical ambiguity covariance in the relatively larger pull-in region, a small shift of the point cloud is tolerable as long as the cloud still resides in the same pull-in region. This means uncorrected small biases may not corrupt the final ambiguity resolution result. It will thus be interesting to know the allowable size of a bias to still allow for reliable integer estimation.

One way to enlarge this allowable size of the bias is to make use of constraints in the ambiguity resolution. As discussed in chapter 3, the length constraint of \mathbf{x}_{LOS} helps to validate the ambiguity candidates in the ambiguity search process, so that not only the ambiguity objective function should be minimized by the integer least squares, but the length of the resultant conditional LOS vector $\check{\mathbf{x}}_{LOS}(\mathbf{a})$ shall also be fulfilled within a predefined threshold δl . The LOS constraint thus works in the way of aiding in accepting only the correct ambiguity solution and rejecting wrong solutions. Equations for constraining the conditional LOS vector include

$$\check{\mathbf{x}}_{LOS}(\mathbf{a}) = \hat{\mathbf{x}}_{LOS} - \mathbf{Q}_{\hat{\mathbf{x}}\hat{\mathbf{a}}} \mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{a}}}^{-1} (\hat{\mathbf{a}} - \check{\mathbf{a}}) \quad (5.73)$$

$$l - \delta l \leq \|\check{\mathbf{x}}_{LOS}(\mathbf{a})\| \leq l + \delta l \quad (5.74)$$

where $\check{\mathbf{a}}$ is a fixed ambiguity candidate.

Since both $\hat{\mathbf{a}}$ and $\hat{\mathbf{x}}_{LOS}$ in Eq.(5.73) are biased by $\delta\hat{\mathbf{a}}$ and $\delta\hat{\mathbf{x}}_{LOS}$, the resultant $\|\check{\mathbf{x}}_{LOS}(\mathbf{a})\|$ is also biased. This requires a larger δl than the unbiased case to avoid wrong rejection of a correct ambiguity. The closed-form expression of δl for the

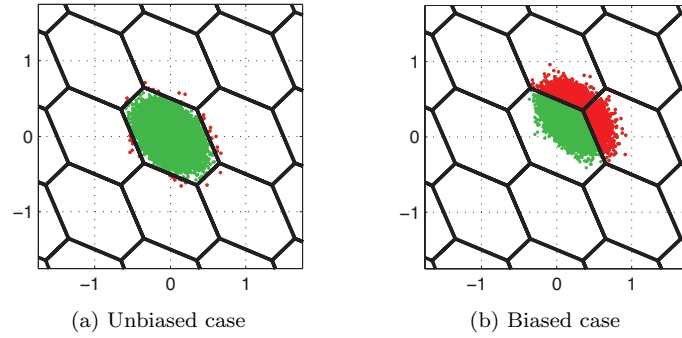


Figure 5.15: Two-dimensional integer least squares pull-in regions (black) with 10^4 float solutions, which are colored green if the corresponding fixed solution is correctly fixed, and red if fixed wrong. (a) is the unbiased case and (b) is the biased case. (Verhagen, 2012)

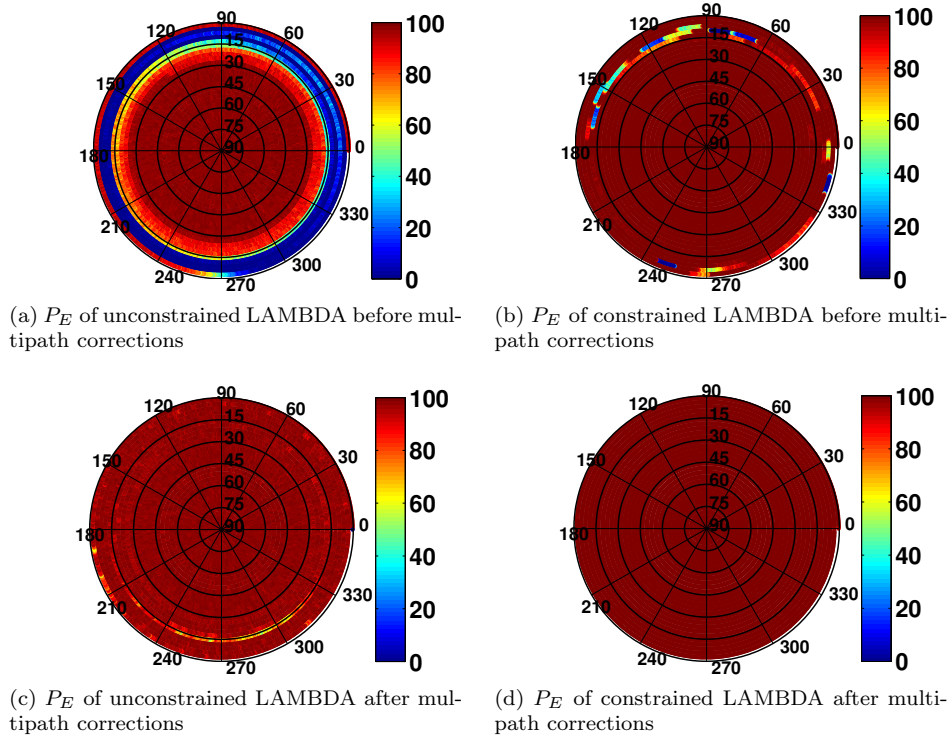


Figure 5.16: Empirical success rate P_E [%] of single-epoch integer ambiguity resolutions before and after multipath corrections

unbiased case has been derived in chapter 3. Here, we assign δl two times larger to address biases.

Figure 5.16 shows the bias-robustness of the integer ambiguity resolution, where the multipath on each antenna is regarded as a bias in the single-epoch dual-frequency measurement. Multipath is generated as the satellite is rotating around the boresight axis of the antenna along with an ascending elevation from 0° to 90° .

The multipath relative amplitude coefficients α_{f_1} and α_{f_2} change according to the antenna gain patterns of the Novatel NOV702GG antenna. The specific explanation on how α_{f_1} and α_{f_2} change as a function of the elevation can be found in Chapter 4. Five antennas are used and their coordinates in the body fixed frame are in Table 5.2. Their z -coordinates represent the perpendicular distances from antennas to the horizontal reflector. Apart from the multipath, the random noise on the undifferenced code and carrier phase measurement in this simulation has the standard deviations of 1 m and 0.003 m, respectively. The empirical success rate P_E of the IAR is obtained by Monte Carlo simulations.

Figure 5.16 (a) and (b) depict the success rate of the unconstrained and constrained LAMBDA methods in the presence of uncorrected multipath errors. It is clear that at low elevations, a low success rate is obtained using the unconstrained LAMBDA method due to the fact that phase measurements are heavily corrupted by the multipath. However, the constrained LAMBDA method can provide a much higher possibility of still maintaining high success rate at low elevations, indicating that it has much better tolerance to multipath biases. Here, the threshold δl in the constrained LAMBDA is chosen two times larger than the unbiased case.

Figure 5.16 (c) and (d) illustrate the success rate of the unconstrained and constrained LAMBDA methods after the multipath is constructed and removed from the carrier phase measurement. The extended complex Kalman filter in section 5.2.5 is used for the multipath parameter estimation. It is clear that after the multipath correction, ambiguities can be more reliably resolved. Some remaining errors when the multipath is not completely or correctly constructed may cause the unconstrained LAMBDA failing in the integer estimation. However, these errors are completely tolerable in the constrained LAMBDA.

Assessing the bias-robustness of the constraint integer ambiguity resolution will be important as the future work in order to guarantee the uncorrected multipath will not corrupt the final ambiguity resolution result. However, this does not mean that the multipath correction process can be skipped, as the multipath does not only impede the ambiguity resolution, but also decreases the positioning accuracy of the final interest.

5.5 Chapter summary

This chapter proposed a promising real-time multipath mitigation method by making use of the SNR data on multiple antennas and their spatial correlations between antennas. Cascaded extended Kalman filters were implemented: the first EKF was used to filter the noise on ratios of SNRs; the second successive EKF (or complex EKF) was used to estimate multipath parameters, which are then reformulated to construct the multipath errors in order to remove these errors from the phase measurements; the integer ambiguity resolution was accelerated due to the multipath removal; after ambiguities were fixed, a third cascaded EKF was used as a combined LOS, LOS rate and multipath estimator, which guaranteed the achievement of mm-order LOS accuracy in the end.

For the multipath parameter estimation, this chapter proposed a real-valued EKF and a complex-valued EKF (ECKF). The ECKF has been found to be insensitive to the initial conditions, while the real-valued EKF was difficult to converge with poorly initial conditions. Moreover, the ECKF has shown better convergence properties for observations with large noise. Both the real-valued EKF and ECKF respond equally

to multi-reflection conditions.

Multipath effects on the integer ambiguity resolution were also examined in this chapter. The time required for the first ambiguity fix has been tremendously reduced after the multipath was estimated and removed from the phase measurement. The LAMBDA method, which is the benchmark in the integer ambiguity resolution domain, has been demonstrated robust to small multipath. This robustness can be improved if the proposed constrained LAMBDA method in chapter 3 is used.

Chapter 6

Network Architecture

As a successor to previous chapters on the RF-based communication and navigation system design in a two-spacecraft formation, this chapter aims at extending previous scenarios and results on systems for a large scale formation with four or more spacecraft. The chapter includes a discussion on potential formation network architectures and investigations of limitations in implementing specific architectures.

In section 6.1 and 6.2, dedicated requirements on the formation network are analysed. Several multiple access network architectures and topologies are discussed and evaluated. CDMA is emphasized in this chapter. In section 6.3, the limitations of CDMA in terms of the multiple access interference and the near-far problem are investigated. Two realistic mission scenarios in the Low Earth Orbit (LEO) and in the Highly Elliptical Orbit (HEO) are analysed in section 6.4 to address the effect of the multiple access interference on the communication performance as well as on the navigation accuracy.

6.1 Dedicated network architecture requirements

Several proposed concepts can be found in the literature about potential network architectures for formation flying missions in space. Bristow et al. (2000) proposed a concept called Operating Missions as a Node on the Internet (OMNI) that regards spacecraft as network nodes and uses TCP/IP protocols to create a robust inter-satellite communication infrastructure. Similar proposals of using internet protocols also include Slywczak (2004), Hogie et al. (2005) and Wood et al. (2007). Vladimirova et al. (2007, 2008) discussed the potential of applying WiFi or WiMax protocols for the establishment of a space wireless sensor network. Clare et al. (2005) suggested to use Ad-hoc in conjunction with WiFi for supporting the high dynamics of spacecraft in formations with large-scale number of nodes. They all take advantages of utilizing existing terrestrial protocols and trying to apply them in space. The benefits are the compatibility with ground infrastructures and the good performance in terms of the large data throughput. However, the above two advantages are not the primary concern in a formation flying mission that requires precise navigation and tight control for the formation acquisition and maintenance.

Most of the flown or proposed formation flying missions involve the acquisition and maintenance of spacecraft in the desired relative geometric configuration in order to create a large virtual spaceborne instrument, such as for applications in

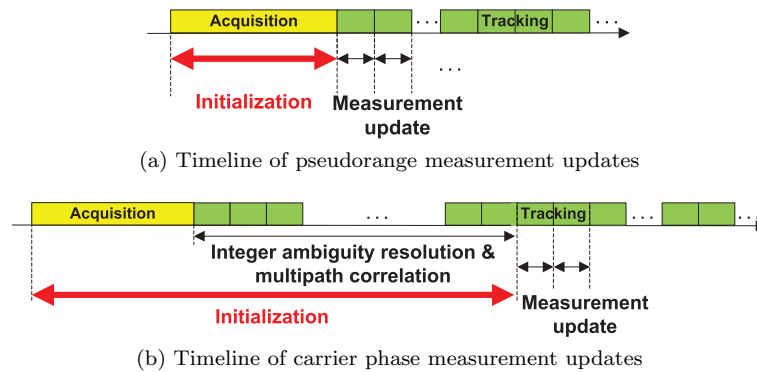


Figure 6.1: Timeline of pseudorange and carrier phase measurement updates

remote sensing and interferometry. The inter-satellite navigation data shall then be exchanged timely to enable the estimation of relative states between two spacecraft. This requirement is referred as to time-critical requirement. In addition, the formation network is expected to operate across various mission phases, ranging from the flexible initial formation built-up phase, the formation acquisition to desired geometry phase to the formation maintenance phase.

Therefore, the time-critical requirement and the flexibly operational requirement across all mission phases are treated as two dedicated formation network requirements, which will be further described in the following.

6.1.1 Time-critical requirement

The time-critical requirement is driven by the nature of tight formation control, collision avoidance or scientific needs in some specific formation operational periods, when a high relative navigation is required.

As introduced in previous chapters, the RF-based inter-satellite ranging system provides measurements for relative navigation. Those measurements include the unambiguous coarse pseudorange measurements and the ambiguous precise carrier phase measurements. Figure 6.1 depicts the measurement update timeline. Pseudorange measurements are used alone in the coarse-mode distance estimation when the navigation acquisition is in the order of several meters, while both the pseudorange and carrier phase measurements shall be utilized for the LOS estimation of less than 1° accuracy and the fine-mode distance estimation of cm-order accuracy.

It can be seen in Figure 6.1 that measurements are yielded after variable initialization periods, including the signal acquisition period for both of the code and carrier phase measurements, and some extra time only for the carrier phase measurement to resolve integer ambiguities and correct phase multipath.

In the acquisition period, a long signal detection process across a bi-dimensional searching space is performed for acquiring coarse code delay and carrier Doppler. For each potential code delay and Doppler combination, a sufficient long signal integration is often needed as well in order to achieve a sufficiently high carrier to noise ratio for signal and noise identification. Therefore, signal acquisition is a time-consuming process in the order of several seconds. However, this time can not be comparable to the time required in the integer ambiguity resolution process,

which may take tens of minutes or even hours and sometimes require relative motions between spacecraft (Barrena et al., 2008). Taking the PRISMA mission for example, it takes around 5 mins and 10 mins to solve the integer ambiguities in the LOS model and in the inter-satellite distance model, respectively, with the help of a sufficiently large relative geometry change between spacecraft (Barrena et al., 2008). A LOS constrained integer ambiguity resolution method, proposed in chapter 3, enables a much faster ambiguity estimation in the absence of large multipath. However, as demonstrated in chapter 5, the first ambiguity fix is still time consuming if the carrier phase measurement is contaminated by multipath.

After initialization, the tracking process will continuously run until the link is ended when switching the communication channel from one pair of spacecraft to another pair. At that time, re-initialization needs to be performed, including the corresponding signal re-acquisition in both the coarse- and fine-mode and the integer ambiguity re-initialization in the fine-mode. This process consumes precious time, especially for the fine-mode operations, and could result in a period that the on-board navigation filter is propagating without the measurement update. Such channel switching and re-initialization will lead to a reduced navigation accuracy.

This issue is referred to as time-critical requirement in the formation network. The network architecture design shall accommodate this time-critical requirement and give a high priority to timeliness rather than the traditional network consideration on the data throughput.

6.1.2 Flexible operations across all mission phases

Recognizing that the relative navigation requirements may change during the course of mission's operations, the network architecture design shall then address various phases of formation accuracy.

Figure 6.2 illustrates the evolutionary phases of a formation flying mission. In the initial deployment phase where the spacecraft are separated by substantial distances from one another, the resolution of relative position and attitude can be preformed based on coarse-mode measurements for collision avoidance, enabling further movement toward the desired configuration to take place safely. Spacecraft can be seen as free flyers located at a wide range of inter-satellite distances to each other. They will randomly access to the network. Neither a solely centralized nor a distributed topology is efficient in such situation, as some spacecraft are possibly out of the communication range of others.

As the spacecraft continue to aggregate into the desired spatial arrangement, they will eventually discover other spacecraft, which may be itself isolated or already be a member of a multi-spacecraft network. This condition is defined as formation acquisition and depicted in the centre of Figure 6.2. The formation acquisition phase is in progress until all spacecraft are connected in a single network and moving towards the desired formation configuration.

Finally, when all spacecraft in the system show a "complete connectivity" and are settled into the desired pattern, the formation maintenance will be performed as shown in the bottom of Figure 6.2. A higher accuracy of their relative position is acquired, enabled by switching the inter-satellite system into the fine-mode. A precise formation can then be achieved using tight control loops. Science operations will take place for, e.g., multiple point remote sensing. At this moment, the mission topology will evolve to a centralized graph with one spacecraft being the functional reference for a certain time period to enable autonomous relative naviga-

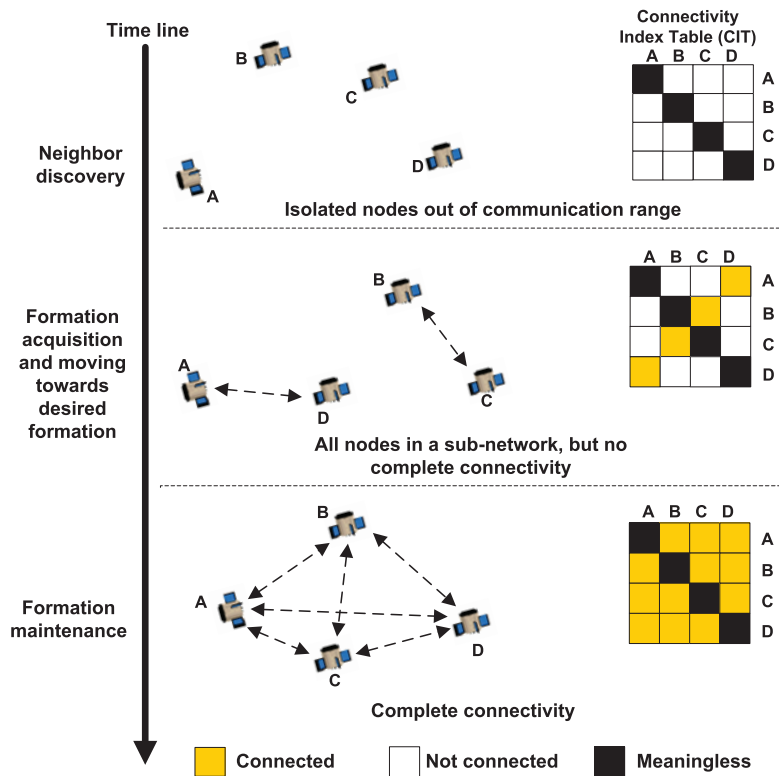


Figure 6.2: Evolutionary phases of a formation flying mission

tion and formation control. It shall be mentioned here that in the final topology, the “complete connectivity” does not mean the spacecraft have to connect to all other spacecraft, but shall connect to the desired spacecraft (often, the closed ones), to satisfy a certain formation configuration.

Subsequently, the formation can be reconfigured to set up a new arrangement for another science objective. The reconfiguration operation will drop back to coarse formation mode to prepare for the new arrangement, whereupon the precise formation can again be executed (Clare et al., 2005).

A connectivity index table (CIT) is proposed to be part of traffic exchanged among spacecraft in order to share the current network condition. The measured relative distance and LOS information can also be included in the CIT, in which way a spacecraft is able to calculate all relative positions among the network even though not all of them are directly connected.

6.2 Candidates for network architectures

As a result of the time-critical requirement, the access to the shared communication channel by multiple nodes in the formation network is suggested to be assigned in a deterministic manner, meaning that the transmission uncertainty, e.g., by the adaptive channel detection, shall be avoided. To this end, multiple access (MA) technologies, including the frequency, time and code division multiple access (FDMA, TDMA and CDMA) are potentially applicable to assure each spacecraft obtaining

Table 6.1: Comparison of different multiple access (MA) technologies for inter-satellite communication and navigation

MA	Advantages	Disadvantages
FDMA	<ul style="list-style-type: none"> • Multiple transmissions among S/C¹ can occur simultaneously • No complex timing schedule is needed 	<ul style="list-style-type: none"> • The larger is the number of S/C, the wider is the frequency band allocation required • Needs frequency isolation between sub-frequency bands to mitigate mutual frequency interference • Requires complex bandpass filters to separate sub-frequencies • Is more costly due to frequency variations • Needs to separate large power signals from the S/C in close proximity • May require power control
TDMA	<ul style="list-style-type: none"> • Single frequency • Easily separates S/C using simple timing logic • Prevents interference from other S/C completely 	<ul style="list-style-type: none"> • Inter-satellite communication can only occur at specific time slots • The overall throughput performance is reduced since each S/C must wait their turn to access the shared frequency • Time synchronization is needed • Signal transmission delay varies along with the different separation distances between S/C • Needs trade-off for the proper time slot to avoid signal collision and guarantee efficient channel occupation as well • The greater is the number of S/C, the longer is the duty cycle of a TDMA sequence
CDMA	<ul style="list-style-type: none"> • Multiple transmissions among S/C can occur simultaneously • GPS-like inter-satellite navigation is easy to operate simultaneously with communication 	<ul style="list-style-type: none"> • Limits the maximum number of S/C due to the cross-correlation interference • Has near-far problem: different separation distances between S/C cause various signal power levels

¹S/C: spacecraft

measurement updates from each of the others equally and timely.

In addition, the network topology, including centralized (star), fully distributed (mesh) and hierarchical forms determine the data flow and the navigation priority within the network. The consideration on topologies shall cope with various phases of formation operations.

Table 6.1 and 6.2 provide brief overviews on different multiple access technologies and different topologies, respectively. Their advantages and disadvantages are listed for further discussions on the design of the network architecture.

As mentioned in Table 6.1, among different MA technologies, FDMA is the most uneconomic choice since it needs a large frequency bandwidth and a complex filtering strategy to isolate sub-frequency bands for mutual frequency interference avoidance. TDMA with a simple timing schedule is suitable for the formation network with a small number of spacecraft. The greatest challenge of TDMA is the time synchronization, which shall be implemented in order for each spacecraft to know when to send and receive data on the TDMA sequence. An alternative solution to the time synchronization in TDMA is to simply reserve a time gap between two slots. A proper time slot and time gap are needed to be chosen to minimize the signal collision avoidance and also improve the channel occupation efficiency. CDMA is

Table 6.2: Comparison of different network topologies used in the formation network

Topologies	Advantages	Disadvantages
Centralized (star)	<ul style="list-style-type: none"> • Simple design • Has $N-1$ connections for N nodes 	<ul style="list-style-type: none"> • Relies on the capability of the central resource • Has a single point of failure: potential faults in the central resource greatly influence the whole mission
Distributed (mesh)	<ul style="list-style-type: none"> • Supports direct interactions among all nodes • Is fault tolerate for each node • Supports real-time communication to each node 	<ul style="list-style-type: none"> • Has $N(N-1)/2$ connections for N nodes • Suffers from a rapid growth in complexity as N increases • Limits in resource such as communication bandwidth and processing capability
Hierarchical¹ (hybrid)	<ul style="list-style-type: none"> • Guarantees robustness 	<ul style="list-style-type: none"> • Control structure complexity depends on the functional relationship between S/C • Needs multilevel approach • Is not necessary for the small scale mission

¹An example of the hierarchical topology in a network with a large number of S/C may consist of two or more centralized sub-networks, where the hubs of sub-networks are connected to each other.

an efficient multiple access method, especially when the inter-satellite communication and navigation are combined. However, CDMA performance decreases as the number of spacecraft increases due to the cross-correlation interference. The same constraint of having limited number of spacecraft also applies to TDMA, but here the reason is that a large number of spacecraft makes the cycle period of the TDMA sequence too long for data to be broadcasted with a high update frequency (especially for navigation data/measurements exchange that may require higher update rate).

Regarding various network topologies listed in Table 6.2, neither a solely centralized nor distributed topology is efficient across all mission phases of operations. For example, during the neighbor discovery and formation acquisition phases, some spacecraft are possibly out of the communication range of others. The centralized or distributed topology is thus not flexible enough to account for the node's joining in or dropping out of the network. On the other hand, as the spacecraft progress towards the desired formation configuration, it is better to evolve to a logical centralized graph in order to enable at least one spacecraft as reference for the precise relative navigation and formation control. The role of reference can rotate from one spacecraft to another to increase robustness and also avoid the single point of failure.

Rotating the role of reference at different time slots can enable robust and efficient connectivity. It can be implemented in a TDMA sequence with a deterministic timing boundary or a CDMA configuration with an adjustable time slot. These two possible arrangements are illustrated in Figure 6.3 and 6.4.

Recalling the system design in chapter 2, a transceiver on each spacecraft will contain a TX/Rx antenna for the dual one-way distance estimation and several additional Rx-only antennas for the LOS estimation. Figures in 6.3 and 6.4 also depict multiple antenna based inter-satellite links in order to clarify signal transmissions for both the distance and LOS estimation during different time slots.

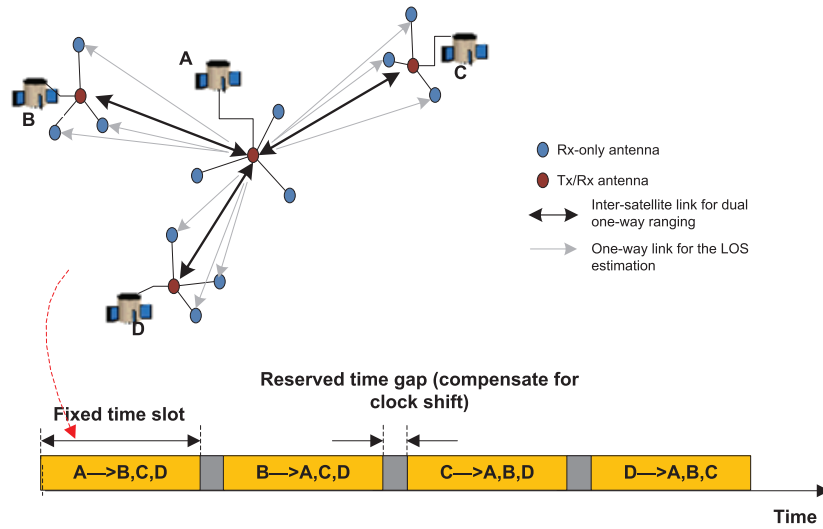


Figure 6.3: Implementation of TDMA with deterministic time slots

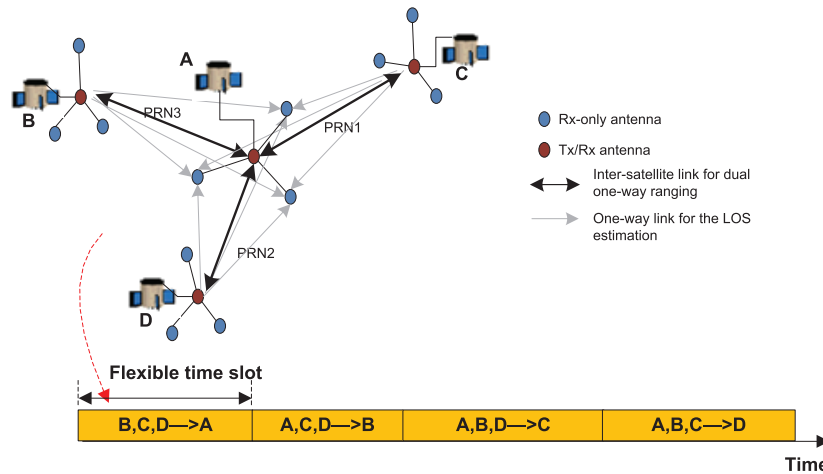


Figure 6.4: Implementation of roles rotating CDMA with flexible time slots

6.2.1 TDMA with deterministic time slot

In the TDMA mode in Figure 6.3, one of the spacecraft is chosen as reference and transmits ranging signals and command data to all other spacecraft in a fixed time slot. The role of being reference is sequentially passed from one spacecraft to another in a TDMA cycle. The complete TDMA cycle consists of equally distributed N time slots for N spacecraft and $N-1$ reserved time gaps between time slots for clock drift compensation. Reserving a time gap is a simple substitute to avoid clock synchronization in the TDMA mode. A proper allocation for the time slot and time gap is crucially important for network operation. On one hand, the time slot shall be large enough to cope with the time-consuming initialization process (especially the fine-mode integer ambiguity resolution process). Although the on-board navigation filter (based on Kalman filtering) can run in the absence of ranging measurement

updates, the price of a decreased navigation accuracy has to be paid. On the other hand, the time slot shall be sufficiently small in order to avoid a too-lengthy TDMA cycle. In addition, a proper time gap between two slots shall also be assigned to account for the clock drift and avoid the waste of channel occupation efficiency.

In this TDMA mode, the time slot is fixed, presenting limited ability of working during different mission phases. This is due to the fact that the signal transmission time may largely change across different mission phases as the change of the accessible number of nodes and the change of the inter-satellite range diversity.

6.2.2 Roles rotating CDMA with flexible time slot

To cope with different mission phases, it is better to have an adjustable time slot. Therefore, a network architecture called roles rotating CDMA with flexible time slots is proposed and illustrated in Figure 6.4. Due to the usage of CDMA strategy, multiple spacecraft can send communication data and ranging signals simultaneously to the single central reference spacecraft in a certain period (also called time slot here) until there is a need to rotate the role of being reference from one spacecraft to another. Within a certain time slot, the reference spacecraft is able to resolve relative positions and attitudes of all other spacecraft simultaneously. The time slot is adjustable to be either a long period to account for fine-mode initialization or a short period to implement only the coarse-mode ranging, depending on the requirements of different mission phases. In addition, signal transmissions in a certain time slot are not necessary to start or stop at the same time due to the code-based multiplexing instead of time-based multiplexing. It is thus tolerable if a spacecraft is joining in or dropping out of the formation, e.g., during the initial neighbour discovery phase.

Considering the robustness and flexibility of the roles rotating CDMA as opposed to the TDMA, this chapter focuses more on the discussion of CDMA. The limitations of CDMA stem from the well-known multiple access interference and near-far problem, which are described in the sequel from both the communication and navigation perspectives in order to provide a thorough recommendation for future missions.

6.3 CDMA limitations: multiple access interference and near-far problem

The multiple access capability of a CDMA network is achieved by using the GNSS-like PRN code. However, as the PRN code is not a completely orthogonal signalling format, the cross-correlation is nonetheless present and induces errors in terms of the multiple access interference (MAI).

6.3.1 Cross correlation without Doppler effect

Assuming two un-correlated PRN ranging signals $c_1(t)$ and $c_2(t)$ being received by a single receiver, they have identical spectrum $G(f)$ if received at the same power level of P (Spilker, 1996). The MAI term is introduced to characterize the cross-correlation (CC) between these two signals. If $c_1(t)$ is the desired signal, $c_2(t)$ can be treated as the interference and vice versa. Disregarding the Doppler effect for simplicity, the MAI term can be written as $c_1(t - \tau_1)c_2(t - \tau_2)$ with the code delay of τ_1 and τ_2 and code offset of $\Delta\tau = \tau_1 - \tau_2$. The MAI power spectrum $G_{MAI}(f)$ can

be obtained by convolving the identical individual signal spectrum $G(f)$ (Spilker, 1996)

$$G_{MAI}(f) = P \int G(v)G(f-v)dv \quad (6.1)$$

where $G(f)$ is in the form of sinc^2 (Spilker, 1996).

Only the MAI spectrum near $f = 0$ is important because the correlation filter has a small bandwidth on the order of Hz. According to (Spilker, 1996),

$$G_{MAI}(0) = P \int G_s^2(v)dv = P \int_0^\infty \left(\frac{\sin \pi f/f_c}{\pi f/f_c} \right)^4 df = \alpha \frac{P}{f_c} \quad (6.2)$$

where f_c is the chipping rate, α is a coefficient as a function of the filtered spectrum of sinc^2 . If the spectrum includes all of its mainlobe and sidelobes, α is 2/3. If the spectrum is filtered to include only the mainlobe, α increases to approximately 0.815 (Spilker, 1996).

Now, assuming M spacecraft at the same separation distance with respect to the reference spacecraft, $M-1$ interfering signals will be received, introducing CC errors in conjunction with the thermal random noise with a noise spectrum density of N_0 . An equivalent noise density $N_{0,eq}$ and an effective energy per bit to noise density ratio $E_b/N_{0,eq}$ can be written as

$$N_{0,eq} = N_0 + \alpha(M-1) \frac{P}{f_c} \quad (6.3)$$

$$\frac{E_b}{N_{0,eq}} = \frac{PT_d}{N_0 + \alpha(M-1)P/f_c} \quad (6.4)$$

where $T_d = 1/f_d$ with f_d as the data bit rate, $E_b/N_{0,eq}$ determines the bit error rate (BER). When the BER is 10^{-5} , $E_b/N_{0,eq}$ is around 10 dB using the BPSK modulation without extra coding for error correction (Wertz and Larson, 1999).

Further taking into account of various separation distances between spacecraft, a so-called near-far problem will show up. More specifically, if interfering signals are sent by near-by spacecraft while the desired signal is sent by a remote spacecraft, the effective $E_b/N_{0,eq}$ will significantly decrease. Since the free space loss is proportional to the square of distance, the MAI spectrum density of Eq. (6.2) for near interfering signals shall be multiplied by a factor of R_f^2/R_n^2 with R_f and R_n as signal transmission distances from the far desired signal and near undesired interference, respectively. The effective $E_b/N_{0,eq}$ can be revised to

$$\begin{aligned} \frac{E_b}{N_{0,eq}} &= \frac{PT_d}{N_0 + \alpha(M-1)(R_f^2/R_n^2)P/f_c} \\ &= \frac{P_s}{N_0 f_d (1 + \alpha(M-1)(R_f^2/R_n^2)P/(f_c N_0))} \\ &= \frac{E_b}{N_0} \left(\frac{1}{1 + \alpha(M-1)(R_f^2/R_n^2)P/(f_c N_0)} \right). \end{aligned} \quad (6.5)$$

Obviously, the multiple access effect of $M-1$ near interferences degrades the original E_b/N_0 by a factor of $1 + \alpha(M-1)(R_f^2/R_n^2)P_s/(f_c N_0)$.

It shall be noted that the signal power P equals to $f_d E_b$. Assuming that the code chipping rate f_c is 1.023 MHz, the data bit rate f_d is 2 kbps, the original E_b/N_0 is

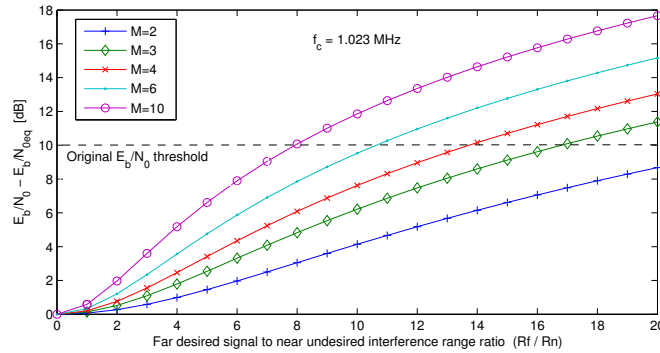


Figure 6.5: Energy per bit to noise density degradation due to the MAI effect using the BPSK-R(1) code

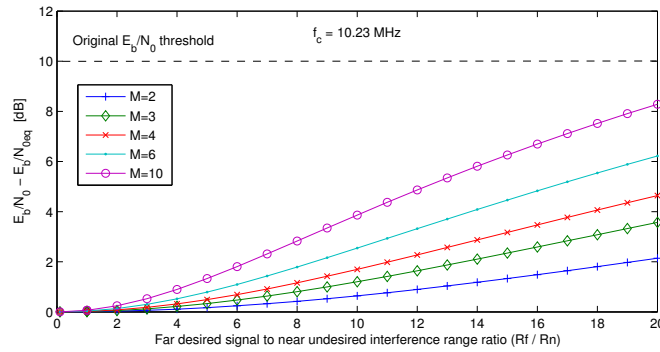


Figure 6.6: Energy per bit to noise density degradation due to the MAI effect using the BPSK-R(10) code

10 dB, and the coefficient α is 0.815 in case of taking account of only the mainlobe spectrum, the difference between the original E_b/N_0 and $E_b/N_{0,eq}$ caused by the MAI effect is calculated and depicted in Figure 6.5.

As shown in Figure 6.5, for a small scale network and small distance diversity, e.g., only one or two interfering spacecraft ($M=2$ or 3) and $R_f/R_n = 2$, the energy per bit to noise density degradation from E_b/N_0 to $E_b/N_{0,eq}$ is less than 1 dB and can be negligible. However, as the number of spacecraft becomes larger or the far desired signal to the near interference range ratio becomes higher, the degradation turns to be unacceptable, which could be beyond the original E_b/N_0 threshold. In other words, the MAI effect in the formation network limits the range diversity as well as the maximum number of spacecraft. As comparison to the BPSK-R(1) code with the chipping rate of 1.023 MHz, the usage of the BPSK-R(10) code in Figure 6.6 exhibits a smaller degradation.

Note that the results here focus on the communication performance and based on the assumption that the Doppler effect on the cross-correlation is ignored. When taking into account the relative navigation performance, the Doppler effect should be accounted in. As will be explained in the sequel, significant navigation errors can show up even within a small number of spacecraft network at low range diversity.

6.3.2 Cross correlation at high Doppler offset

Taking into account of Doppler frequencies $f_{D,1}$ and $f_{D,2}$ on the assumed desired signal $c_1(t)$ and interference $c_2(t)$, the Doppler offset refers to as the difference between these two Doppler frequencies $\Delta f_{D,12} = f_{D,1} - f_{D,2}$. The cross-correlation term $R_{cc}(\Delta\tau)$ and its spectrum density $G_{MAI}(f)$ will turn to

$$R_{cc}(\Delta\tau) = \int_0^T c_1(t)c_2(t - \Delta\tau) \cos(2\pi\Delta f_{D,12}t + \Delta\phi_{12})dt \quad (6.6)$$

$$G_{MAI}(f) = P \int G(v)G(f + \Delta f_{D,12} - v)dv \quad (6.7)$$

where T is the integration time, $\Delta\tau$ and $\Delta\phi_{mk}$ are the code delay offset and phase offset. Due to Doppler effect, the interference signal structure $c_2(t)$ has been changed to a multiplication of $c_2(t)$ and $\cos(2\pi\Delta f_{D,12}t + \Delta\phi_{12})$, indicating that the interference spectrum is shifted by $\Delta f_{D,12}$. The CC error becomes significant when $\Delta f_{D,12}$ is close to zero (same as Eq.(6.2)) or an integer multiple (n times) of f_c/L ($n = \pm 1, \pm 2 \dots$), where L is the sequence length. This phenomenon has been observed and demonstrated by researchers and was referred to as Doppler crossover (Kaplan and Hegarty, 2006; Misra and Enge, 2001; van Dierendonck et al., 1999). For the C/A code, as $f_c/L = 1$ kHz ($f_c/L = 1.023\text{MHz}/1023$), the CC error is significant with zero and n-kHz Doppler crossover.

The phenomenon of Doppler crossover stems from the fact that a PRN code (e.g., C/A code) has a limited sequence length (e.g., $L = 1023$) and is periodically repeated (e.g., every 1 millisecond). The C/A code spectrum is actually not continuous but composed by 1 kHz separated line components within the sinc² envelope (Spilker, 1996). The line component in the center is at zero frequency, while the others are symmetrically located at the positive and negative frequencies with the 1 kHz spacing between adjacent lines. Therefore, for the Doppler offset $\Delta f_{D,12} \neq n f_c/L$, the line components of the desired signal spectrum $G(v)$ and the shifted interference signal spectrum $G(f + \Delta f_{D,12} - v)$ do not overlap, thus the cross correlation spectrum by multiplying them will diminish. On the contrary, mixing at the existing line frequencies at $n f_c/L$ will result in the interference being minimally suppressed. That is, if the Doppler offset is an integer multiple of the line component spacing f_c/L , the cross-correlation noise energy ‘‘leaks’’ through the correlation process and will thus result in large cross-correlation errors.

In order to estimate the magnitude of the CC error and its influence on the tracking accuracy in the receiver, the discriminator output as the function of the code tracking error needs to be analysed. Under the assumption that the auto-correlation is much larger than the cross-correlation, the normalized early-minus-late coherent discriminator D can be written as (Zhu and van Graas, 2005)

$$\begin{aligned} D &= \frac{\sqrt{I_E^2 + Q_E^2} - \sqrt{I_L^2 + Q_L^2}}{\sqrt{I_E^2 + Q_E^2} + \sqrt{I_L^2 + Q_L^2}} \\ &\approx \frac{R_{cc}(-dT_c/2) - R_{cc}(dT_c/2)}{2R_{ac}(-dT_c/2)} \\ &= \frac{\int_0^T c_1(t) [c_2(t + dT_c/2) - c_2(t - dT_c/2)] \cos(2\pi\Delta f_{D,12}t + \Delta\phi_{12})dt}{2R_{ac}(-dT_c/2)} \end{aligned} \quad (6.8)$$

where d is the early-late spacing in chips, I and Q denote the in-phase and quadrature arms, and subscriptions E and L indicate the early and late correlators, $R_{cc}(\pm dT_c/2)$

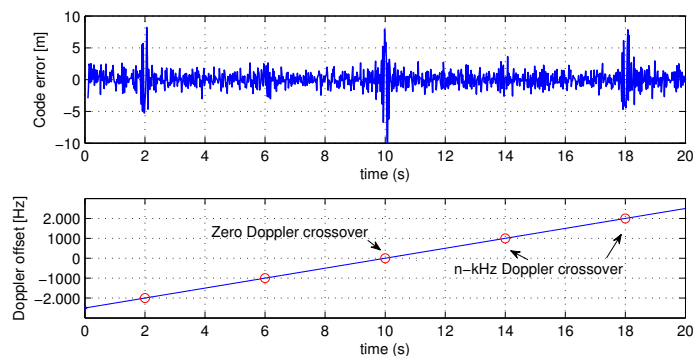


Figure 6.7: Code error in the presence of cross-correlation (top); Doppler offset between the desired and interfering signals is assumed changing over time from -2500 Hz to 2500 Hz (bottom)

and $R_{ac}(\pm dT_c/2)$ are the cross-correlation and auto-correlation with early or late delays of $\pm dT_c/2$. It is obvious that the discriminator output is related to the early-late spacing d . Similar to the error pattern of multipath, for a small d , $c_2(t + dT_c/2)$ and $c_2(t - dT_c/2)$ are dependent and the common part could be cancelled out (Zhu and van Graas, 2005).

Simulations via the software-defined receiver are performed to demonstrate the Doppler crossover effect on the code tracking accuracy. Assume the Doppler offset linearly increases over time from -2500 Hz to 2500 Hz. Only one interference signal is assumed, which is at the same power level as the desired signal. C/A code is used in the simulation. The receiver is configured to work with an integration time of 20 ms and an early-late spacing of 0.8 chips.

From Figure 6.7, the zero and n-kHz Doppler crossover phenomenon are easily observed. Different crossover points have different error magnitudes. At the zero and 2-kHz crossover points, the code error suddenly changes up to 5 m, while the 1-kHz crossover point only introduces errors of approximately 2 m. The error pattern follows a sinusoidal oscillation around the crossover point. A sensitive zone of ± 25 Hz around the crossover point can be regarded as the largest cross-correlation error zone (Zhu and van Graas, 2005). According to Eq.(6.6), the cross-correlation error is dependent on the PRN code pattern and the code delay. RRN 7 and 22 are used in this simulation with the code delay of 923 chips and 204 chips in the C/A code sequence of in total 1023 chips.

It should be noted that the error magnitude is also affected by the code Doppler. Similar to the carrier Doppler offset, the code Doppler offset slowly changes the relative delay between the desired and interfering signals, resulting in a slightly enlarged or lessened error magnitude as opposed to the case without the consideration of the code Doppler.

6.3.3 Near-far problem at Doppler crossover

The well-known near-far problem not only deteriorates the E_b/N_0 performance as discussed in section 6.3.1, but more seriously exacerbates the navigation accuracy, especially when it occurs at the moment of the Doppler crossover.

Suppose the Doppler offset is 1 Hz, which is inside the sensitive zone of the zero-Doppler crossover. Figure 6.8 illustrates the simulation results of the cross-

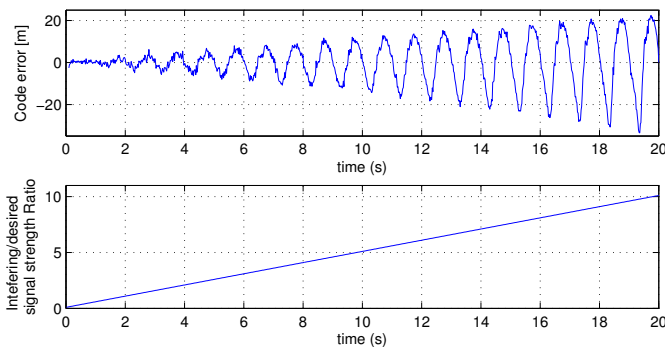


Figure 6.8: Code error in the presence of the cross-correlation at 1 Hz Doppler offset (top); Interference-to-desired signal strength ratio is assumed changing over time from 0.1 to 10 (bottom)

correlation error when the interfering signal strength to the desired signal strength ratio linearly increases over time. It is obvious that the magnitude of the cross-correlation error increases linearly and it has a sinusoidal pattern at the frequency of 1 Hz.

This result indicates that the magnitude of the cross-correlation error is proportional to the interfering to desired signal strength ratio. To this end, a high range diversity in the formation flying network will lead to a large cross-correlation error when it occurs within the Doppler crossover sensitive zone.

6.4 Case-studies

6.4.1 Case-study set-up

The software-defined signal simulator and receiver are used to investigate MAI errors in pseudorange measurements.

The simulator, with its architecture depicted in Figure 6.9, generates desired and interfering signals from four satellites. The Doppler offset and the interfering to desired signal strength ratio are determined by specific formation scenarios according to their relative dynamics. All interferences are summed up onto the desired signal. White noise is also added before the compound signal enters the bandpass filter and quantization chain. This simulator emulates the MAI scenario and also considers the receiver front-end conditioning process so that a digitalized noisy intermediate-frequency (IF) signal can be produced for signal processing in a software receiver.

The acquisition and tracking are implemented in multiple channels in the software receiver, see Figure 6.10. Channels are isolated by orthogonally multiplexed PRN codes. Code errors in presence of MAI effects can be extracted from tracking loops. Extensive descriptions on the acquisition and tracking can be found in chapter 2. Basic parameters used to configure the software simulator and receiver are specified in Table 6.3, while the MAI-dependent parameters, including the Doppler offset and the relative signal strength ratio between satellites, will be determined by specific mission scenarios.

In the following, we will analyse two realistic formation scenarios as examples to investigate how the navigation performance is degraded by MAI effects.

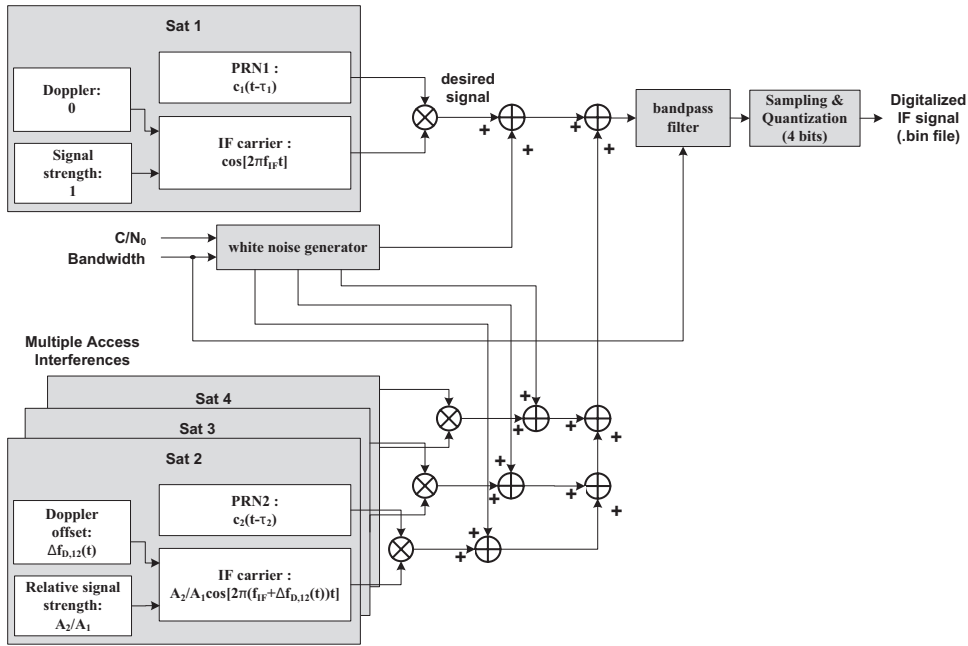


Figure 6.9: Software simulator setup for the demonstration of MAI effects. Here, the Doppler shift on the desired signal (Sat 1) and interference signal (Sat 2 for example) are assumed 0 and $\Delta f_{D,12}(t)$, respectively, while in reality they shall be $f_{D1}(t)$ and $f_{D2}(t)$ and satisfy $\Delta f_{D,12}(t) = f_{D2}(t) - f_{D1}(t)$.

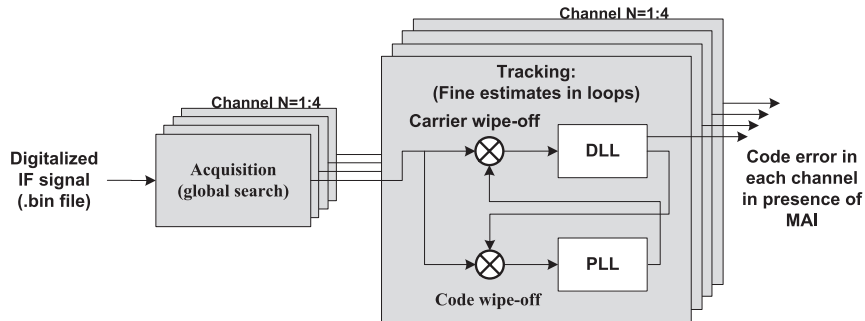


Figure 6.10: Software receiver setup for the demonstration of MAI effects

6.4.2 Circular low earth orbit formation scenario

A formation geometry in the Low Earth Orbit (LEO) with a chief satellite and several deputy satellites is a commonly used relative orbit geometry. In case of a Keplerian two-body motion, if the chief orbit is circular and the inter-satellite distance is much smaller than the chief's semi-major axis, their relative dynamics can be expressed in Clohessy-Wiltshire (CW) equations in a linear form in the Hill frame (Alfriend et al., 2010). The relative motion of the deputy with respect to the chief can be expressed using

$$\mathbf{x} = (\mathbf{r} \quad \mathbf{v})^T = (x \quad y \quad z \quad \dot{x} \quad \dot{y} \quad \dot{z})^T \quad (6.9)$$

Table 6.3: Specifications in the software simulator and receiver for the demonstration of MAI effects

Simulator	PRN code	BPSK-R(1)
	IF frequency [MHz]	9.5485
	Sampling frequency [MHz]	38.192
	Filtering Bandwidth [MHz]	4
	C/N_0 [dB-Hz]	60
	Doppler offset and relative signal strength	Determined by mission scenarios
Receiver	DLL discriminator	Normalized early-minus-late
	DLL early-late spacing [chips]	0.8
	DLL noise bandwidth [Hz]	2
	PLL discriminator	arctan
	PLL noise bandwidth [Hz]	20
	Integration time [ms]	1

where vectors \mathbf{r} and \mathbf{v} denote the relative positions and velocities in radial x , along-track y and cross-track z directions. A relative ellipse orbit can be created in a closed form periodic solution when the initial orbit elements satisfy (Montenbruck and Gill, 2000; Alfriend et al., 2010)

$$4x_0 + 2\dot{y}_0/n = 0 \quad (6.10)$$

$$y_0 - 2\dot{x}_0/n = 0 \quad (6.11)$$

where n is the orbital mean motion according to $n = \sqrt{\mu/a^3}$ with μ as the Earth's gravitational coefficient and a as the semi-major axis of the chief.

Suppose there are five satellites in the formation, one is chief and the others are deputies in two safe elliptical orbits. The initial relative orbit elements are given in Table 6.4. The orbit of the chief is circular with a semi-major axis of 7000 km. After the relative orbit propagation using CW equations, it can be seen in Figure 6.11 that two relative orbits of deputy satellites are coplanar, which have elliptical projections in the xy - and xz -plane and linear motions in the yz -plane. The ellipse for Sat 1 and Sat 2 has dimensions of $1 \times 1 \times 1$ km, while Sat 3 and Sat 4 are in a $1 \times 2 \times 2$ km ellipse. The inter-satellite distance of each deputy with respect to the chief is depicted in Figure 6.12 and has a sinusoidal pattern.

When the chief satellite receives signals from deputies at the same time using the CDMA technology, multiple access interference will occur that results from cross-correlation effects. As analysed in the last section, the cross-correlation error is dependent on the signal strength and Doppler offset. The signal strength ratio between interfering and desired signals can be easily calculated by inversely scaling the inter-satellite distance ratio, while the Doppler offset can be obtained according to

$$\begin{aligned} \Delta f_D &= \frac{f_{RF}}{c} v_p \\ &= \frac{f_{RF}}{c} \frac{\mathbf{r} \cdot \mathbf{v}}{\|\mathbf{r}\|} = \frac{f_{RF}}{c} \frac{\dot{x}x + \dot{y}y + \dot{z}z}{\sqrt{x^2 + y^2 + z^2}} \end{aligned} \quad (6.12)$$

where v_p is the relative velocity projected in the inter-satellite link direction. Only this part of velocity introduces Doppler effects to the inter-satellite communication. The carrier frequency of the signal f_{RF} is assumed 2271.06 MHz in the S-band, and c is the velocity of light.

Table 6.4: Initial relative orbit elements

	x_0 [m]	y_0 [m]	z_0 [m]	\dot{x}_0 [m/s]	\dot{y}_0 [m/s]	\dot{z}_0 [m/s]
Sat 1	1000	0	0	0	-2.156	-0.178
Sat 2	578.78	-1618.03	-809.02	-0.872	-1.267	-0.634
Sat 3	-1000	0	0	0	2.156	2.156
Sat 4	-578.78	1618.03	1618.03	0.872	1.267	1.267

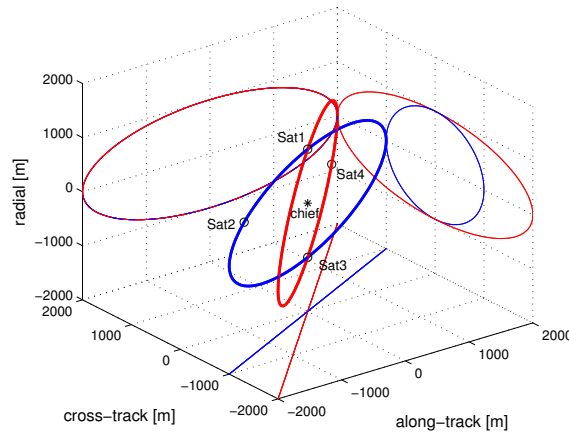


Figure 6.11: Sate ellipse trajectories of Sat 1, 2, 3 and 4 with respect to the chief. The cross and circle denote positions of the chief and deputies at their initial orbits

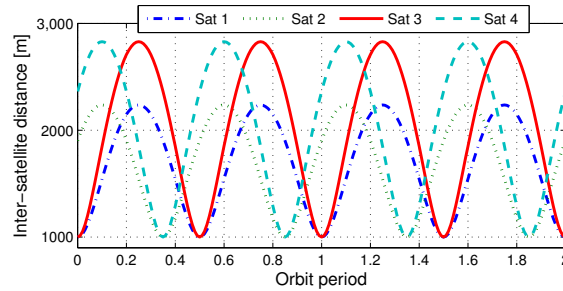
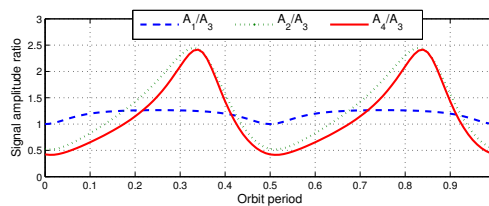


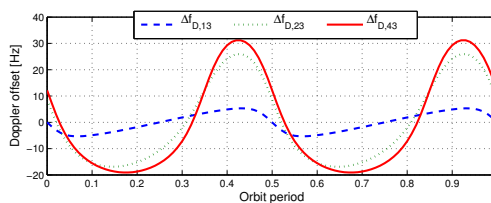
Figure 6.12: Inter-satellite distance of each deputy with respect to the chief

For the receiver on the chief satellite, multiple channels are allocated to track different PRN codes from deputies. Suppose extracting measurements for the relative motion estimation between Sat 3 and the chief, then the signal from Sat 3 is the desired signal and signals from Sat 1, 2, 4 are regarded as interferences. Figure 6.13 (a)(b) depict the signal strength ratio and Doppler offsets between interfering and desired signals. Figure 6.13 (c) illustrates the cross-correlation error extracted from the software receiver.

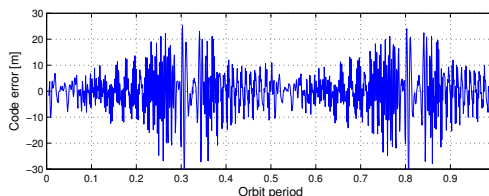
As can be seen, during the whole orbit period, the Doppler offset is within or quite close to the Doppler crossover sensitive zone of ± 25 Hz, resulting in large cross-correlation errors in general. More specifically, the error frequency is chang-



(a) Interfering to desired signal magnitude ratio caused by distance diversity



(b) Doppler offset



(c) Code errors in presence of cross-correlations from three interfering satellites

Figure 6.13: MAI effects in LEO safe ellipse formation geometry

ing faster as approaching to zero-Doppler crossover points. The error magnitude increases when the zero-Doppler crossover occurs at a high interfering to desired signal strength ratio. Maximum errors in Figure 6.13(c) show up in the measurement batch of $[0.3, 0.4]$ and $[0.7, 0.9]$ of an orbital period when zero-Doppler crossovers in $\Delta f_{D,23}$ and $\Delta f_{D,43}$ occur at the same moment of the signal strength ratio closing to the highest point. On the contrary, although zero-Doppler crossovers in $\Delta f_{D,23}$ and $\Delta f_{D,43}$ also occur at around 0.02 and 0.52 fraction of the orbital period, large cross-correlation errors did not show up due to the fact that the interfering signal strength is smaller than the desired signal strength at these moments.

6.4.3 Highly elliptical orbit scenario: MMS mission

The MMS (Magnetospheric Multiscale) formation is a NASA mission, which uses four identical satellites to make three-dimensional measurements of magnetospheric boundary regions and examine the process of magnetic reconnection (Volle et al., 2007).

Four identical satellites in the MMS mission are arranged in a tetrahedral geometry. The inter-satellite communication will be conducted in a distributed topology. Two distinct phases are divided in this mission. For Phase 1, the MMS will be in a 1.2×12 Earth Radii highly elliptical orbit with an orbital period of approxi-

Table 6.5: Orbital elements for Phase 1 in the MMS mission (Volle et al., 2007)

	Semi-major axis [km]	Eccentricity	Inclination [°]
Sat 1	42095.7	0.81818	27.8
Sat 2	42095.7000043072	0.81719081297	27.80015587911
Sat 3	42095.7000019023	0.81749305346	27.80520233720
Sat 4	42095.7000026211	0.81750706118	27.79359055330
	Argument of perigee [°]	Right ascension of the ascending node [°]	True anomaly [°]
Sat 1	15.000001	0	180
Sat 2	15.0184660490	0.00076380491	179.9921269275
Sat 3	15.0026369333	359.94611	180.0188885580
Sat 4	14.9042680950	0.060163692	180.0179095340

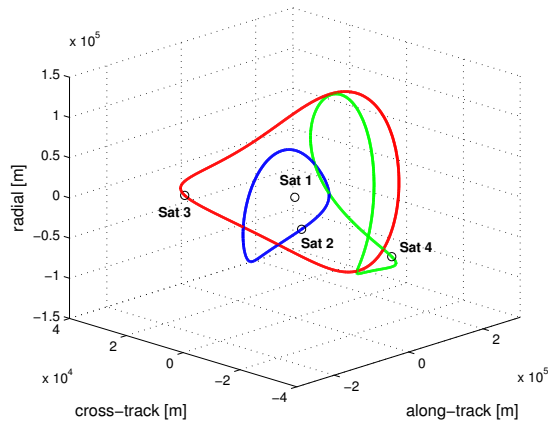


Figure 6.14: MMS mission relative trajectories of Sat 2, 3, 4 with respect to Sat 1. The circle denotes positions of four identical satellites at the initial orbit

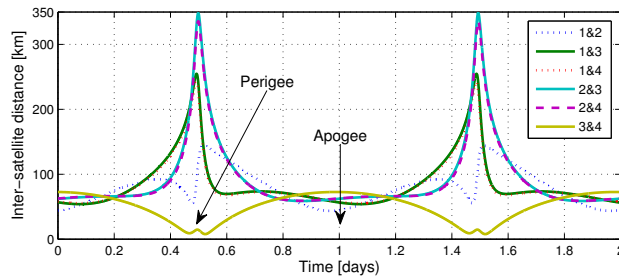


Figure 6.15: Inter-satellite distance between each pair of spacecraft in MMS mission

mately one day. The initial orbital elements are given in Table 6.5. Unlike the low earth circular orbit in the last scenario, the MMS mission cannot use linearized CW equations for relative orbit determination, but by propagating absolute orbits using absolute Keplerian dynamics and then determining relative motions in the Hill frame by transformation.

Figure 6.14 displays the relative trajectories of Sat 2, 3, 4 with respect to Sat 1 in order to provide a basic overview on how the tetrahedral formation is established. Figure 6.15 illustrates inter-satellite separations between each pair of spacecraft over two complete orbits. Near the apogee, the inter-satellite distance is about 60 km, while as approaching to the perigee, spacecraft separations vary dramatically, ranging from 10 km to 350 km.

This MMS mission, although physically having four purely identical and distributed spacecraft, can be a good example of implementing the CDMA with roles rotating architecture. Within a flexible time slot, one of the spacecraft can be regarded as the functional reference and receiving signals from the other three spacecraft simultaneously. The turns of being the reference rotate from one spacecraft to another. Figure 6.16 illustrates two examples of Sat 1 and Sat 4 as references, respectively, at two distinct time slots. The first example in Figure 6.16 (a) assumes that Sat 1 requires the desired signal from Sat 4 while is interfered by signals from Sat 2 and 3. The second example in Figure 6.16(b) assumes that Sat 4 receives the desired signal from Sat 1 and is interfered by Sat 2 and 3. These two examples have dramatically different interfering to desired signal strength ratios and Doppler offsets, thus may introduce different levels of MAI errors.

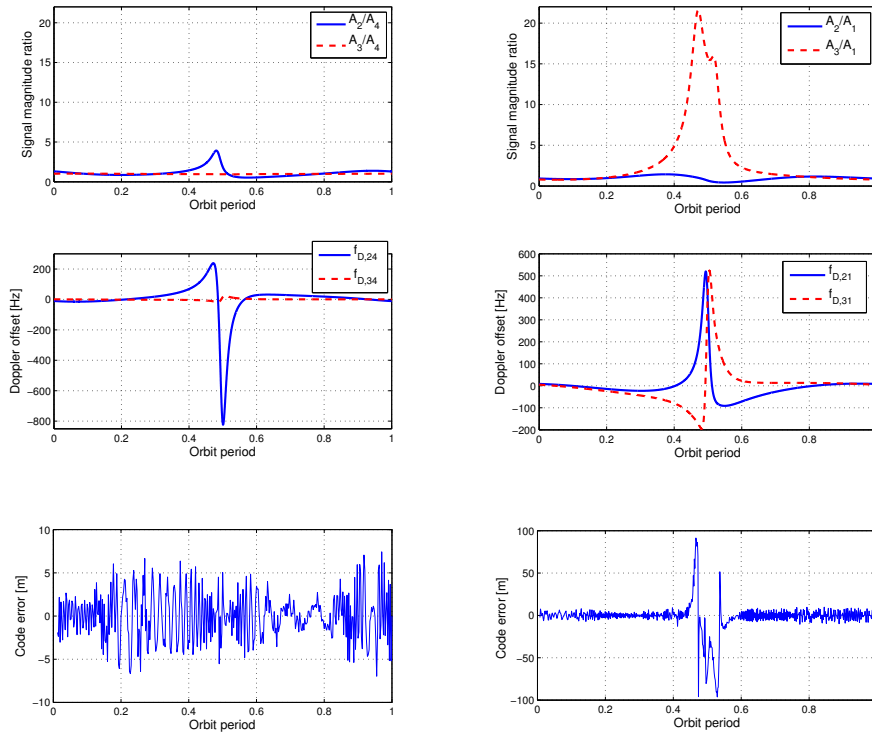
In Figure 6.16 (a), it is clear that the Doppler offset (in blue solid line) ranges from -800 Hz to 220 Hz, which is far beyond the sensitive zone of the Doppler crossover most of the time except for the crossing moments at around 0.20, 0.49, 0.58 and 0.92 fraction of orbital period, when large cross-correlation errors are observed in the bottom figure. Although the interference has a strength of up to 4 times larger than the desired signal at the duration of [0.38, 0.46] orbits, it does not introduce more errors since its Doppler offset is extremely far from crossovers. The visible errors at this duration actually result from another interference contributor (green dashed line) with its Doppler crossover at the moment of 0.5 fraction of orbital period. The Doppler offset of this interference is within the sensitive zone in a complete orbit period.

MAI effects in 6.16 (b) are more severe because of the widely diverse inter-satellite separations that can lead to severe near-far problem. Up to 100 m errors are visible for a measurement batch of [0.45, 0.55] orbits. During this period, a Doppler crossover at around 0.50 fraction of orbital period does happen, but luckily, it is very instantaneous and shall not introduce severe consequences to the period far from the crossing point. However, the first error spike shows up when the Doppler offset is around 200 Hz and the signal strength ratio reaches 13. This means that significant errors can occur at a high range diversity even if the corresponding Doppler offset is beyond its crossover sensitive zone.

6.4.4 Case-study summary

The effects of multiple access interference play important roles in defining the CDMA capability for the combined inter-satellite communication and navigation.

In the scenario of the safe ellipse orbit in the LEO where the corresponding Doppler offset is within the Doppler crossover sensitive zone for the whole orbit



(a) MAI effects in the time slot when Sat 1 is regarded as the reference, requiring the desired ranging signal from Sat 4 but interfered by Sat 2 and 3

(b) MAI effects in the time slot when Sat 4 is regarded as the reference, requiring the desired ranging signal from Sat 1 but interfered by Sat 2 and 3

Figure 6.16: MAI effects in the MMS mission

period, considerable cross-correlation errors will be introduced.

In the scenario of the MMS mission in the highly elliptical orbit, the Doppler offset is much higher, which is beyond the crossover sensitive zone for a substantial time of an orbit period. This shall result in smaller cross-correlation errors in general. However, during the period when the inter-satellite separation diversity is extremely high, severe errors can occur regardless of Doppler crossovers. In this period, an adaptive power control mechanism shall be implemented to minimize the impact of this near-far problem. It is necessary to lower the transmitting power of interfering signals when they are in close proximity.

For both scenarios, there is no n-kHz Doppler crossover taking place. However, if the system works in a higher carrier frequency, e.g., K-band or Ku-band, the chance of n-kHz Doppler crossover will be higher, which is also a source of large MAI errors.

The methods of mitigating MAI errors include an improvement of the code delay loop inside the receiver by using a smaller correlator spacing or a longer integration time. Long time carrier smoothing of the pseudorange measurements can also be used, but should be carefully taken into account in some tight control mission phases when a high measurement update rate is required.

6.5 Chapter summary

In this chapter, several network architectures were presented to support the RF-based inter-satellite communication and relative navigation in a large scale formation with four or more spacecraft. Two dedicated network requirements, being the time-critical requirement and the flexible operations across various mission phases, were proposed to drive the investigation and evaluation of different architectures.

The CDMA technology was modified to work in a centralized topology with the central reference being rotated from one spacecraft to another in adjustable time slots. This modification enables the system working with a wide range of flexibility, such as enabling both the coarse- and fine-mode navigation at different mission phases, allowing to detect some spacecraft while tracking others, and being insensitive to a spacecraft joining in or dropping out of the formation.

The limitation of using CDMA was also investigated in terms of the multiple access interference (MAI). The MAI was found Doppler dependent and suffers as well from the near-far problem. From a communication performance point of view, the energy per bit to noise density ratio is reduced as compared to the case without the MAI, leading to a limited operational range diversity and a limited maximum number of spacecraft in a formation. Furthermore, the MAI error worsens the navigation performance, especially at the moment of Doppler crossovers or in case of signals being corrupted by the near-by interferences.

Two case-study scenarios, one of a circular Low Earth Orbit mission and another for a highly elliptical orbit mission, were provided that demonstrated the severe effects MAI errors and the high probability of its occurrence within an orbit period. MAI errors easily exceed the meter level, which is suggested to be mitigated using a smaller correlator spacing or a longer integration time. A long time carrier smoothing also aids in minimizing MAI errors. However, it needs to be carefully taken into account in tight control periods when a high measurement update rate is required.

Chapter 7

Conclusions and Outlook

This final chapter provides a summary and conclusions from previous chapters. This is followed by an outlook, where topics for further research are proposed and discussed to enhance the development of future formation flying missions.

7.1 Summary

Satellite formation flying is an enabling revolutionary technology for scientific and commercial applications in space. The distribution of functions and payloads among fleets of coordinated small satellites offers the possibility to overcome the classical limitations of traditional single-satellite systems. The science return can be dramatically enhanced through observations made with larger and configurable baselines between/among satellites. Coordinating the alignment of baselines, i.e. estimating and controlling baselines, is typically required for formation acquisition and maintenance. This requirement is triggered by the needs of collecting multi-point scientific data or creating large spaceborne instruments such as telescopes and interferometers. A high level of accuracy for relative navigation and control is thus of critical importance.

A common way to perform relative navigation for formation flying missions is to utilize differential Global Navigational Satellite System (GNSS) measurements. This configuration could enable an accuracy better than one centimeter in certain cases, but is generally limited to Low Earth Orbit (LEO) missions. As opposed to the GNSS-based relative navigation that relies on the visibility of GNSS constellations, self-contained relative navigation metrology, i.e., through the transmission/reception of radio frequency (RF) signals via inter-satellite links, has attracted much attention recently. This RF-based metrology can serve as an argumentation or even substitution to GNSS-based metrology in high Earth orbits where GNSS constellations are poorly visible. A formation flying RF (FFRF) sensor has been developed, embarked, tested and validated on the PRISMA mission in LEO. This sensor achieved centimeter level accuracy for the inter-satellite distance estimation and one degree level accuracy for the line-of-sight estimation. The success of FFRF on PRISMA mission has boosted its future usage on high Earth orbit missions, e.g., PROBA-3.

Considering FFRF as a benchmarking system, this research aims at investigating key technologies of the self-contained RF-based relative navigation system to improve FFRF performance. Research questions (RQs) are proposed as follows.

- RQ1:** What is the architecture and functionality of an inter-satellite RF system?
- RQ2:** What algorithms shall be developed to enable relative navigation?
- RQ3:** How to improve the relative navigation performance in terms of accuracy, efficiency and reliability?
- RQ4:** How to apply relative navigation in a large-scale formation with four or more satellites?

The RF metrology in this research inherits classical GNSS technologies by transmission and reception locally generated GNSS-like pseudo random noise (PRN) ranging signals and carrier phases via inter-satellite links. The system architecture and functionality were extensively discussed in Chapter 2 (**RQ1**). The system is designed to comprise of one transmitter and one receiver with multiple antennas in S band. Two different PRN signal structures, BPSK-R and BOC, were comprehensively analysed and evaluated in terms of the lower bound accuracy, multipath performance and bandwidth occupation. Signal processing architecture was introduced through the signal conditioning, acquisition and tracking chain. A multi-antenna set-up was designed to enable both the inter-satellite distance estimation and the line-of-sight estimation (**RQ2**). The LOS estimation algorithms were elaborated in Chapter 3 (**RQ2**) together with the algorithms to improve associated carrier phase integer ambiguity resolution (IAR) efficiently and reliably (**RQ3**). An unaided, fast and reliable IAR was proposed, analysed, and validated by both numerical simulations and field tests. The antenna geometry impact on the algorithm accuracy and efficiency was also investigated in Chapter 3. From the PRISMA experience, multipath has been found a dominating error source, which may cause the failure of IAR and decrease the LOS estimation accuracy (**RQ3**). Potential solutions to deal with multipath were mainly discussed in Chapter 4 and 5 with respect to pseudorange and carrier phase multipath, respectively. Chapter 6 extended the previous scenarios and results for a large scale formation with four and more satellites (**RQ4**), and proposed network architectures including a discussion on their potentials and limitations. The novelties of this research will be specifically highlighted in the sequel.

7.2 Conclusions

Conclusions of this research are presented in the following by categorizing three main challenges with respect to the technologies:

- 1) to deal with multipath,
- 2) to perform an unaided, fast and reliable carrier phase integer ambiguity resolution, and
- 3) to share channels among multiple spacecraft.

Multipath: Potential solutions to deal with multipath were comprehensively discussed in Chapter 2, 4, and 5, respectively, in terms of evaluating different signal structures to assure multipath immunity, designing receiver-internal methods to mitigate pseudorange multipath, and developing a cascaded extended Kalman filter to estimate carrier phase multipath.

- Two different types of signal structures, BPSK-R and BOC, were discussed in Chapter 2. They exhibit various multipath immunity capabilities. Although

the BOC signal structure is a new generation of a PRN ranging code and has been demonstrated in the literature with a general better performance against multipath, it requires an advanced tracking strategy to distinguish multiple peak ambiguities in its auto-correlation function and is therefore computationally intensive. This research found that the usage of a conventional BPSK-R modulation with higher chipping rate, i.e. 10 Mcps, can not only avoid multi-peak ambiguity but also guarantee comparable multipath immunity capability as using BOC. This is especially true in space environments where typically short-delay multipath is introduced by the reflections from the dimension-limited spacecraft structures.

- A novel method, termed “Multipath Envelope Curve Fitting”, was proposed in Chapter 4 to mitigate short-delay pseudorange multipath. Compared to other state-of-the-art methods, this method is effective in mitigating short-delay pseudorange multipath by approximately 50%, and also shows comparable performance for medium and large delayed multipath. The method is proposed based on the fact that the signal strength, compound of a LOS signal and a single multipath component, exhibits an in-phase correlation with the multipath error. By linearly combining multiple signal strength estimators (from multipath correlators in the receiver), the pseudorange multipath error can be accurately estimated. A simple implementation strategy was also proposed that enables a receiver-internal multipath estimation process operated in conjunction with the tracking loop with a minimum computational effect. This method is specifically designed for a relatively “clean” multipath environment like Space. In an environment consisting of a large number of multipath signals, if there is not a dominating component among them, the superposition of multiple multipath with different delays and different phases will break the in-phase correlation. In this case, the curve fitting method will lose its effectiveness.
- Cascaded extended Kalman filters (EKF) were designed in Chapter 5 to estimate the carrier phase multipath on the fly. Most of the existing methods either assume fixed carrier phase integer ambiguities so as to use the phase residual for multipath estimation, or make use of signal-to-noise ratio (SNR) in a post-processing manner for multipath estimation. The proposed method in this research also uses SNR data as measurements, yet can process them in three cascaded EKFs in real time. The first EKF is used to filter out the noise on SNRs. The second successive EKF is used to estimate multipath parameters, which are then reformulated to construct the multipath errors in order to remove these errors from the phase measurements. The integer ambiguity resolution can be dramatically accelerated due to the multipath removal. After ambiguities are fixed, a third cascaded EKF is used as a combined LOS, LOS rate and multipath estimator, which guarantees an achievement of sub-degree LOS accuracy. Numerical simulations have shown that this proposed method will accelerate the multipath contaminated IAR process (Time to first fix) dramatically and allows for removing multipath from phase measurements almost completely.
- For the carrier phase multipath estimation, both the real-valued and complex-valued EKF were proposed to apply in the second and third cascaded EKFs. The complex-valued EKF has been found to be insensitive to poor initial condi-

tions when the real-valued EKF fails to fast converge. Moreover, the complex-valued EKF has shown better tolerance to large noise SNR observations. Both real-valued and complex-valued EKFs showed good response in multi-reflection environment.

Integer ambiguity resolution: The research in this thesis aims at providing a fast, unaided and non-motion-based IAR to allow for future autonomous formation flying in a tightly controlled time-critical mode. A precise LOS estimation after IAR not only needs to be provided timely to the subsequent relative orbit or attitude propagation but shall also avoid any a-priori information from other sensors, so that the RF-based system can be foreseen as the first-stage metrology in autonomous navigation before its incorporation with other systems for more accurate and robust navigation. The well-known LAMBDA method was used as a fundamental solution for integer ambiguities. The modification of LAMBDA in this thesis copes with involving redundancies to the observation model with more frequencies, antennas and constraints. Ambiguity resolutions and performances were discussed in Chapter 2, 3 and 5, respectively, in terms of designing an equivalent multi-frequency PRN structure to facilitate IAR, developing the LOS constrained single-epoch IAR, and evaluating IAR robustness to multipath.

- Three carrier frequencies in the S-band, S1, S2 and S3, were suggested to facilitate IAR in section 2.2. However, the BPSK-R ranging code only needs to be modulated onto the S1 frequency (spreading only central spectrum), while other frequencies are modulated by low rate communication data or stay unmodulated (regarded as separated tones away from the central spectrum). This can maximally avoid the code despreading process in the receiver while maintain the capability of extra-frequency assisted fast IAR.
- In Chapter 3, single-epoch IAR algorithms were proposed by taking advantage of a nonlinear quadratic LOS length constraint. Two proposed methods, namely, the validation method and the subset ambiguity bounding method, achieved remarkable improvements with up to 80% higher success rates than the original LAMBDA in numerical simulations. The validation method showed a slightly better performance than the subset ambiguity bounding method as they differ in utilizing all-ambiguity-set and subset-ambiguity, respectively. The IAR performance was also demonstrated by field tests using GPS satellite signals with various occurrences of signals disturbances, e.g., cycle slips and losses of lock due to multipath or signal blockage.
- A constrained ambiguity dilution of precision ($ADOP_{\delta l}$) measure was analytically derived to serve as an indicator in characterizing the expected success rate of ambiguity resolution. $ADOP_{\delta l}$ clearly shows that the ambiguity resolution capability depends on the the constraint threshold δl , noise levels on code and phase observations, the number of frequencies, the number of antennas and their relative geometry. For $ADOP_{\delta l}$ values smaller than 0.12 cycles, the approximately success rate can be expected higher than 99%.
- The antenna geometrical information (the number, length and relative orientations of antenna baselines), if properly arranged, can enhance ambiguity resolution and improve LOS accuracy. For better observability and higher LOS accuracy, antenna baselines are recommended having longer lengths and

larger angular separations. Nonetheless, for superior ambiguity resolution performance, antenna baselines with different lengths showed higher success rates than the baselines with all long lengths. Therefore, the impact of badly arranged angular separations between baselines can be dramatically compensated by properly determining the threshold δl . Field tests in section 3.4 has also demonstrated that cycle slips and losses of lock in the carrier phase measurements can sometimes be tolerable in the IAR process as long as four of antennas (three baselines) have uncorrupted observations.

- The multipath effects on the integer ambiguity resolution were extensively examined in section 5.4. The time required for the first ambiguity fix can be tremendously reduced after the multipath removed via the proposed cascaded EKFs. It has also been demonstrated that the classical ambiguity resolution method, namely the LAMBDA method, has a certain tolerance to multipath. This tolerance can be improved if the LOS constraint is taken into account.

Multiple access technology: Multiple access technologies were discussed in Chapter 6 in order to extend the RF-based navigation system to a large scale formation with four or more spacecraft. The CDMA technology is emphasized in terms of its capabilities and limitations.

- In order to fulfil dedicated network requirements of supporting time-critical relative navigation, the CDMA is modified to work in a centralized topology with the central target vehicle being rotated from one spacecraft to another within adjustable time slots, such that various navigation accuracy requirements in different missions phases can be fulfilled in various time slots and a problem called single point of failure can be avoided.
- The limitation of CDMA was extensively investigated in terms of the multiple access interference (MAI). MAI was found Doppler dependent and suffers as well from the near-far problem. Two case-study scenarios, one of a circular Low Earth Orbit mission and another for a highly elliptical orbit mission, were provided that demonstrated the severe effects MAI errors and the high probability of its occurrence within an orbit period. For the inter-satellite communication performance, it was found that the energy per bit to noise density ratio is reduced as compared to the case without the MAI, leading to a limited operational range diversity and a limited maximum number of spacecraft in a formation. Furthermore, the MAI error worsens the navigation performance, especially at the moment of Doppler crossovers or in case of signals being corrupted by the near-by interferences. MAI errors in pseudorange observations can easily exceed the meter level which is suggested to be mitigated using a smaller correlator spacing, a longer integration time or carrier smoothing.

7.3 Outlook

The challenges with respect to multipath, IAR and multiple access technologies have been addressed in this thesis. However, the performance and validation of the proposed solutions can be further improved in the future.

The constrained integer ambiguity resolution methods, proposed in chapter 3, utilize an approximate *soft* inequality constraint. Similar research such as con-

strained LAMBDA (C-LAMBDA) method is present in the literature which rigorously integrate *hard* equality constraints into the ambiguity objective function. Future work is thus needed to implement a comparison between these two types of methods in terms of achievable success rate and computational efficiency. In addition, the statistical theory of ambiguity validation (acceptance tests) using, e.g., ratio tests, has been analytically derived for the standard LAMBDA method. However, it has not yet been built for the constrained version of LAMBDA. This is important to be covered in future work.

Like the integer ambiguity resolution that would utilize the a-priori constraint to strengthen the underlying model, a-priori constraints in the Kalman filter would also improve the convergence properties of the filter. It is thus suggested in future work to include an inequality constraint on the carrier phase multipath parameter estimation process, such as $0 \leq \alpha < 1$, where α is the relative amplitude ratio between the multipath and LOS signals. This could constrain the propagation of α and help to distinguish the real origin (amplitude or phase) that causes the sinusoid-like oscillation in the multipath contaminated observations.

The research in this thesis focused primarily on the system architecture design, the LOS estimation, associated error reduction and integer ambiguity resolution. More work has to be done in the future on the dual one-way inter-satellite distance estimation in order to enable a complete relative navigation functionality. The inter-satellite distance can be estimated in the first instance by using only the pseudorange measurements at meter level accuracy in the coarse-mode. As shown by the PRISMA mission, the LOS ambiguity resolution results can greatly assist in the distance ambiguity resolution process, leading to a final centimeter level accuracy in fine-mode.

To verify the system performance and proposed solutions, this research included extensive tests using numerical simulations, field experiments, software-defined simulator and receiver based verifications as well as case studies. However, for space applications, the most convincing test is the in-orbit validation. The basic RF metrology has been embarked, tested and validated on the PRISMA mission. The improvements in this research are expected to be further verified in future space missions, especially the multipath mitigation performance due to the different ground and space environments. Experience from the PRISMA mission tells us that the multipath ground tests, even in an anechoic chamber, will never represent a full warranty of good in-orbit performance.

Appendix A

Covariance Matrices of $\mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{a}}}$, $\mathbf{Q}_{\hat{\mathbf{x}}(\mathbf{a})\hat{\mathbf{x}}(\mathbf{a})}$ and $\mathbf{Q}_{\hat{\mathbf{x}}(\mathbf{a}_p)\hat{\mathbf{x}}(\mathbf{a}_p)}$

To calculate the covariance matrices $\mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{a}}}$, $\mathbf{Q}_{\hat{\mathbf{x}}\hat{\mathbf{x}}}$ and $\mathbf{Q}_{\hat{\mathbf{x}}(\mathbf{a})\hat{\mathbf{x}}(\mathbf{a})}$, the rules of Kronecker product will be used

$$(\mathbf{M} \otimes \mathbf{N})^T = \mathbf{M}^T \otimes \mathbf{N}^T \quad (\text{A.1})$$

$$(\mathbf{M} \otimes \mathbf{N})^{-1} = \mathbf{M}^{-1} \otimes \mathbf{N}^{-1} \quad (\text{A.2})$$

$$(\mathbf{M} \otimes \mathbf{N})(\mathbf{P} \otimes \mathbf{Q}) = \mathbf{MP} \otimes \mathbf{NQ}, \quad (\text{A.3})$$

and the inverse of matrices will also be applied

$$\begin{bmatrix} \mathbf{M} & \mathbf{N} \\ \mathbf{P} & \mathbf{Q} \end{bmatrix}^{-1} = \begin{bmatrix} (\mathbf{M} - \mathbf{NQ}^{-1}\mathbf{P})^{-1} & -(\mathbf{M} - \mathbf{NQ}^{-1}\mathbf{P})^{-1}\mathbf{NQ}^{-1} \\ -\mathbf{Q}^{-1}\mathbf{P}(\mathbf{M} - \mathbf{NQ}^{-1}\mathbf{P})^{-1} & (\mathbf{Q} - \mathbf{PM}^{-1}\mathbf{N})^{-1} \end{bmatrix}. \quad (\text{A.4})$$

The covariance matrix of the float solutions for the ambiguities and the LOS vector is

$$\begin{bmatrix} \mathbf{Q}_{\hat{\mathbf{x}}\hat{\mathbf{x}}} & \mathbf{Q}_{\hat{\mathbf{x}}\hat{\mathbf{a}}} \\ \mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{x}}} & \mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{a}}} \end{bmatrix} = \begin{bmatrix} \mathbf{B}^T \mathbf{Q}_{yy}^{-1} \mathbf{B} & \mathbf{B}^T \mathbf{Q}_{yy}^{-1} \mathbf{A} \\ \mathbf{A}^T \mathbf{Q}_{yy}^{-1} \mathbf{B} & \mathbf{A}^T \mathbf{Q}_{yy}^{-1} \mathbf{A} \end{bmatrix}^{-1} \quad (\text{A.5})$$

with

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_m \end{bmatrix} \otimes \mathbf{I}_n \quad (\text{A.6})$$

$$\mathbf{B} = [1 \ 1 \ \dots \ 1]^T \otimes \mathbf{G} \quad (\text{A.7})$$

$$\mathbf{Q}_{yy}^{-1} = \begin{bmatrix} \sigma_\rho^2 & & & \\ & \sigma_\phi^2 & & \\ & & \ddots & \\ & & & \sigma_\phi^2 \end{bmatrix}^{-1} \otimes \mathbf{W} \quad (\text{A.8})$$

where \mathbf{n} and m are the number of frequencies and baselines. If an ultra-BOC signal structure is assumed, code measurements can only be demodulated from the centre carrier while phase measurements exist on both central carrier and separate tones. Provided the total number of frequencies is m , there are $m+1$ ones in the expression of \mathbf{B} . We assume the phase noise variance σ_ϕ^2 on different frequencies are the same. Matrix \mathbf{W} represents how the noise is magnified by the single differenced operator

$$\mathbf{W} = (\mathbf{D}\mathbf{D}^T)^{-1} \quad (\text{A.9})$$

$$\mathbf{D} = [\mathbf{I}_n, -\mathbf{e}_n] \quad (\text{A.10})$$

with \mathbf{e}_n representing an unit column matrix of order n .

Following the rules of Kronecker product, we can obtain

$$\mathbf{A}^T \mathbf{Q}_{\mathbf{y}\mathbf{y}}^{-1} \mathbf{A} = \frac{1}{\sigma_\phi^2} \mathbf{\Lambda}_1 \otimes \mathbf{W} \quad (\text{A.11})$$

$$\mathbf{B}^T \mathbf{Q}_{\mathbf{y}\mathbf{y}}^{-1} \mathbf{A} = \frac{1}{\sigma_\phi^2} \mathbf{\Lambda}_2 \otimes (\mathbf{G}^T \mathbf{W}) \quad (\text{A.12})$$

$$\mathbf{B}^T \mathbf{Q}_{\mathbf{y}\mathbf{y}}^{-1} \mathbf{B} = \left(\frac{1}{\sigma_\rho^2} + \frac{m}{\sigma_\phi^2} \right) (\mathbf{G}^T \mathbf{W} \mathbf{G}) \quad (\text{A.13})$$

where $\mathbf{\Lambda}_1$ and $\mathbf{\Lambda}_2$ are defined as

$$\mathbf{\Lambda}_1 = \begin{bmatrix} \lambda_1^2 & & & \\ & \lambda_2^2 & & \\ & & \ddots & \\ & & & \lambda_m^2 \end{bmatrix} \quad (\text{A.14})$$

$$\mathbf{\Lambda}_2 = [\lambda_1 \quad \lambda_2 \quad \cdots \quad \lambda_m] \quad (\text{A.15})$$

Thus, the covariance matrices for $\mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{a}}}$ and $\mathbf{Q}_{\hat{\mathbf{x}}\hat{\mathbf{x}}}$ are

$$\begin{aligned} \mathbf{Q}_{\hat{\mathbf{a}}\hat{\mathbf{a}}} &= \left(\mathbf{A}^T \mathbf{Q}_{\mathbf{y}\mathbf{y}}^{-1} \mathbf{A} - (\mathbf{A}^T \mathbf{Q}_{\mathbf{y}\mathbf{y}}^{-1} \mathbf{B})(\mathbf{B}^T \mathbf{Q}_{\mathbf{y}\mathbf{y}}^{-1} \mathbf{B})(\mathbf{B}^T \mathbf{Q}_{\mathbf{y}\mathbf{y}}^{-1} \mathbf{A}) \right)^{-1} \\ &= \left(\frac{1}{\sigma_\phi^2} \mathbf{\Lambda}_1 \otimes \mathbf{W} - \frac{\sigma_\rho^2}{\sigma_\phi^2(\sigma_\phi^2 + m\sigma_\rho^2)} (\mathbf{\Lambda}_2^T \mathbf{\Lambda}_2) \otimes (\mathbf{W} \mathbf{G} (\mathbf{G}^T \mathbf{W} \mathbf{G})^{-1} \mathbf{G}^T \mathbf{W}) \right)^{-1} \end{aligned} \quad (\text{A.16})$$

$$\begin{aligned} \mathbf{Q}_{\hat{\mathbf{x}}\hat{\mathbf{x}}} &= \left(\mathbf{B}^T \mathbf{Q}_{\mathbf{y}\mathbf{y}}^{-1} \mathbf{B} - (\mathbf{B}^T \mathbf{Q}_{\mathbf{y}\mathbf{y}}^{-1} \mathbf{A})(\mathbf{A}^T \mathbf{Q}_{\mathbf{y}\mathbf{y}}^{-1} \mathbf{A})(\mathbf{A}^T \mathbf{Q}_{\mathbf{y}\mathbf{y}}^{-1} \mathbf{B}) \right)^{-1} \\ &= \sigma_\rho^2 (\mathbf{G}^T \mathbf{W} \mathbf{G})^{-1}. \end{aligned} \quad (\text{A.17})$$

The conditional LOS covariance $\mathbf{Q}_{\hat{\mathbf{x}}(\mathbf{a})\hat{\mathbf{x}}(\mathbf{a})}$ and $\mathbf{Q}_{\hat{\mathbf{x}}(\mathbf{a})\hat{\mathbf{x}}(\mathbf{a}_p)}$ are conditioned on an all-ambiguity-set and a subset-of-ambiguity, respectively. These covariance matrices

are thus

$$\begin{aligned}
\mathbf{Q}_{\hat{\mathbf{x}}(\mathbf{a})\hat{\mathbf{x}}(\mathbf{a})} &= \left(\mathbf{B}^T \mathbf{Q}_{yy}^{-1} \mathbf{B}\right)^{-1} \\
&= \frac{1}{\frac{1}{\sigma_\rho^2} + \frac{m}{\sigma_\phi^2}} \left(\mathbf{G}^T \mathbf{W} \mathbf{G}\right)^{-1} \\
&\approx \frac{\sigma_\phi^2}{m} \left(\mathbf{G}^T \mathbf{W} \mathbf{G}\right)^{-1} \tag{A.18}
\end{aligned}$$

$$\begin{aligned}
\mathbf{Q}_{\hat{\mathbf{x}}(\mathbf{a}_p)\hat{\mathbf{x}}(\mathbf{a}_p)} &= \left(\mathbf{G}_p^T \mathbf{Q}_{\Delta\Phi_p\Delta\Phi_p}^{-1} \mathbf{G}_p\right)^{-1} \\
&= \sigma_\phi^2 \left(\mathbf{G}_p^T \mathbf{W}_p \mathbf{G}_p\right)^{-1} \tag{A.19}
\end{aligned}$$

where $\mathbf{W}_p = (\mathbf{D}_p \mathbf{D}_p^T)^{-1}$ with the single difference operator $\mathbf{D}_p = [\mathbf{I}_3, -\mathbf{e}_3]$ for the primary three observation equations.

Appendix B

The Determinant of $\mathbf{Q}_{\hat{a}\hat{a}}$

Before deviating the determinant of $\mathbf{Q}_{\hat{a}\hat{a}}$ in Eq.(A.16), the properties of the determinant for an identity matrix \mathbf{I} and for square matrices \mathbf{M} of size $m \times m$ and \mathbf{N} of size $n \times n$ shall be known

$$|\mathbf{I}| = 1 \quad (\text{B.1})$$

$$|\mathbf{M}^T| = |\mathbf{M}| \quad (\text{B.2})$$

$$|\mathbf{M}^{-1}| = \frac{1}{|\mathbf{M}|} \quad (\text{B.3})$$

$$|c\mathbf{M}| = c^m |\mathbf{M}| \quad (\text{B.4})$$

$$|\mathbf{M} \otimes \mathbf{N}| = |\mathbf{M}|^n |\mathbf{N}|^m. \quad (\text{B.5})$$

The factorization rule for the determinant of matrix $\mathbf{M} + \mathbf{UN}^{-1}\mathbf{V}$ is

$$|\mathbf{M} + \mathbf{UN}^{-1}\mathbf{V}| = |\mathbf{M}| |\mathbf{N}|^{-1} |\mathbf{N} + \mathbf{VM}^{-1}\mathbf{U}| \quad (\text{B.6})$$

where \mathbf{M} , \mathbf{N} , \mathbf{U} and \mathbf{V} are all square matrices.

To derive the determinant of $\mathbf{Q}_{\hat{a}\hat{a}}$, we assume two coefficients in $\mathbf{Q}_{\hat{a}\hat{a}}$ as $a = \frac{1}{\sigma_\phi^2}$ and $b = \frac{\sigma_\rho^2}{\sigma_\phi^2(\sigma_\phi^2 + m\sigma_\rho^2)}$ for convenience. According to the factorization rule for the determinant of Eq.(B.6), the inverse of the determinant of $\mathbf{Q}_{\hat{a}\hat{a}}$ is

$$\begin{aligned} |\mathbf{Q}_{\hat{a}\hat{a}}|^{-1} &= |a\mathbf{\Lambda}_1 \otimes \mathbf{W} - b(\mathbf{\Lambda}_2^T \mathbf{\Lambda}_2) \otimes (\mathbf{W}\mathbf{G}(\mathbf{G}^T \mathbf{W}\mathbf{G})^{-1} \mathbf{G}^T \mathbf{W})| \\ &= a^{mn} |\mathbf{\Lambda}_1 \otimes \mathbf{W}| |\mathbf{G}^T \mathbf{W}\mathbf{G}|^{-1} |\mathbf{G}^T \mathbf{W}\mathbf{G} - \\ &\quad \frac{b}{a} (\mathbf{\Lambda}_2 \otimes (\mathbf{G}^T \mathbf{W})) \cdot (\mathbf{\Lambda}_1 \otimes \mathbf{W})^{-1} \cdot (\mathbf{\Lambda}_2^T \otimes \mathbf{W}\mathbf{G})| \\ &= a^{mn} |\mathbf{\Lambda}_1 \otimes \mathbf{W}| |\mathbf{G}^T \mathbf{W}\mathbf{G}|^{-1} |\mathbf{G}^T \mathbf{W}\mathbf{G} - \\ &\quad \frac{b}{a} (\mathbf{\Lambda}_2 \mathbf{\Lambda}_1^{-1} \mathbf{\Lambda}_2^T \otimes (\mathbf{G}^T \mathbf{W}\mathbf{G}))| \\ &= a^{mn} |\mathbf{\Lambda}_1 \otimes \mathbf{W}| |\mathbf{G}^T \mathbf{W}\mathbf{G}|^{-1} \left(1 - \frac{mb}{a}\right) |\mathbf{G}^T \mathbf{W}\mathbf{G}| \\ &= a^{mn} \left(1 - \frac{mb}{a}\right)^3 |\mathbf{\Lambda}_1 \otimes \mathbf{W}| \\ &= a^{mn} \left(1 - \frac{mb}{a}\right)^3 |\mathbf{\Lambda}_1|^n |\mathbf{W}|^m \\ &= \frac{1}{\sigma_\phi^{2mn} \left(1 + \frac{m\sigma_\rho^2}{\sigma_\phi^2}\right)^3} \left(\prod_{i=1}^m \lambda_i\right)^{2n} \frac{1}{(n+1)^m} \end{aligned} \quad (\text{B.7})$$

Thus,

$$\begin{aligned}
|\mathbf{Q}_{\hat{a}\hat{a}}| &= \sigma_\phi^{2mn} \left(1 + \frac{m\sigma_\rho^2}{\sigma_\phi^2}\right)^3 \left(\prod_{i=1}^m \lambda_i\right)^{-2n} (n+1)^m \\
&\approx \sigma_\phi^{2mn} m^3 \left(\frac{\sigma_\rho^2}{\sigma_\phi^2}\right)^3 \left(\prod_{i=1}^m \lambda_i\right)^{-2n} (n+1)^m \quad (\text{B.8}) \\
&= m^3 (n+1)^m \left(\frac{\sigma_\rho^2}{\sigma_\phi^2}\right)^3 \left(\frac{\sigma_\phi^m}{\prod_{i=1}^m \lambda_i}\right)^{2n} \\
&= m^3 (n+1)^m \frac{\sigma_\rho^6 \sigma_\phi^{2mn-6}}{\left(\prod_{i=1}^m \lambda_i\right)^{2n}}.
\end{aligned}$$

Appendix C

Correlation Coefficient between Early/Late Correlators

The noise on correlators is white with respect to time, while correlators with different delays I_i and I_j along the delay dimension are highly correlated. In the following, the correlation coefficient between I_i and I_j will be derived.

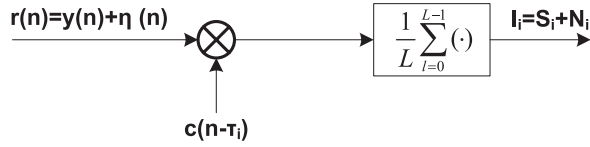


Figure C.1: Scheme of correlation

The correlation process is shown in Figure C.1. The received signal after quantization $r(n)$ consists of a real signal $y(n)$ and a noise $\eta(n)$. The correlator output is also the sum of a deterministic correlation value S_i and a noise N_i . The mean and the variance of N_i are $E(N_i) = 0$ and $\text{Var}(N_i) = \sigma_{IF}^2/L$, where σ_{IF}^2 is the signal variance after down-conversion from the carrier to the intermediate frequency band, and L is the number of samples in the code sequence (Borio, 2012).

The covariance between N_i and N_j is

$$\begin{aligned} \text{cov}(I_i, I_j) &= E[(S_i + N_i)(S_j + N_j)^*] - E[(S_i + N_j)] E^*[(S_j + N_j)] \\ &= E[S_i S_j^* + S_i N_j^* + S_j^* N_i + N_i N_j^*] - E[S_i] E^*[S_j]. \end{aligned} \quad (\text{C.1})$$

Since

$$\begin{aligned} E[S_i N_j^*] &= S_i E[N_j^*] = 0 \\ E[S_j^* N_i] &= S_j^* E[N_i] = 0 \\ E[S_i S_j^*] &= E[S_i] E^*[S_j], \end{aligned} \quad (\text{C.2})$$

the covariance $\text{cov}(I_i, I_j)$ then equals to

$$\begin{aligned}
\text{cov}(I_i, I_j) &= E [N_i N_j^*] \\
&= E \left[\frac{1}{L} \sum_{n=0}^{L-1} \eta(n) c(n - \tau_0) \cdot \frac{1}{L} \sum_{m=0}^{L-1} \eta^*(m) c(m - \tau_1) \right] \\
&= \frac{1}{L^2} \sum_{n=0}^{L-1} \sum_{m=0}^{L-1} E [\eta(n) \eta^*(m)] c(n - \tau_0) c(m - \tau_1) \\
&= \frac{1}{L^2} \sum_{n=0}^{N-1} \sum_{m=0}^{L-1} \sigma_{IF}^2 \delta(m - n) c(n - \tau_0) c(m - \tau_1) \\
&= \frac{1}{L} \sigma_{IF}^2 \cdot \frac{1}{L} \sum_{n=0}^{L-1} c(n - \tau_0) c(n - \tau_1) \\
&= \frac{\sigma_{IF}^2}{L} R(\Delta\tau)
\end{aligned} \tag{C.3}$$

where $\Delta\tau = \tau_i - \tau_j$, $\delta(m - n) = 1$ only when $m = n$. Then, the correlation coefficient between I_i and I_j is

$$\begin{aligned}
\rho_{I_i, I_j} &= \frac{\text{cov}(I_i, I_j)}{\sqrt{\sigma_{I_i}^2 \sigma_{I_j}^2}} = \frac{\frac{\sigma_{IF}^2}{L} R(\Delta\tau)}{\frac{\sigma_{IF}^2}{L} R(0)} \\
&= R(\Delta\tau) .
\end{aligned} \tag{C.4}$$

Bibliography

- Alfriend, K., Vadali, S. R., Gurfil, P., How, J., and Breger, L. (2010). *Spacecraft Formation Flying - Dynamics, Control and Navigation*. Butterworth-Heinemann, 1st edition.
- Aung, M., Purcell, G. H., Tien, J. Y., and et al (2002). Autonomous Formation Flying Sensor for the Starlight Mission. In *the International Symposium on Formation Flying Missions and Technologies*, Toulouse.
- Avila-Rodriguez, J. A., Hein, G. W., Wallner, S., Issler, J. L., Ries, L., Lestarquit, L., de Latour, A., Godet, J., Bastide, F., Pratt, A. R., and Owen, J. (2008). The MBOC Modulation: A Final Touch for the Galileo Frequency and Signal Plan. *Journal of the Institute of Navigation*, 55(1):14–28.
- Axelrad, P., Comp, C. J., and Macdoran, P. F. (1996). SNR-based Multipath Error Correction for GPS Differential Phase. *IEEE Transactions on Aerospace and Electronic Systems*, 32(2):650–660.
- Axelrad, P., Gold, K., Madhani, P., and Reichert, A. (1999). Analysis of Orbit Errors Induced by Multipath for the ICESat Observatory. In *the 12nd ION GPS*, pages 875–883, Nashville, TN.
- Barrena, V., Suatoni, M., Flores, C., Thevenet, J., and Mehlen, C. (2008). Formation flying rf ranging subsystem for prisma: Navigation algorithm design and implementation. In *the 3rd International Symposium on Formation Flying Missions and Technologies*, Noordwijk, the Netherlands.
- BarSever, Y., Srinivasan, J., and et al (2001). From AFF to CCNT: JPL's Evolving Family of Multifunction Constellation Transceivers. In *the 2nd International Workshop on Satellite Formation Flying*, Isreal.
- Bastide, F., Akos, D., Macabiau, C., and Roturier, B. (2003). Automatic Gain Control (AGC) as An Interference Assessment Tool. In *16th ION*, pages 2042–2053, Portland, OR.
- Bertiger, W., BarSever, Y., Desai, S., and et al (2002). GRACE: Millimeters and Microns in Orbit. In *the 15th ION GPS*, pages 2022–2029, Portland.
- Betz, J. W. (2001). Binary Offset Carrier Modulations for Radionavigation. *Journal of The Institute of Navigation*, 48(4):227–246.
- Betz, J. W. and Kolodziejwski, K. R. (2000). Extended Theory of Early-Late Code Tracking for a Bandlimited GPS Receiver. *Journal of The Institute of Navigation*, 47(3):211–226.

- Betz, J. W. and Kolodziejcki, K. R. (2009). Generalized Theory of Code Tracking with an Early-Late Discriminator Part I: Lower Bound and Coherent Processing. *IEEE Transactions on Aerospace and Electronic Systems*, 45(4):1538–1556.
- Bilich, A. (2006). *Improving the Precision and Accuracy of Geodetic GPS: Applications to Multipath and Seismology*. Doctoral dissertation, University of Colorado.
- Bilich, A. and Larson, K. M. (2007). Mapping the GPS Multipath Environment using the Signal-to-Noise Ratio. *Radio Science*, 42(RS6003).
- Bilich, A., Larson, K. M., and Axelrad, P. (2008). Modeling GPS Phase Multipath with SNR: Case Study from the Salar de Uyuni, Bolivia. *Journal of Geophysical Research*, 113(B04401).
- Borio, D. (2008). *A Statistical Theory for GNSS Signal Acquisition*. Doctoral dissertation, POLITECNICO DI TORINO.
- Borio, D. (2012). Gns Receiver Design. Technical Report Lecture notes ENGO 638, University of Calgary.
- Borre, K., Akos, D. M., Bertelsen, N., and et al. (2007). *A Software-Defined GPS and Galileo Receiver: A Single-Frequency Approach*. Birkhauser Boston, 1st edition.
- Bourga, D., Mehlen, C., and et al (2002). A Formation Flying RF Subsystem for DARWIN and SMART-2. In *the International Symposium Formation Flying Missions and Technologies*, Toulouse, France.
- Bristow, J., Folta, D., and Hartman, K. (2000). A Formation Flying Technology Vision. In *AIAA Space Conference and Exposition*, Long Beach, CA.
- Buist, P. J., Teunissen, P. J. G., and Giorgi, G. (2011). Multivariate Bootstrapped Relative Positioning of Spacecraft Using GPS L1/Galileo E1 Signals. *Advances in Space Research*, 47(5):770–785.
- Buist, P. J., Teunissen, P. J. G., Giorgi, G., and Verhagen, S. (2009). Multiplatform Instantaneous GNSS Ambiguity Resolution for Triple- and Quadruple-Antenna Configurations with Constraints.
- Busking, E. B., Elferink, F. H., and van, H. (2011). Method for Measuring the Distance Between Tags. EP Patent App. EP20,090,173,013.
- Byun, S. H., Hajj, G. A., and Young, L. E. (2002). Development and Application of GPS signal Multipath Simulator. *Radio Science*, 37(6):10–1–10–23.
- Cantrell, P. E. and Ojha, A. K. (1987). Comparison of Generalized Q-Function Algorithms. *IEEE Transactions on Information Theory*, 33:591–596.
- Centrella, J. and Reddy, F. (2011). Goddard’s Astrophysics Science Division Annual Report 2010. Technical Report NASA/TM-20110215870, NASA Goddard Space Flight Center.
- Chen, D. and Lachapelle, G. (1995). A Comparison of the FASF and Least-squares Search Algorithms for On-the-fly Ambiguity Resolution. *Journal of the Institute of Navigation*, 42(2):371–391.

- Clare, L. P., Gao, J. L., Jennings, E. H., and Okino, C. (2005). A Network Architecture for Precision Formation Flying Using the IEEE 802.11 MAC Protocol. In *IEEE Aerospace Conference*, Big Sky, MT.
- Cohen, C. E. (1992). *Attitude Determination Using GPS*. Doctoral dissertation, Stanford University.
- Comp, C. J. and Axelrad, P. (1998). Adaptive SNR-based Carrier Phase Multipath Mitigation Technique. *IEEE Transactions on Aerospace and Electronic Systems*, 34(1):264–276.
- Dai, Z. (2012). MATLAB Software for GPS Cycle-slip Processing. *GPS Solution*, 16:267–272.
- D’Amico, S. (2010). *Autonomous Formation Flying in Low Earth Orbit*. Doctoral dissertation, Delft University of Technology.
- Daneshmand, S. (2013). *GNSS Interference Mitigation Using Antenna Array Processing*. Doctoral dissertation, University of Calgary.
- Dash, P. K., Jena, R. K., Panda, G., and Routray, A. (2000). An Extended Complex Kalman Filter for Frequency Measurement of Distorted Signals. *IEEE Transactions on Instrumentation and Measurement*, 49(4):746–753.
- de Jonge, P. and Tiberius, C. (1996). The LAMBDA Method for Integer Ambiguity Estimation: Implementation Aspects. Technical Report LGR Series, Delft University of Technology.
- Delpech, M., Guidotti, P. Y., Grelier, T., and et al (2011). First Formation Flying Experiment based on a Radio Frequency Sensor: Lessons Learned and Perspectives for Future Missions. In *the 4th International Symposium on Formation Flying Missions and Technologies*, Quebec, Canada.
- Dierendonck, A., Fenton, P., and Ford, T. (1992). Theory and Performance of Narrow Correlator Spacing in a GPS Receiver. *Journal of the Institute of Navigation*, 39(3):265–283.
- Duncan, S. M., Unwin, M., Sweeting, M., and Hodgart, S. (2008). In-Orbit Validation of A GPS Attitude Sensor. In *4th ESA Workshop on Satellite Navigation Technologies (NAVITEC)*.
- Edwards, B. L. (2002). Distributed Spacecraft Crosslink Study Part 1: Spectrum Requirements and Allocation Survey Report and Recommendations. Technical report, Goddard Space Flight Center.
- Elferink, F. H. and Hoogeboom, P. (2013). Method for Determining Distance and Speed of FMCW Radar Terminals. WO Patent App. PCT/NL2012/050,859.
- Ercek, R., Doncker, P. D., and Grenez, F. (2006). Statistical Determination of the PR Error Due to NLOS-Multipath in Urban Canyons. In *the 19th ION GNSS*, pages 1771–1777, Forth Worth, TX.
- ESA (2007). DARWIN: Study ended, no further activities planned. http://www.esa.int/Our_Activities/Space_Science/Darwin_overview.

- Fan, K. and Ding, X. (2006). Estimation of GPS Carrier Phase Multipath Signals Based on Site Environment. *Journal of Global Positioning Systems*, 5(1-2):22–28.
- Fan, S., Zhang, K., and Wu, F. (2005). Ambiguity Resolution in GPS-based Low-Cost Attitude Determination. *Journal of Global Positioning Systems*, 4(1-2):207–214.
- Fention, P. C. and Jones, J. (2005). The theory and performance of novatel inc.’s vision correlator. In *the 18th ION GNSS*, pages 2178–2186, Long Beach, CA.
- Filippov, V., Sutiadin, I., and Ashjaee, J. (1999). Measured characteristics of dual depth dual frequency choke ring for multipath rejection in gps receivers. In *the 12nd ION GPS*, pages 793–796, Nashville.
- Fine, P. and Wilson, W. (1999). Tracking Algorithm for GPS Offset Carrier Signal. In *National Technical Meeting of ION*, pages 671–676, San Diego, CA.
- Fisman, P. M., Betz, J. W., Clark, J. E., and et al. (2000). Predicting Performance of Direct Acquisition for the M-code Signal. In *ION National Technical Meeting (NTM)*, pages 574–582, Anaheim, CA.
- Frei, E. and Beutler, G. (1990). Rapid Static Positioning Based on the Fast Ambiguity Resolution Approach (FARA): Theory and First Results. *Manuscripta Geodaetica*, 15(4):325–356.
- Garin, L., van Diggelen, F., and Rousseau, J. M. (1996). Strobe and Edge Correlator Multipath Mitigation for Code. In *the 9th ION GPS*, pages 657–664, Kansas City.
- Gill, E., Montenbruck, O., and D’Amico, S. (2007). Autonomous Formation Flying for the PRISMA Mission. *Journal of Spacecraft and Rockets*, 44(3):671–681.
- Giorgi, G. (2011). *GNSS Carrier Phase-Based Attitude Determination: Estimation and Applications*. Doctoral dissertation, Faculty of Aerospace Engineering, Delft University of Technology.
- Giorgi, G. and Teunissen, P. J. G. (2012). Instantaneous Global Navigation Satellite System (GNSS)-Based Attitude Determination for Maritime Applications. *IEEE Journal of Oceanic Engineering*, 37(3):348–362.
- Giorgi, G., Teunissen, P. J. G., Verhagen, S., and Buist, P. J. (2010). Testing A New Multivariate GNSS Carrier Phase Attitude Determination Method for Remote Sensing Platforms. *Advances in Space Research*, 46(2):118–129.
- Giorgi, G., Teunissen, P. J. G., Verhagen, S., and Buist, P. J. (2012). Instantaneous Ambiguity Resolution in GNSS-based Attitude Determination Applications: A Multivariate Constraint Approach. *Journal of Guidance, Control, and Dynamics*, 35(1):51–67.
- Grelier, T., Delpech, M., Guidotti, P. Y., and et al. (2011). Flight Results of FFIORD Radio Frequency Sensor. In *the 4th International Symposium on Formation Flying Missions and Technologies*, Quebec, Canada.
- Grelier, T., Garcia, A., Peragin, E., Lestarquit, L., Harr, J., and et al (2008). GNSS in Space Part 1: Formation Flying Radio Frequency Missions, Techniques and Technology. Technical report, Inside GNSS.

- Grelier, T., Garcia, A., Peragin, E., Lestarquit, L., Harr, J., and et al (2009). GNSS in Space Part 2: Formation Flying Radio Frequency Techniques and Technology. Technical report, Inside GNSS.
- Hatch, R. (1986). Dynamic Advanced GPS at the Centimeter Level. In *the 4th International Geodetic Symposium on Satellite Positioning*, Austin, Texas.
- Heckler, G., Winternitz, L., and Bamford, W. (2008). Mms-IRAS TRL-6 Testing. In *the 21st ION GNSS*, Savannah, GA.
- Hein, G. W., Avila-Rodriguez, J. A., Wallner, S., Pratt, A. R., Owen, J., Issler, J., Betz, J. W., Hegarty, C. J., Lenahan, S., and et al. (2006). MBOC: The New Optimized Spreading Modulation Recommended for Galileo L1 OS and GPS L1C. In *IEEE/ION Symposium on Position, Location and Navigation*, pages 883–892, San Diego, CA.
- Hein, G. W., Irsigler, M., Avila-Rodriguez, J. A., and Pany, T. (2004). Performance of Galileo L1 Signal Candidates. In *Proceedings of European Navigation Conference GNSS (ENC-GNSS04)*, Rotterdam, the Netherlands.
- Hodgart, M. S. and Blunt, P. D. (2007). Dual Estimate Receiver of Binary Offset Carrier Modulated Signals for Global Navigation Satellite Systems. *Electronics Letters*, 43(16):877–878.
- Hodgart, M. S., Blunt, P. D., and Unwin, M. (2008). Double Estimator-A New Receiver Principle for Tracking BOC Signals. *Inside GNSS*, pages 26–36.
- Hogie, K., Criscuolo, E., and Parise, R. (2005). Using Standard Internet Protocols and Applications in Space. *Journal of Computer Networks*, 47:603–650.
- Huisman, L., Teunissen, P. J. G., and Odijk, D. (2010). On the Robustness of Next Generation GNSS Phase-only Real-Time Kinematic Positioning. In *the International Federation of Surveyors Congress*, Sydney, Australia.
- Hwu, S. U. and Loh, Y. C. (1999). Space station gps multipath analysis and validation. In *the 49th IEEE Vehicular Technology Conference*, pages 757–761, Houston, TX.
- Irsigler, M. (2008). *Multipath Propagation, Mitigation and Monitoring in the Light of Galileo and the Modernized GPS*. Doctoral dissertation, Bundeswehr University Munich.
- Irsigler, M. and Eissfeller, B. (2003). Comparison of Multipath Mitigation Techniques with Consideration of Future Signal Structures. In *the 16th ION GPS/GNSS*, pages 2584–2592, Portland.
- Jansen, F., Lumb, D., Altieri, B., and et al (2001). XMM-Newton Observatory. *Astronomy and Astrophysics*, 365(1):1–6.
- Jazwinski, A. H. (1970). *Stochastic Processes and Filtering Theory*. Academic Press, 1st edition.
- Jones, J., Fenton, P., and Smith, B. (2004). Theory and Performance of the Pulse Aperture Correlator. Technical report, Alberta, Canada.

- Joosten, P. and Irsigler, M. (2003). Gns Ambiguity Resolution in the Presence of Multipath. In *the European Navigation Conference GNSS*, Graz, Austria.
- Jung, J. (1999). High Integrity Carrier Phase Navigation for Future LAAS using Multiple Civilian GPS Signals. In *the 12nd ION GPS*, pages 727–736, Nashville.
- Kaplan, E. D. and Hegarty, C. J. (2006). *Understanding GPS: Principle and Applications*. Artech House, 2nd edition.
- Keong, J. H. (1999). GPS/GLONASS Attitude Determination with a Common Clock using a Single Difference. In *the 12nd ION GPS*, pages 1941–1950, Nashville, TN.
- Kim, D. and Langley, R. B. (1999). An Optimized Least-squares Technique for Improving Ambiguity Resolution Performance and Computational Efficiency. In *the 12nd ION GPS*, pages 1579–1588, Nashville, Tennessee.
- Kim, D. and Langley, R. B. (2000). GPS Ambiguity Resolution and Validation: Methodologies, Trends and Issues. In *7th Workshop on GNSS-International Symposium on GPS/GNSS*, Seoul, Korea.
- Kim, J. and Lee, S. W. (2009). Flight Performance Analysis of GRACE K-band Ranging Instrument with Simulation Data. *Acta Astronautica*, 65:1571–1581.
- Kim, J. and Tapley, B. D. (2002). Error Analysis of A Low-Low Satellite-to-Satellite Tracking Mission. *Journal of Guidance, Control and Dynamics*, 25(6):1100–1106.
- Kim, J. and Tapley, B. D. (2003). Simulations of Dual One-Way Ranging Measurements. *Journal of Spacecraft and Rockets*, 40(3):419–425.
- Krieger, G., Moreira, A., Fiedler, H., and et al (2007). TanDEM-X: A Satellite Formation for High-resolution SAR Interferometry. *IEEE Transactions on Geoscience and Remote Sensing*, 45(11):3317–3341.
- Kubo, N. and Yasuda, A. (2003). How Multipath Error Influences on Ambiguity Resolution. In *the 16th ION GPS/GNSS*, pages 2142–2150, Portland, OR.
- Kuylen, L. V., Boon, F., and Simsky, A. (2005). Attitude Determination Methods used in the Polarrx2@multi-antenna GPS Receiver. In *18th ION GPS*, page 125–135, Long Beach, CA.
- Landgraf, M. and Mestreau-Garreau, A. (2013). Formation Flying and Mission Design for PROBA-3. *Acta Astronautica*, 82(1):137–145.
- Lau, L. (2005). *Phase Multipath Modelling and Mitigation in Multiple Frequency GPS and Galileo Positioning*. Doctoral dissertation, University of London.
- Leitner, J. (2004). Formation Flying - the Future of Remote Sensing from Space. Technical Report NASA-Techdoc-20040171390, NASA Goddard Space Flight Center.
- Lestarquit, L., Harr, J., Trommer, G. F., and et al. (2006). Autonomous Formation Flying RF Sensor Development for the PRIMSA Mission. In *19th ION GNSS*, pages 2571–2578, Fort Worth, TX.

- Llorente, J. S., Agenjo, A., Carrascosa, C., de Negueruela, C., Mestreau-Garreau, A., and et al (2013). PROBA-3: Precise Formation Flying Demonstration Mission. *Acta Astronautica*, 82(1):38–46.
- Lunot, V., Seyfert, F., Bila, B., and Nasser, A. (2008). Certified Computation of Optimal Multiband Filtering Functions. *IEEE Transactions on Microwave Theory and Techniques*, 56(1):105–112.
- MacGougan, G., O’Keefe, K., and Chiu, D. S. (2008). Multiple UWB Range Assisted GPS RTK in Hostile Environments. In *the 21st ION GNSS*, pages 3020–3035, Savannah, GA.
- Martin, N., Leblond, V., Guillotel, G., and et al. (2003). BOC(x,y) Signal Acquisition Techniques and Performance. In *the 16th ION GPS/GNSS*, pages 188–198, Portland, OR.
- Martin-Neira, M., Toledo, M., and Pelaez, A. (1995). The Null Space Method for GPS Integer Ambiguity Resolution. In *Proceedings of DSNS’95*, pages 170–178, Bergen, Norway.
- Mcgraw, G. A. and Braash, M. S. (1999). GNSS Multipath Mitigation Using Gated and High Resolution Correlator Concepts. In *ION National Technical Meeting (NTM)*, pages 333–342, San Diego.
- Mendel, J. M. (1995). *Lessons in Estimation Theory for Signal Processing, Communication and Control*. Prentice Hall PTR, 2nd edition.
- Misra, P. and Enge, P. (2001). *Global Positioning System Signals Measurements and Performance*. Ganga-Jamuna Press, 2nd edition.
- Mohiuddin, S. and Psiaki, M. L. (2008). Carrier-Phase Differential Global Positioning System Navigation Filter for High-Altitude Spacecraft. *Journal of Guidance, Control and Dynamics*, 31(4):801–814.
- Monikes, R., Wendel, J., and Trommer, G. F. (2005). A Modified LAMBDA Method for Ambiguity Resolution in the Presence of Position Domain Constraints. In *the 18th ION GNSS*.
- Montenbruck, O. and Gill, E. (2000). *Satellite Orbits: Models, Methods and Applications*. Springer, 1st edition.
- Mullen, L. (2011). Rage Against the Dying of the Light. *Astrobiology Magazine, Exploring the Solar System and Beyond*.
- Nishiyama, K. (1997). A Nonlinear Filter for Estimating a Sinusoidal Signal and its Parameters in White Noise. *IEEE Transactions on Signal Processing*, 45(4):970–981.
- NovAtel (2009). GPS-701-GG and GPS-702-GG Antennas. Technical report. <http://www.novatel.com/assets/Documents/Papers/GPS701-702GG.pdf>.
- Odiijk, D., Teunissen, P. J. G., and Amiri-Simkooei, A. R. (2008). Closed-Form ADOP Expressions for Single-Frequency GNSS-based Attitude Determination. In *Proc. VI Hotine-Marussi Symposium of Theoretical and Computational Geodesy: Challenge and Role of Modern Geodesy*, pages 200–206, Wuhan, China.

- O'Keefe, K., Petovello, M., Cao, W., Lachapelle, G., and Guyader, E. (2009). Comparing Multicarrier Ambiguity Resolution Methods for Geometry-based GPS and Galileo Relative Positioning and Their Applications to Low Earth Orbiting Satellite Attitude Determination. *International Journal of Navigation and Observation*, 2009.
- Pany, T., Forster, F., and Eissfeller, B. (2004). Real-time Processing and Multipath Mitigation of High-Bandwidth L1/L2 GPS Signals with a PC-based Software Receiver. In *the 17th ION GNSS*, pages 971–985, Long Beach, CA.
- Pany, T., Irsigler, M., and Eissfeller, B. (2005). Optimum Discriminator Based Code Multipath Mitigation. In *the 18th ION GNSS*, pages 2139–2154, Long Beach, CA.
- Park, C., Kim, I., Jee, G. I., and Lee, J. G. (1996). Efficient Ambiguity Search using Constraints Equation. In *IEEE Position, Location and Navigation Symposium*, Atlanta, US.
- Park, C. and Teunissen, P. J. G. (2003). A New Carrier Phase Ambiguity Estimation for GNSS Attitude Determination Systems. In *Proceedings of the International Symposium on GPS/GNSS*, pages 283–290, Tokyo, Japan.
- Park, C. and Teunissen, P. J. G. (2009). Integer Least-squares with Quadratic Equality Constraints and Its Application to GNSS Attitude Determination Systems. *International Journal of Control, Automation, and Systems*, 7(4):566–576.
- Park, C. W. (2001). *Precise Relative Navigation using Augmented CDGPS*. Doctoral dissertation, Stanford University.
- Pasetti, A. and Giulicchi, L. (1999). Experimental results on three multipath compensation techniques for gps-based attitude determination. Technical report, American Astronautical Society.
- Petovello, M. G. (2003). *Real-Time Integration of A Tactical-Grade IMU and GPS for High-Accuracy Positioning and Navigation*. Doctoral dissertation, University of Calgary.
- Ray, J. K. (1999). Use of Multiple Antennas to Mitigate Carrier Phase Multipath in Reference Stations. In *the 12nd ION GPS*, pages 269–280, Nashville.
- Ray, J. K. (2000). *Mitigation of GPS Code and Carrier Phase Multipath Effects using a Multi-antenna System*. Doctoral dissertation, University of Calgary.
- Reichert, A. (1999). *Correction Algorithms for GPS Carrier Phase Multipath Utilizing the Signal-to-Noise Ratio and Spatial Correlation*. Doctoral dissertation, University of Colorado.
- Reichert, A. and Axelard, P. (1999). Gps Carrier Phase Multipath Reduction Using SNR Measurements to Characterize an Effective Reflector. In *the 12nd ION GPS*, pages 1951–1959, Nashville, TN.
- Renga, A., Grassi, M., and Tancredi, U. (2013). Relative Navigation in LEO by Carrier-Phase Differential GPS with Intersatellite Ranging Augmentation. *International Journal of Aerospace Engineering*, 2013:1–11.

- Rosello, J., Silvestrin, P., Weigand, R., and et al. (2012). Next Generation of ESA's GNSS Receivers for Earth Observation Satellites. In *6th ESA Workshop on Satellite Navigation Technologies (NAVITEC)*.
- Rost, C. and Wanninger, L. (2009). Carrier Phase Multipath Mitigation based on GNSS Signal Quality Measurements. *Journal of Applied Geodesy*, 3(2009):1–8.
- Roy, R. and Kailath, T. (1989). Esprit - Estimation of Signal Parameters via Rotational Invariance Techniques. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 37(7):984–995.
- Rupp, T., D'Amico, S., Montenbruck, O., and Gill, E. (2007). Autonomous Formation Flying at DLR's German Space Operations Center (gsoc). In *the 58th International Astronautical Congress (IAC)*, Hyderabad, India.
- Scappuzzo, F. (1997). *Phase Multipath Estimation for Global Positioning System (GPS) using Signal-to-Noise Ratio (SNR) Data*. Master thesis, Massachusetts Institute of Technology.
- Scherzinger, B. M. (2000). Precise Robust Positioning with Inertial/GPS RTK. In *13rd ION GPS*, page 155–162, Salt Lake City, UT.
- Scherzinger, B. M. (2002). Robust Positioning with Single Frequency Inertially Aided RTK. In *Proceedings of ION NTM*, pages 911–917, San Diego, CA.
- Schmidt, R. (1981). *A Signal Subspace Approach to Multiple Emitter Location and Spectral Estimation*. Doctoral dissertation, Stanford University.
- Seltman, H. (2012). Approximations for Mean and Variance of a Ratio. Technical report. <http://www.stat.cmu.edu/hselman/files/ratio.pdf>.
- Sleewaegen, J. M. (1997). Multipath Mitigation, Benefits From Using the Signal-to-Noise Ratio. In *the 10th ION GPS*, pages 531–540, Kansas City, MO.
- Sleewaegen, J. M. and Boon, F. (2001). Mitigating Short-Delay Multipath: A Promising New Technique. In *the 14th ION GPS*, pages 204–213, Salt Lake City, UT.
- Slywczak, R. (2004). Low Earth Orbit Satellite Internet Protocol Communications Concept and Design. Technical Report 2004-212299, NASA/TM technical report.
- Smyrnaio, M., Schon, S., and Nicolas, M. L. (2013). *Multipath Propagation, Characterization and Modeling in GNSS, Geodetic Sciences - Observations, Modeling and Applications*. InTech, 1st edition.
- Spilker, J. (1996). *Global Positioning System: Theory and Applications, Vol. 1*. American Institute of Aeronautics and Astronautics, Inc., 1st edition.
- Sutton, E. (2002). Integer Cycle Ambiguity Resolution Under Conditions of Low Satellite Visibility. In *IEEE Symposium on Position Location and Navigation*, pages 91–98, Palm Springs, CA.
- Teunissen, P. J. G. (1995). The Least-squares Ambiguity Decorrelation Adjustment: A Method for Fast GPS Integer Ambiguity Estimation. *Journal of Geodesy*, 70(1-2):65–82.

- Teunissen, P. J. G. (1997). A Canonical Theory for Short GPS Baselines. Part I: The Baseline Precision; Part II: The Ambiguity Precision and Correlation; Part III: The Geometry of the Ambiguity Search Space; Part IV: Precision versus Reliability. *Journal of Geodesy*, 71(6):32–336, 389–401, 486–501, 513–525.
- Teunissen, P. J. G. (1998). Success Probability of Integer GPS Ambiguity Rounding and Bootstrapping. *Journal of Geodesy*, 72(10):606–612.
- Teunissen, P. J. G. (1999). An Optimality Property of the Integer Least-squares Estimator. *Journal of Geodesy*, 73(11):587–593.
- Teunissen, P. J. G. (2002). The Parameter Distributions of the Integer GPS Model. *Journal of Geodesy*, 76(1):41–48.
- Teunissen, P. J. G. (2006). The LAMBDA Method for the GNSS Compass. *Artificial Satellites*, 41(3):89–103.
- Teunissen, P. J. G. (2007). A General Multivariate Formulation of the Multi-antenna GNSS Attitude Determination Problem. *Artificial Satellites*, 42(2):97–111.
- Teunissen, P. J. G. (2010). Integer Least-squares Theory for the GNSS Compass. *Journal of Geodesy*, 84(7):433–447.
- Teunissen, P. J. G. (2011). The Affine Constrained GNSS Attitude Model and Its Multivariate Integer Least-squares Solution. *Journal of Geodesy*, 86(7):547–563.
- Teunissen, P. J. G., Joosten, P., and Tiberius, C. (2000). Bias Robustness of GPS Ambiguity Resolution. In *the 13th ION GPS*, pages 104–112, Salt Lake City, UT.
- Teunissen, P. J. G., Joosten, P., and Tiberius, C. (2002). A Comparison of TCAR, CIR and LAMBDA GNSS Ambiguity Resolution. In *the 15th ION GPS*, pages 2799–2808, Portland, OR.
- Teunissen, P. J. G. and Kleusberg, A. (1998). *GPS for Geodesy*. Springer, 2nd edition.
- Teunissen, P. J. G. and Odijk, D. (1997). Ambiguity Dilution of Precision: Definition, Properties and Application. In *the 10th ION GPS*, page 891–899, Kansas City, MO.
- Teunissen, P. J. G., Simons, D. G., and Tiberius, C. C. J. M. (2009). Probability and Observation Theory. Technical Report Lecture notes AE2E01, Delft University of Technology.
- Thevenet, J. B. and Grelier, T. (2012). Formation Flying Radio-Frequency Metrology Validation and Performance: The PRISMA Case. *Acta Astronautica*, 82:2–15.
- Tien, J. Y., Srinivasan, J. M., Young, L. E., and et al. (2004). Formation Acquisition Sensor for the Terrestrial Planet Finder (TPF) Mission. In *IEEE Aerospace Conference*, pages 2680–2690, Big Sky, Montana.
- Townsend, B., Fenton, P., Dierendonck, K. V., and van Nee, R. (1995a). L1 carrier phase Multipath error reduction using medll technology. In *the 8th ION GPS*, pages 1539–1544, Palm Springs, California.

- Townsend, B., van Nee, R., Fenton, P., and Dierendonck, K. V. (1995b). Performance Evaluation of the Multipath Estimating Delay Lock Loop. *Navigation*, 42(3):503–514.
- Townsend, B. R. and Fenton, P. C. (1994). A Practical Approach to the Reduction of Pseudorange Multipath Errors in a L1 GPS Receiver. In *the 7th ION GPS*, pages 143–148, Salt Lake City.
- Unwin, M., Jales, P., Blunt, P., and Duncan, S. (2012). Preparation for the First Flight of SSTL’s Next Generation Space GNSS Receivers. In *6th ESA Workshop on Satellite Navigation Technologies (NAVITEC)*.
- van Dierendonck, A. J., McGraw, G. A., Erlandson, R. J., and Coker, R. (1999). Cross-correlation of C/A Codes in GPS/WAAS Receivers. In *the 12nd ION GPS*, Nashville, TN.
- van Nee, R. (1995). *Multipath and Multi-Transmitter Interference in Spread-Spectrum Communication and Navigation Systems*. Doctoral dissertation, Delft University of Technology.
- van Nee, R., Sierveld, J., Fenton, P., and Townsend, B. (1994). The Multipath Estimating Delay Lock Loop: Approaching Theoretical Accuracy Limits. In *IEEE Position Location and Navigation Symposium*, pages 246–251, Las Vegas, NV.
- Verhagen, S. (2005). On the Reliability of Integer Ambiguity Resolution. *Navigation*, 52(2):98–110.
- Verhagen, S. (2012). Challenges in Ambiguity Resolution: Biases, Weak Models, and Dimensional Curse. In *the 6th ESA workshop on Satellite Navigation Technologies*, Noordwijk, The Netherlands.
- Verhagen, S. and Joosten, P. (2004). Analysis of Integer Ambiguity Resolution Algorithms. *European Journal of Navigation*, 2(4):40–52.
- Verhagen, S., Odijk, D., Boon, F., and Almansa, J. M. L. (2007). Reliable Multi-Carrier Ambiguity Resolution in the Presence of Multipath. In *the 20th ION GNSS*, pages 339–350, Fort Worth, TX.
- Verhagen, S. and Teunissen, P. J. G. (2006). New Global Navigation Satellite System Ambiguity Resolution Method Compared to Existing Approaches. *AIAA Journal of Guidance, Control, and Dynamics*, 29(4):981–991.
- Vladimirova, T., Bridges, C. P., and Prassions, G. (2007). Characterising Wireless Sensor Motes for Space Applications. In *the 2nd NASA/ESA Conference on Adaptive Hardware and System*, pages 43–50, Edinburgh, UK.
- Vladimirova, T., Wu, X., and Bridges, C. P. (2008). Development of A Satellite Sensor Network for Future Space Missions. In *IEEE Aerospace Conference*, pages 1–10, Big Sky, MT.
- Volle, M., Lee, T., and Long, A. (2007). Maneuver Recovery Analysis for the Magnetospheric Multiscale Mission. Technical report, NASA Technical Reports Server.
- Wang, B., Miao, L., Wang, S., and Shen, J. (2009). A Constrained LAMBDA Method for GPS Attitude Determination. *GPS Solution*, 13(2):97–107.

- Weill, L. R. (2002). Multipath Mitigation using Modernized GPS Signals: How Good Can it Get? In *the 15th ION GPS*, pages 493–505, Portland, OR.
- Weiss, J. P., Axelrad, P., and Anderson, S. (2007). A GNSS Code Multipath Model for Semi-Urban, Aircraft, and Ship Environments. *Journal of The Institute of Navigation*, 54(4):293–307.
- Weisskopt, M. C., Tananbaum, H. D., and et al (2000). Chandra X-ray Observatory (CXO): Overview. In *Conference of X-Ray Optics, Instruments and Mission III*, Munich.
- Wertz, J. R. and Larson, W. J. (1999). *Space Mission Analysis and Design*. Space technology library, 3rd edition.
- Wikipedia (2013). Satellite Formation Flying. http://en.wikipedia.org/wiki/Satellite_formation_flying.
- Wikipedia (2014a). Friis transmission equation. http://en.wikipedia.org/wiki/Friis_transmission_equation.
- Wikipedia (2014b). Spread Spectrum. http://en.wikipedia.org/wiki/Spread_spectrum.
- Wood, L., Ivancic, W., and Hodgson, D. (2007). Using Internet Nodes and Routers onboard Satellites. *International Journal of Satellite Communications and Networking*, 25(2):195–216.
- Xu, G. (2007). *Adjustment and Filtering Methods, GPS: Theory, Algorithm and Applications*. Springer, 2nd edition.
- Zenick, R. G. and Kohlhepp, K. (2000). GPS Micro Navigation and Communication System for Clusters of Micro and Nanosatellites. In *Small Satellite Conference*, UT.
- Zhang, W., Cannon, M. E., Julien, O., and Alves, P. (2003). Investigation of Combined GPS/Galileo Cascading Ambiguity Resolution Schemes. In *the 16th ION GPS/GNSS*, pages 2599–2610, Portland, OR.
- Zhu, Z. and van Graas, F. (2005). Effects of Cross-correlation on High Performance C/A Code Tracking. In *ION NTM*, San Diego, CA.

List of Author's Publications

Journals

1. **R. Sun**, K. O'Keefe, J. Guo, E.K.A. Gill (2014), Precise and fast GNSS signal direction of arrival estimation, *Journal of Navigation*, 67(1):17-35.
2. **R. Sun**, J. Guo, E.K.A. Gill (2013), Precise line-of-sight vector estimation based on an inter-satellite radio frequency system, *Advances in Space Research*, 51(7):1080-1095. **Outstanding Paper Award for Young Scientists**, awarded by **COSPAR** (Committee on Space Research) Bureau and announced in the COSPAR Technical Panel on Satellite Dynamics at the 40th COSPAR Scientific Assembly, Aug. 2014
3. **R. Sun**, J. Guo, E.K.A. Gill, D.C. Maessen (2012), Potentials and limitations of CDMA networks for combined inter-satellite communication and relative navigation, *International Journal on Advances in Telecommunications*, 5(1-2)

International Conferences and Workshops

1. **R. Sun**, J. Guo, E.K.A. Gill (2012), Antenna array based line-of-sight estimation using a GNSS-like inter-satellite ranging system, *6th ESA workshop on Satellite Navigation Technologies and European Workshop on GNSS Signals and Signal Processing (NAVITEC)*, Noordwijk, the Netherlands
2. **R. Sun**, J. Guo, E.K.A. Gill (2011), Signal amplitude-based multipath envelope fitting: A promising method for short-delay multipath mitigation in space applications, *the 24th ION GNSS*, Portland, Oregon
3. **R. Sun**, J. Guo, E.K.A. Gill, and D.C. Maessen (2011), Characterizing network architecture for inter-satellite communication and relative navigation in precise formation flying, *the 3rd International Conference on Advances in Satellite and Space Communications (SPACOMM)*, Budapest, Hungary, **Best Paper Award**
4. **R. Sun**, J. Guo, E.K.A. Gill (2011), Opportunities and challenges of wireless sensor networks in space, *the 61st International Astronautical Congress (IAC)*, Prague, CZ
5. **R. Sun**, D.C. Maessen, J. Guo and E.K.A. Gill (2010), Enabling inter-satellite communication and ranging for small satellites, *Small Satellite Systems and Services Symposium (4S)*, Funchal, Portugal

6. M.D. Castera, M. Beekema, E. van Breukelen, J.Guo, **R. Sun** et al. (2010), Reducing wire harness in AIT phase and intra-spacecraft communication using wireless sensor networks. *7th ESA Round Table on Micro and Nano Technologies for Space Applications*, Noordwijk, the Netherlands

Curriculum Vitae

Rui Sun was born in 1984 in Harbin, China. In 2007 and 2009, she received her B.E. degree and Master degree in Electronic Information Engineering at Harbin Institute of Technology, China. She conducted her master thesis in Micro-satellite Research Center in Harbin about the satellite communication and ranging subsystem design. From 2009 to 2014, she was a PhD candidate under the co-supervision of Prof. Eberhard Gill and Dr. Jian Guo in the group of Space System Engineering, faculty of Aerospace Engineering in Delft University of Technology, the Netherlands. Between July to November in 2012, she was invited as a visiting scientific researcher to the Positioning, Location and Navigation group in University of Calgary, Canada, worked with Prof. Kyle O'Keefe.

During the PhD period, Rui received the best paper award at the 3rd International Conference on Advances in Satellite and Space Communications (SPACOMM) in Hungary, for a paper titled "Characterizing network architecture for inter-satellite communication and relative navigation in precise formation flying". She was also the session chair for the Space Communications Session.

One of her Journal papers, entitled "Precise line-of-sight vector estimation based on an inter-satellite radio frequency system", published in the Advances in Space Research, was awarded by COSPAR Bureau as "Outstanding Paper Award for Young Scientists". The announcement of the award took place during the COSPAR Technical Panel on Satellite Dynamics at the 40th COSPAR Scientific Assembly in Moscow.