POLITECNICO DI TORINO

SCUOLA DI DOTTORATO

Research Doctorate Course in
Electronics and Communications
XXV ciclo

Final Dissertation

High-performance velocity, frequency and time estimation using GNSS

Sabrina Ugazio

Tutor
prof. Letizia Lo Presti
prof. Frank van Graas

Coordinator of the Research Doctorate Course
prof. Ivo Montrosset

2013

To my beloved parents

Alla mia mamma e al mio babbo

Nos esse quasi nanos, gigantium humeris insidentes.
We are like dwarfs on the shoulders of giants.
Bernardus Carnotensis

Acknowledgments

I am very grateful to those people, professors and colleagues, who encouraged me to do the Ph.D. course, because it is a unique opportunity, and first of all I would like to thank my supervisor, Prof. Letizia Lo Presti, for her guidance and continuous support. She is my landmark and my model.
I would like to thank Prof. Frank van Graas, the co-supervisor of this thesis, who was my tutor during my staying at Ohio University. His kindness and professionalism are unique, and I will never end to thank him for the opportunity that he gave me and for everything he taught to me.
Aknownledgements have do be done to Dr. Jade Morton (Miami University, Ohio) and Dr. Michael Sulzer (Arecibo observatory) to provide the data from the Incoherent Scatter Radar in Arecibo, and to Dr. Frank van Graas and Dr. Wouter Pelgrum that collected all the data from GNSS receivers used for the TEC measurements described in this thesis.
I deeply acknowledge the supports from my colleagues and friends from the NavSAS group in Torino from the Ohio University, which allowed me to have the most beautiful time in US. In particular, I thank Dr. Maurizio Fantino and Dr. Marco Pini, and all the lab-mate from the NavSAS group, that I do not cite not to forget someone, but they were a precious help and support, with their professionalism and their friendship.
I thank my colleagues also for their work, in particular to realize the N-Gene software receiver. I had the access to the C code of the receiver, in which I could implement the tracking loop that I developed.
During my staying in the US, I found wonderful people that gave me kind support and friendship, and among them, I would like to mention Dr. Wouter Pelgrum and my friend Samantha Craig.
Furthermore, I would like to thank Dr. van Graas, Dr. Pelgrum and Samantha Craig to have given me the possibility to participate to the work on the Rubidium

oscillator analysis. I performed analysis with the focus on the possibility to include the Rubidium measurement in the GNSS solution to improve the velocity estimate and all the work and to do this they provided me all the data and the results that they collected before my arrival.
A particular thank to my friends, that have been a precious support during all these years, starting from the University and during my Ph.D. years. In particular, I would like to mention Michele and Paolo, that were always there, for listening and helping. I am deeply thankful to all my dear friends, they always encouraged me not to give up and to believe in myself.
I have intentionally left as last ones the most important people in my life. All my thanks and all my love go to them: my family, all my family. And especially my parents, and my brother also, which have always believed in me, supported my decisions, understood, bear and encouraged me. I am afraid I will never be able to show my parents how much I love them and how much I am thankful to them. Everything was possible only thanks to you.
Intenzionalmente ho lasciato per ultime le persone piu' importanti della mia vita. I miei piu' sentiti ringraziamenti e il mio piu' profondo affetto vanno a loro: la mia famiglia, tutta la mia famiglia. E in particolare ai miei genitori, e a mio fratello, che hanno sempre creduto in me, supportato le mie decisioni, mi hanno sopportata e incoraggiata. Temo che non saro' mai in grado di mostrare ai miei genitori quanto li amo e quanto sono loro grata. Se tutto questo e' stato possibile, e' soltanto grazie a voi.
And a special thank to Adrian. Even when you were far away, you were always there.

Introduction

GNSS (Global Navigation Satellite System) receivers provide PVT solution, where PVT stands for Position, Velocity and Time. In general, the main interest of the common GNSS user is on the position solution and as a consequence the main focus of the research is the improvement of the position solution accuracy. However, many applications exist in which the measurement of both velocity and/or time is crucial and this is the reason why the focus of this thesis is on the Velocity and Time solution.

The PVT solution is computed through trilateration techniques, based on a TOA (Time Of Arrival) ranging technique, therefore the PVT solution is correlated to the measurement of time. In particular, the position solution is related to a time measurement while the velocity solution is correlated to a frequency measurement.

Different factors that affect the velocity estimate on one side and the time estimate on the other side are taken into account in this thesis, that in classic PVT solution are usually neglected.

In the velocity/frequency estimate, the significant measurement is the change in the user-satellite distance, i.e. a relative measurement, thus the measurements errors that remain constant during the time interval over which the velocity is estimated cancel out. Carrier-phase difference solution enables velocity accuracy in the order of 1 mm/s, a high-level accuracy which is crucial for many applications, including Inertial Measurement Unit (IMU) calibration, motion compensation for Synthetic Aperture Radar (SAR) and flight reference systems. Thanks to the cancellation of the common errors, that in the position solution represents the very larger error sources, in the velocity solution other minor effects become the limiting error sources. The first goal investigated in this thesis is to look for the accuracy limit that can be achieved in the velocity/frequency solution.

The second objective is to investigate the problem of high- accuracy time solution. As well as the position, the time is an absolute measurement, affected by large error sources. Furthermore, the clock error is in common to all the satellite measurements, and due to this, the common errors among the satellites are not told apart and are in general attributed to the clock measurement. As a result, lots of error sources that are not involved in the position solution become dominant in the time solution. A main limiting factor in the timing accuracy is represented by the errors in ionospheric delay estimate, where many error sources are involved, in particular the unknown bias ([1]) due to the receiver hardware.
After a part to introduce GNSS and its basic principles, with the focus on the aspects that are more relevant for the dissertation and that allow one to outline the motivations of the work, the thesis is divided in three main parts, two regarding in particular the velocity/frequency solution and the last one focused on the high-accuracy time solution.
The first step to improve the velocity solution was to notice how the performance is much worse on the vertical solution than on the horizontal and how highly correlated the vertical solution is to the local frequency estimate. This is due to the geometry of GNSS, that implies that users on the Earth or close to the Earth (as aircrafts) can see satellites all around them on the horizontal direction, but they cannot see satellites under them, which is reflected in a poorer geometry on the vertical direction. Due to this characteristic, an error on the pseudorange, as the clock error is, reflects on the vertical solution more heavily than on the other dimensions. As a result, the vertical solution can be about three times worse than the horizontal and from the covariance matrix of the solution it can be seen how the correlation is high in particular between the vertical and the clock solution. This fact is true both for the position and for the velocity solution, which means that the vertical velocity accuracy is highly correlated to the local oscillator frequency. As a result, a way to improve the vertical velocity accuracy is to obtain a better estimate of the local frequency. In this thesis, models for the local oscillators and ways to integrate the frequency estimate in the GNSS solution are investigated.
Another important aspect to improve the performance of the velocity measurement is to improve the accuracy of the GNSS measurement. Since the measurement used to obtain precise velocity is the carrier phase, which enables accuracy in the order of 1 mm/s, the goal to improve the accuracy on the carrier-phase measurement is crucial. With this objective, novel Digital Phase Lock Loops

(DPLLs) has been designed, both of second and third order, with an adaptive bandwidth algorithm. The objective was to tune the loop bandwidth according to the input signal dynamics and noise, and use a bandwidth small enough to reduce the noise effects as much as possible, but wide enough to properly track the input dynamics. Since the PLL is designed for precise velocity measurement, the performance in terms of dynamics tracking ability is crucial.

The last part of the analysis concerns the time solution. In most of cases in GNSS, high importance is given to the position accuracy, while the residual common biases are included in the receiver clock error. This approach makes the time solution not very accurate. Since the main bias which affects the time solution is the ionosphere delay, in this thesis the accuracy of the Total Electron Content (TEC) estimate is investigated, with the focus on the measurement bias. All the measurements which this thesis refers to are made using GPS (Global Positioning System) only, nevertheless sometimes in the thesis it is talked about GNSS in general. This is because the approaches considered in this thesis are tested here using GPS, but they can be applied to all the GNSSs.

Thesis outline

The thesis is organized in four parts with eleven chapters as follows:

Part I - Fundamentals on GNSS and Motivations of the Work : this part gives a general overview on how a GNSS works and outlines the motivations of the work, explaining how the proposed methods are important in order to improve GNSS accuracy.

- Chapter 1 - GNSS overview: provides a brief introduction on GNSS, including a historical overview, basic working principles and main aspects of the system's structure.
- Chapter 2 - Error Budget Overview and Dilution of Precision: includes an overview on the error budget in GNSS and on the effects of pseudorange errors on the PVT solution.

Part II - Frequency Aided Precise Velocity : in this part the precise velocity algorithm [2] is summarized and novel techniques to aid the GNSS velocity solution by means of the receiver frequency estimate are presented. An entire chapter is dedicated to the analysis of the local oscillator (in particular a Rubidium), because the modelling of the oscillator is a crucial step to improve the frequency estimate and then , thanks to it, the velocity estimate.

- Chapter 3 - Precise Velocity Algorithm: makes a summary of how the velocity solution can be computed precisely using carrier-phase single-differences and then correcting for the changes in the geometry and in the satellite Doppler, over the time interval considered to measure the velocity.
- Chapter 4 - Rubidium Oscillator Analysis: since the estimate of the frequency error affecting the local oscillator is fundamental in the computation of the velocity solution, the oscillator characteristics are very important. A Rubidium oscillator is considered and the main features to characterize an oscillator are recalled. This chapter shows tests on the Rb oscillator and investigates different models that can be applied to the Rb cell;

- Chapter 5 - Frequency Aided Precise Velocity: this chapter presents methods to integrate higher accuracy frequency estimate to the GNSS velocity solution.

Part III - Adaptive Phase Lock Loop : in this part the work on adaptive PLLs is described. The PLL is important because it measures the carrier-phase, and the precise velocity algorithm used in Part II is based on the carrier-phase single differences.

- Chapter 6 -Digital Phase Lock Loops: describes the structure of Phase Lock Loops, in particular the second and third order loops, in the digital domain.
- Chapter 7 - Adaptive Bandwidth Algorithm: presents the adaptive bandwidth algorithm that has been developed.
- Chapter 8 - Adaptive Loop Analysis: contains the simulation results obtained with the adaptive bandwidth digital PLL.

Part IV - Total Electron Content Measurements : while Part II and Part III are related to the velocity solution, this part is related to the third term of the PVT solution, i.e. the time. While the velocity is a relative measurement, the time is an absolute measurement, therefore the errors involved in the time estimate are of different nature. Since the ionospheric TEC error is a main component of the time error, an analysis of the TEC measurement is done and it is described in this last part of the thesis.

- Chapter 9 - Earth ionosphere: introduces the ionosphere and its effects on the propagation of electromagnetic waves.
- Chapter 10 - GNSS TEC Measurement and Related Error Sources: focus on the double frequency technique to estimate the ionospheric TEC using GNSS and analyzes the main error sources affecting this method.
- Chapter 11 - Comparisons of Different Methods: different method to estimate the ionospheric TEC are compared, with the main attention paid first of all on the noise and multipath effects and then on the residual error biases.

Each part contains its conclusions, then general conclusions and future activities are drawn at the last part of the thesis.

Thesis contributions

The main contributions of the thesis can be summarized as follows:

- Investigate some possible solution to improve the velocity estimate with the local frequency model.
- Provide a technique for the software receiver to minimize the carrier phase noise, which affects the performance of the velocity solution, by means of a digital PLL adapted in real-time according to the input signal characteristics.
- Investigate the performance of absolute measurements of the ionospheric TEC, which represent a main limiting factor for high accuracy time GNSS solution.

Table of contents

List of tables

List of figures

List of acronyms

AFB	Air Force Base
AWGN	Additive White Gaussian Noise
BOC	Binary Offset Carrier
C/N_0	Carrier-to-Noise-Density Ratio
CDMA	Code Division Multiple Access
CMC	Code Minus Carrier
CNMP	Code Noise and Multi-Path
COSMIC	Constellation Observing System for Meteorology Ionosphere and Climate
CS	Control Segment
DLL	Delay Lock Loop
DME	Distance Measuring Equipment
DPLL	Digital PLL
DoD	Department of Defense
ECEF	Earth Centered Earth Fixed
ECI	Earth Centered Inertial
EGNOS	European Geostationary Navigation Overlay Service
EIL	Effective Ionisation Level
ENB	Equivalent Noise Bandwidth
ENU	East North Up
ESA	European Space Agency
EU	European Union
FAA	Federal Aviation Administration

FIR	Finite Impulse Response
FLL	Frequency Lock Loop
FOC	Full Operational Capability
GA	Ground Antennas
GAGAN	GPS Aided Geo Augmented Navigation
GEO	GEostationary Orbit
GDOP	Geometric Dilution of Precision
GLONASS	GLobal Orbiting Navigation Satellite System
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
GUI	Graphical User Interface
HW	HardWare
I	In-Phase Component
IEEE	International Electrical and Electronic Engineers
IF	Intermediate Frequency
IIR	Infinite Impulse Response
ILS	Instrument Landing System
INS	Inertial Navigation System
IOC	Initial Operational Capability
ION	Institute Of Navigation
IPP	Ionospheric Pierce Point
ISMB	Istituto Superiore Mario Boella
JPL	Jet Propulsion Laboratory
LAAS	Local Area Augmentation System
LEO	Low Earth Orbit
LMS	Least Mean Square
LoS	Line of Sight
LS	Least Square
MCS	Master Control Station
MEO	Medium Earth Orbit
MLS	Microwave Landing System
MS	Monitor Stations
MSAS	Multi-functional Satellite Augmentation System
MW	MicroWave
NED	North East Down
NASA	National Aeronautics and Space Administration

NAVSAT	navy NAVigation SATellite system
NAVSTAR	NAVigation System with Timing And Ranging
NDB	Non Directional Beacon
PLL	Phase Lock Loop
PPS	Precision Positioning System
PR	Pseudo-Range
PRN	Pseudo-Random Noise
PVT	Position, Velocity and Time
Q	Quadrature Component
RCS	Radar Cross Section
RF	Radio Frequency
SA	Selective Availability
SBAS	Satellite-based Augmentation System
SD	Single Difference
SPS	Standard Positioning System
SNR	Signal-to-Noise Ratio
SR	Scatter Radar
SV	Space Vehicle
TEC	Total Electron Content
TOA	Time Of Arrival
UERE	User Equivalent Range Error
UHF	Ultra High Frequency
US	United States
USD	United States Dollar
UTC	Coordinated Universal Time
VOR	VHF Omnidirectional Range
WAAS	Wide Area Augmentation System
WMM	World Magnetic Model

Part I

Fundamentals on GNSS and Motivations of the Work

Chapter 1

GNSS Overview

1.1 Background and history

1.1.1 GPS: the first GNSS

Navigation is defined as the science of getting a craft or a person from one place to another [3]. Just after the Sputnik I launch by the Soviet Union in 1957, in the 1960s the first satellite navigation systems, the US Navy NAVigation SATellite System (NAVSAT, better known as TRANSIT) was developed [14]. The experimental satellites were launched in 1961-1962 and the final system became operational in 1964, using a constellation of only five satellites. The TRANSIT project started in particular because in that period the US Navy had a vital need for a navigation system to guide a new class of submarine carrying Polaris missiles, then it was also used by the Navy's ships. Other militar applications included hydrographic and geodetic surveying. TRANSIT provided continuous navigation satellite service from 1964, initially only for Polaris submarines and later also for civilians.

TRANSIT could only provide high-accuracy positioning in two dimensions and this is the reason why the Global Positioning System (GPS) was developed later. The US DoD (Departmente of Defense) approved the basic GPS architecture in 1973 and the first satellite was launched in 1978 ([15], [16]).

Today GPS uses 24 satellites orbiting around the earth, constellation completed in July 1993. On December 8 1993 the IOC (Initial Operational Capability) was announced officially, while FOC (Full Operational Capability) was declared in 1995, on July 17. The cost of development of GPS has been reported to be about

10 billion USD, while the annual operation and maintenance cost is estimated to be about 500 million USD.
The main objective of the DoD in developing GPS was to offer the US military accurate estimates of position, velocity and time which, in quantitative terms, meant to provide estimates with position error of 10 m, velocity error of 0.1 m/s and time error of 100 ns, all in a Root-Mean-Square (RMS) sense. Although the system was developed essentially for military scope, in 1991 US opened the service to the civilian users of the world [17]. Two kinds of service were offered. The civil one was called SPS (Standard Positioning System), to differentiate it from the militar one, denominated PPS (Precision Positioning System).
In fact the open-service signal was intentionally corrupted with the aim to allow to civil users precision of only about 100-150 metres, while military could reach precision of 10 metres. The DoD decreed that the civil users of GPS would be provided with a *reasonable accuracy* consistent with the national security considerations. The initial view was to limit the accuracy of vertical position estimates to about 500 m, then greater accuracy was allowed with horizontal position error generally below 100 m. Basically, the constraint on the vertical accuracy was decided in order to forbid non-US-military users to drive ballistic devices such as missiles using GPS.
Until 2000, the precision of the signal available to non-US-military users was limited due to a timing pulse distortion process known as Selective Availability (SA), that was an intentional noise introduced in order to add errors to the satellite signals [18]. In 2000, thanks to a US President Bill Clinton decree, that signal degradation was disabled, permitting to civilian GPS users precision of less than 10 metres.
Furthermore, in 1998, a GPS modernization program started, the so-called GPS III. The project involved not only new ground stations and new satellites, but also additional navigation signals for both civilian and military users to enhance accuracy availability for all users.

1.1.2 Other GNSS

US GPS is the first navigation satellite system to have been developed, but nowadays it is not the only one.
In the 1970s, the USSR started its GNSS, GLObal NAvigation Satellite System (GLONASS) [19]. GLONASS was intended for use by the Soviet military, for

navigation and ballistic, but on May 18, 2007, the Russian president signed a decree officially providing open access (free and without limitations) to the civilian navigation signals, both to Russian and foreign consumers. Since 2011, the system is fully operative with 24 satellites, providing a standard precision service that offers horizontal positioning accuracy within 5-10 meters. The GLONASS program involves now also India as a partner.
With the intention to fulfill GLONASS and GPS inter-operability, since 2006 Russia and the US have organized several meetings.
Galileo is the GNSS currently being built by the European Union (EU) and European Space Agency (ESA). The project is alternative and complementary to the US GPS and the Russian GLONASS. In 1999, the different concepts (from Germany, France, Italy and the United Kingdom) for Galileo were compared and reduced to one by a joint team of engineers from all four countries. Moreover, Galileo is an international project which involves also Countries out of EU, including China, Israel, Ukraine, Morocco, South Korea and Norway. In March 2002, the EU and the European Space Agency (ESA) agreed to fund the Galileo project which is intended to be the first civilian GNSS and the first stage of the Galileo programme was agreed upon officially in 2003 by the EU and the ESA.
With the scope to obtain inter-operability with GPS, in 2004 EU decided to use as signal modulation the Binary Offset Carrier (BOC), so that coexistence and the combined use of both systems will be possible.
Due to financial difficulties and arguments on the system structure, the Galileo project has been delayed several times. At the moment, two operative satellites are in orbit, and the forecast is to have a complete operativity with 30 satellites by 2018.
Unlike GPS, which the US military runs and uses on a primary basis, Galileo is as said intended primarily for civilian use. It has to be noted that the US reserves the right to limit the signal strength or accuracy of the GPS systems, or to shut down public GPS access completely (although it has never been done), so that only the US military and its allies would be able to use it in time of conflict. The European system will only be subject to shutdown for military purposes in extreme circumstances, and it will be available at its full precision to both civil and military users, providing also higher accuracy with respect to the existing systems.
The Galileo open service will offer a horizontal accuracy of 4 meters and a commercial service is in program too, which will provide better performance.

China is building its own GNSS, Compass (called also Beidou-2), which is the development of Beidou-1, a geostationary system composed by 4 satellites, providing only limited coverage and applications. Compass should be fully operational, with 35 satellites, by 2015, providing an open service with horizontal accuracy of 10 meters.

1.1.3 Augmentation systems

In addition to the cited satellite systems, augmentation systems have been built. These can be divided in two categories:

- local area augmentation systems with the limited coverage
- wide area augmentation systems, which is a SBAS (Satellite-Based Augmentation System).

The Local Area Augmentation System (LAAS) allows to correct some GPS errors by means of real time differential corrections and in particular it is used to land aircrafts, thus it is located in proximity of the airports.
The wide area augmentation system makes use of geostationary satellites. Different SBAS are operative worldwide: WAAS (Wide Area Augmentation Systems) in US (developed by the FAA) , EGNOS (European Geostationary Navigation Overlay Service) in Europe, GAGAN (GPS Aided Geo Augmented Navigation) in India and MSAS (Multi-functional Satellite Augmentation System) in Japan. WAAS an LAAS may substitute, in aircraft navigation, the use of ILS (Instrument Landing System), MLS (Microwave Landing System), DME (Distance Measuring Equipment), VOR (VHF Omnidirectional Range) and NDB (Non Directional Beacon), which requires different (and therefore more expensive) equipments on-board.

1.2 Basic principles of satellite navigation

1.2.1 Basic idea

After the launch of Sputnik I, it was discovered that the Doppler shift pattern in the transmitted signal, measured from a single ground station in a know position, was enough to determine the satellite's orbit. This discovery made the

scientists think that the concept could be applied also in the opposite, in order to determine a position on the earth measuring the transmitted signal Doppler shift, if the transmitter orbit was known.
However, it was not immediately clear if and how the idea could be realizable. Practice problems existed, such as the choice of the kind of signal to be transmitted. In fact, the transmitted signal needed to be transmittable in every location, without interference to other systems, and continuous transmission was needed at a sufficiently high power to be received from the receivers on the earth.
The solution was found in the transmission of a spread spectrum signal.
A PRN (PseudoRandom Code) code is transmitted, with power below the noise threshold, that thanks to its pseudo-random nature can be recovered from the noise (see [18]) using a correlation technique, as briefly described in Section 1.5.1.

1.2.2 Time Of Arrival ranging technique

Satellite navigation is based on the trilateration technique, consisting in solving a system containing as unknowns the user coordinates and as known parameters the satellite positions (provided by the ephemeris, contained in the navigation data) and the measured distances between the user and each of those satellites. Since the position in 3 dimensions is needed, at least 3 satellites are required, in order to have three independent equations. As shown in Figure 1.1, to find its position the receiver needs to measure its distance from three known points (3 satellites), representing the center of three spheres having as ray the distances from the user, respectively.
This technique is known as trilateration, a process that allows to determine unknown locations by measuring distances and then applying the geometry of circles, spheres or triangles. In satellite navigation, in order to find the solution another issue needs to be solved. This is due to the fact that the user-satellites distances are estimated by means of TOA (Time Of Arrival) ranging, which means to measure the propagation time, defined as the time that a signal takes to reach a receiver from a transmitter (at a known location). This time has then to be multiplied by the propagation velocity, which for GNSS signals is approximated with the speed of light ($c = 299{,}792{,}458 m/s$), to obtain the transmitter-user distance.
The issue of this technique is that it involves the knowledge of the times of transmission and of receipt respectively, information that could be perfectly known

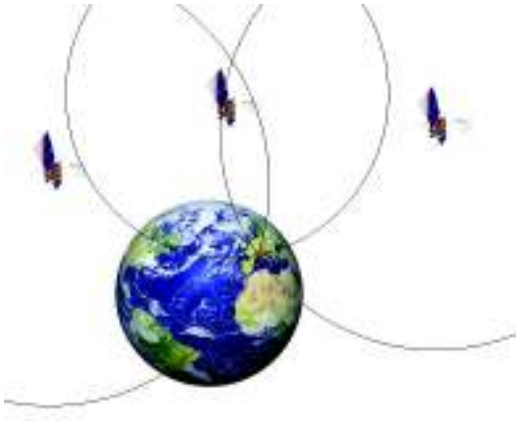

Figure 1.1: Trilateration technique. In order to compute the unknown position of an object in the 3D space, it is necessary to know its distance from at least 3 known points.

only if both the transmitter and the receiver time reference were perfectly synchronized.
Satellite frequency and time are based on highly accurate free running Cesium (Cs) or Rubidium (Rb) atomic clocks, furthermore, satellite clock corrections are computed by the ground stations and included in the navigation data.
The receiver clock instead is in general of lower quality (commercial receiver include a simple, cheap, crystal oscillator), therefore the clock error cannot be neglected. Note that at the speed of light an error on the arrival time measurement of 1 μs would cause a ranging error of 300 meters.
To handle this problem, the clock bias is treated as an unknown to solve for, therefore, since the unknowns are four, including the three spacial coordinates and the clock bias, at least four satellites in view are needed to find a solution. Figure 1.2 shows the trilateration method including the time unknown.

1.2.3 Range and pseudorange

Because of the TOA technique used, the range measured by the receiver is referred to as *pseudorange*, which is the range computed by multiplying the signal propagation velocity (speed of light) by the time difference between two non-synchronized clocks, the satellite clock and the receiver clock ([3]). Some definitions are important in GNSS not to be confused:

- true range: true satellite-user distance;

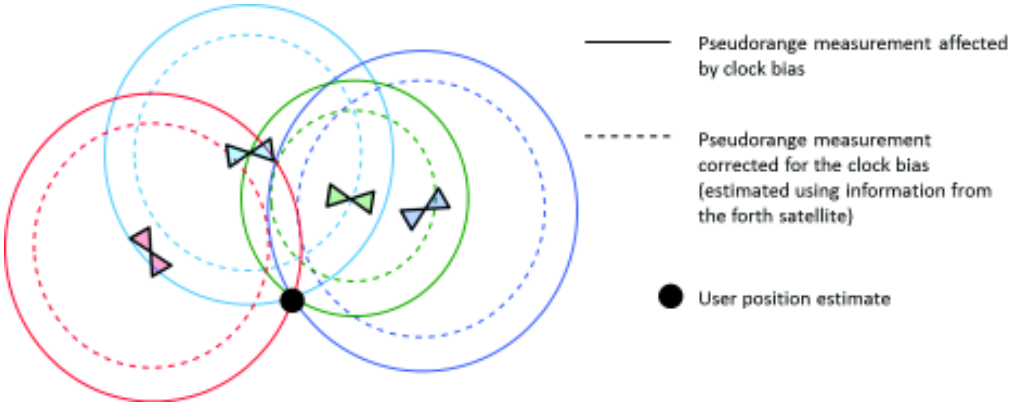

Figure 1.2: GNSS trilateration technique: four satellites in view are needed at least, to solve for the receiver clock bias.

- true pseudorange: true range affected by receiver clock unknown bias;
- pseudorange or measured pseudorange: true pseudorange affected by all the error sources.

In Figure 1.3, the basic relationship between time and range measurement are shown ([3]), where:

- T_s = transmission time (in system time)
- T_u = receiving time (in system time)
- δt = satellite clock bias (with respect to system time)
- t_u = receiver clock bias (with respect to system time)
- $T_s + \delta t$ = satellite clock reading at the time when signal leaves the satellite
- $T_u + t_u$ = receiver clock reading at the time when signal arrives to the user

Note the notation: t_u will be indicated in this thesis as ΔB. This notation that makes use of a Δ stresses the fact that the user time bias is a difference between the user time and the true time reference. The unknown is represented by this time difference (user clock error). The true transmission time will be denoted as $\tau = \Delta t$.
The true range is the true propagation time τ times the propagation velocity, assumed to be the velocity of light in the vacuum, c:

$$R_j = c \cdot \tau \tag{1.1}$$

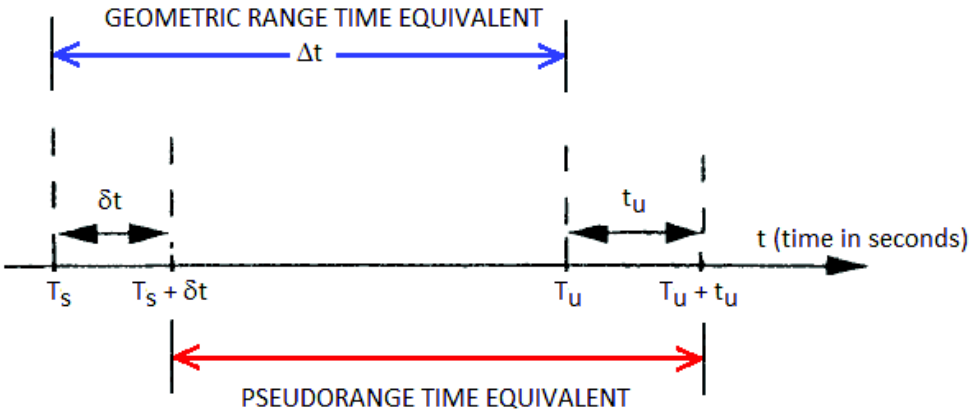

Figure 1.3: Time and range measurement. Source: [3]

The true pseudorange is defined as the true range plus an error due to the receiver and to the satellite clock error. Neglecting other error and noise sources, the true pseudorange is:

$$\rho_j = c \cdot (\tau - \Delta B + \delta t) \tag{1.2}$$

where δt is known from the navigation data, $\tau + \Delta B$ is the flying time measured by the receiver, that includes the clock error ΔB, common to all the satellite, that can be solved for so that τ can be estimated.

1.2.4 User position computation using trilateration technique

As explained in Subsection 1.2.2, in order to determine the user position (x_u,y_u,z_u) and the pseudorange measurement offset due to the unknown clock bias, at least four pseudorange measurements are needed, referred to four satellites in view. $\mathbf{s} = (\mathbf{x_j},\mathbf{y_j},\mathbf{z_j})$ represents the location of the $j-th$ satellite in view, in ECEF (Earth Centered Earth Fixed) coordinates, as well as the user location is represented by $\mathbf{u} = (\mathbf{x_u},\mathbf{y_u},\mathbf{z_u})$. The true range from a user to a satellite j is defined as:

$$\mathbf{R_j} = \mathbf{s_j} - \mathbf{u} \tag{1.3}$$

and the range magnitude is defined as the norm:

$$R_j = ||\mathbf{R_j}|| = ||\mathbf{s_j} - \mathbf{u}|| \tag{1.4}$$

representing the geometric distance from the satellite to the receiver. In (1.4), the unknown is the vector $\mathbf{u}$ representing the user position, while $\mathbf{s_j}$ is the position of the $j-th$ satellite, information contained in the navigation message from the satellite. R_j is the true range as in 1.1, but it can only be measured as pseudorange (1.2), since the receiver clock error is included in the TOA measurement. The basic system of equations that allows to get a solution for the user position and for the receiver clock bias ΔB can be expressed as:

$$\begin{cases} \rho_1 &= \sqrt{(x_{s_1}-x_u)^2+(y_{s_1}-y_u)^2+(z_{s_1}-z_u)^2}+c\Delta B \\ \rho_2 &= \sqrt{(x_{s_2}-x_u)^2+(y_{s_2}-y_u)^2+(z_{s_2}-z_u)^2}+c\Delta B \\ \rho_3 &= \sqrt{(x_{s_3}-x_u)^2+(y_{s_3}-y_u)^2+(z_{s_3}-z_u)^2}+c\Delta B \\ \rho_4 &= \sqrt{(x_{s_4}-x_u)^2+(y_{s_4}-y_u)^2+(z_{s_4}-z_u)^2}+c\Delta B \end{cases} \tag{1.5}$$

This system of equations can be solved analytically, but, due to the computational difficulties involved ([20]), an iterative method using linearization is applied.

The idea is to start from a rough estimate for both the user coordinates and the receiver clock bias and then refine the estimate using a Newton-Raphson method. A generic pseudorange can be approximated as Taylor expansion around a point, representing a known location $\hat{\mathbf{u}} = (\hat{x}_u,\hat{y}_u,\hat{z}_u)$ and a known receiver clock bias $\Delta\hat{B}$

$$\rho_j \simeq \sqrt{(x_{s_j}-\hat{x}_u)^2+(y_{s_j}-\hat{y}_u)^2+(z_{s_j}-\hat{z}_u)^2}+c\cdot\Delta\hat{B} \tag{1.6}$$

Using this approximation for the pseudoranges, (1.5) becomes

$$\begin{cases} \Delta\rho_1 &= e_{x_1}\Delta x_u + e_{y_1}\Delta y_u + e_{z_1}\Delta z_u - c\Delta B \\ \Delta\rho_2 &= e_{x_2}\Delta x_u + e_{y_2}\Delta y_u + e_{z_2}\Delta z_u - c\Delta B \\ \Delta\rho_3 &= e_{x_3}\Delta x_u + e_{y_3}\Delta y_u + e_{z_3}\Delta z_u - c\Delta B \\ \Delta\rho_4 &= e_{x_4}\Delta x_u + e_{y_4}\Delta y_u + e_{z_4}\Delta z_u - c\Delta B \end{cases} \tag{1.7}$$

where $\Delta\rho_j$ and $\Delta\mathbf{u}$ are the difference in pseudorange and in user position between the actual location and the linearization point:

$$\Delta\rho_j = \rho_j - \hat{\rho_j} \tag{1.8}$$

$$\Delta\mathbf{u} = \mathbf{u} - \hat{\mathbf{u}} \tag{1.9}$$

and the vector $\mathbf{e_j}$ represents the unary vector of the Line of Sight (LOS) between the user and the $j-th$ satellite:

$$\mathbf{e_j} = \frac{\mathbf{s}-\hat{\mathbf{u}}}{||\mathbf{s}-\hat{\mathbf{u}}||} \tag{1.10}$$

Using an array notation, it is:

$$\Delta\rho = \begin{bmatrix} \Delta\rho_1 \\ \Delta\rho_2 \\ \Delta\rho_3 \\ \Delta\rho_4 \end{bmatrix} \tag{1.11}$$

$$\mathbf{H} = \begin{bmatrix} e_{x_1} & e_{y_1} & e_{z_1} & 1 \\ e_{x_2} & e_{y_2} & e_{z_2} & 1 \\ e_{x_3} & e_{y_3} & e_{z_3} & 1 \\ e_{x_4} & e_{y_4} & e_{z_4} & 1 \end{bmatrix} \tag{1.12}$$

and

$$\mathbf{\Delta x} = \begin{bmatrix} \Delta x_u \\ \Delta y_u \\ \Delta z_u \\ -c\Delta B \end{bmatrix} \tag{1.13}$$

then

$$\Delta\rho = \mathbf{H\Delta x} \tag{1.14}$$

If only four satellites are used to compute the solution, the system is determined and the solution is:

$$\mathbf{\Delta x} = \mathbf{H^{-1}}\mathbf{\Delta}\rho \tag{1.15}$$

If more than 4 satellites are in view, it is useful to use all the available satellites (see Chapter 2), but in this case an over-determined system is considered and a simple matrix inversion cannot be used to find a solution.
In that case a Least Mean Square (LMS) approach can be applied, that gives the solution as:

$$\mathbf{\Delta x} = (\mathbf{H^T H})^{-1}\mathbf{H^T}\mathbf{\Delta}\rho \tag{1.16}$$

1.3 Coordinate systems

A note needs to be made about the coordinate frame used. So far, the ECEF coordinates have been considered, so that the user coordinates were expressed in x, y, z components. However, since this coordinate system is not easy to be

used, usually the geodetic coordinates are considered, i.e. latitude, longitude and height ([21]). Another frame very useful is the ENU, East-North-Up, or NED, North-East-Down, where a tangent plane is considered to a certain point of the earth and the coordinates are referred to that plane. The north-east directions lay on the plane, while the up (or down) direction is along the perpendicular. In this thesis, the ENU reference system will be considered also, in particular when the interest will be on the improvement of the vertical estimate (Part II). To pass from a coordinate frame to another one see as e.g. the method in [11].

1.4 GPS signal structure

Each GPS satellite transmits continuously using two radio frequencies in the L-band, which is a sub-band of the UHF (Ultra High Frequency) band (1 - 2 GHz). The frequencies used by GPS are

$$f_{L_1} = 1575.42 \text{ MHz} \tag{1.17}$$

for the so called Link 1 (L1) and

$$f_{L_2} = 1227.60 \text{ MHz} \tag{1.18}$$

for Link 2 (L2).
On L1 two signals are transmitted, one for civil users and the other for military users (DoD-authorized). The signal on L2 used to be is for DoD-authorized users only, while recently also a civilian L2 signal is available.
Also other frequencies are transmitted in GPS under the modernization program (L3, L4, L5), but in this thesis only the standard links on L1 and L2 are used.
GPS signals are made of three components:

- Carrier: a RF sinusoidal signal (frequency f_{L_1} or f_{L_2})
- PRN code (ranging code): a binary sequence (0 and 1), that repeats for a certain length. A different PRN sequence is assigned to each satellite, so that it can be univocally identified by its transmitted code. Each GPS satellite transmits two different codes:
  - Coarse/Acquisition (C/A) code, composed by a sequence of 1023 bits called chips.The duration of each C/A code is about 1μ s (chip width is about 300 m). C/A code chip rate is called *chipping rate*

  - Precision (encrypted) [P(Y)] code

- Navigation data: binary code containing different information:
  - ephemeris (satellite position and velocity)
  - satellite status (health)
  - satellite clock bias corrections
  - almanac (reduced-precision ephemeris data on all satellites in the constellation)

  The navigation message has a length of 12.5 minutes, but the essential information about satellite ephemeris and clock parameters are repeated every thirty seconds.

Figure 1.4 shows the general architecture for the generation of the L1 and L2 GPS links.

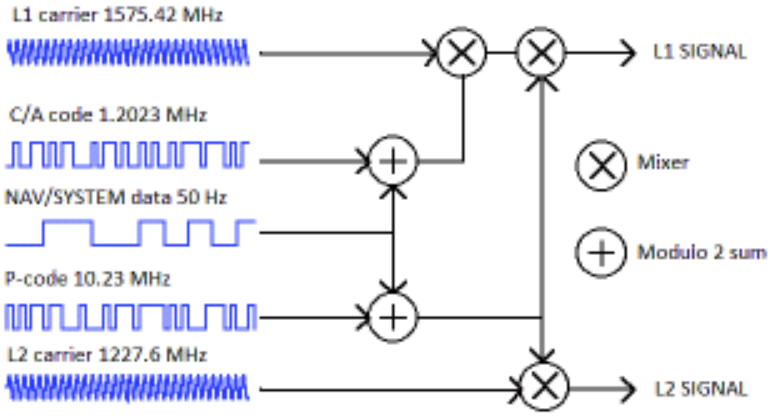

Figure 1.4: Time and range measurement.

1.5 Code and carrier range measurements

The fundamental navigation technique for GPS is to use one-way ranging from the GPS satellites that are also broadcasting their estimated positions. Ranges are measured to four satellites simultaneously in view by matching (correlating) the incoming signal with a user-generated replica signal and measuring the received phase against the user's (relatively crude) crystal clock. [18]

There are two ways for a GPS receiver to perform ranging measurements. The standard one is based on code pseudorange measurements, but also carrier-phase measurements are possible, as explained in the follow. In both cases, both the code and the carrier must be in track of the receiver, because the two components of the signal must be separated.
In the next, code will be considered cleaned from any carrier residuals, as well as for the carrier that will be considered without any residual of the code sequence. In real life, some residuals of carrier will be present on the code, as well as a residual of code modulation affects the carrier also after the code and carrier wipe-off (see Section 1.5.3).

1.5.1 Code measurement

The PRN sequence allows to identify the delay between the transmission and the receiving of the signal. The receiver generates a local replica of the PRN, and then it has to find the unknown delay, as shown in Figure 1.5. The technique to identify the delay is to correlate the incoming signal from the satellite with the local replica delayed with different delays. Thanks to the binary structure of the code, which is pseudo-random, the correlation result is almost nil (apart the noise) when the delay is uncorrect, and a correlation peak appears when the local code is delayed of the correct time.
The code measurement is quite simple, but it is not very accurate. For a typical receiver using the C/A code, the standard deviation is about 3 meters (if using P(Y) code, 0.3 meters), which corresponds to 1/100 of the code chip length.

1.5.2 Carrier phase measurement

Better accuracy than the one reached with the code measurement in Subsection 1.5.1 can be achieved with the use of carrier-phase measurements [22]. In fact, thanks to its high frequency, the carrier-phase has very fine tick marks in unit of cycles. The resolution, as well for the code measurement, is typically 1/100 of the cycle, which in this case means millimeter level accuracy. For example on GPS L1 1 cycle corresponds to $\lambda \simeq 19$ cm and 24 cm on L2, which allow resolution of about 2 mm.
Unfortunately, the carrier does not contain any information about the number of whole cycles, therefore the desired pseudorange measurement is available as a

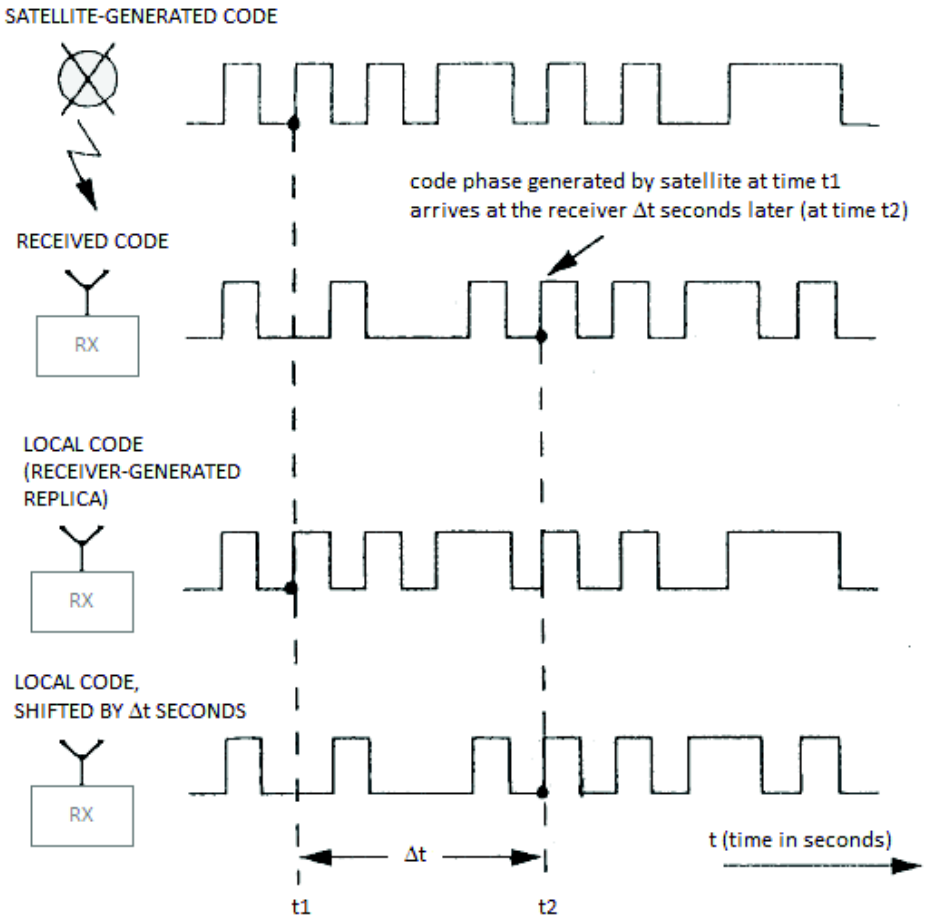

Figure 1.5: Pseudorange measurement through the measurement of the code transmission time. Source: [3]

very accurate measurement of the fractional number of cycles plus an unknown integer number N, called integer ambiguity. This can be seen as follows: at the starting up of a receiver, the carrier phase measurement starts, but before the starting up, a number of cycles already occurred along the line of sight satellite-receiver, as shown in Figure 1.6.

Methods exist to solve the integer ambiguity issue, which are not trivial. However, once the integer ambiguity is solved, as long as a continuous tracking is maintained with a certain satellite the number N does not change, therefore there is no need to solve for N again, as shown in Figure 1.6. More details about the carrier phase measurement are given in Chapter 6. The carrier phase measurement results to be very effective when the absolute range measurement is

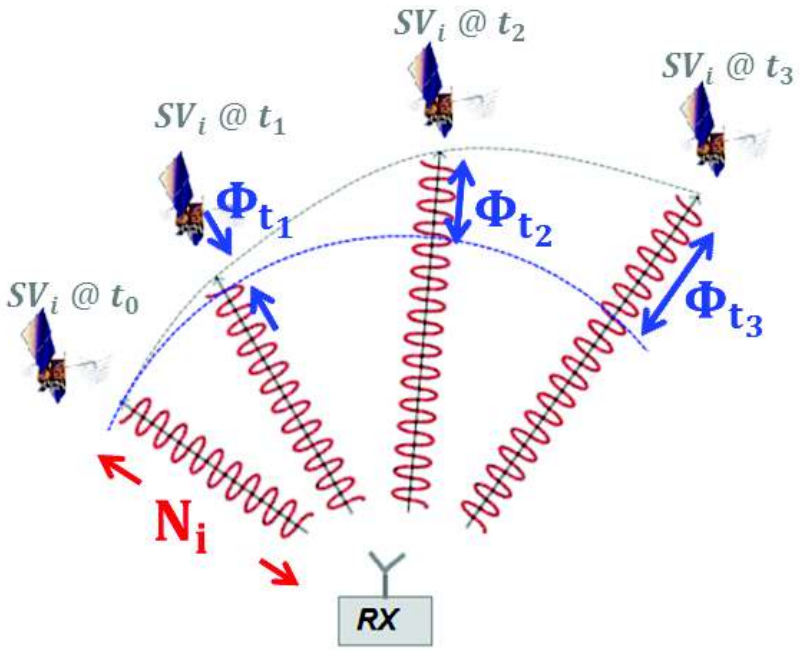

Figure 1.6: Carrier phase measurement and integer ambiguity.

.

not needed at all and therefore the integer ambiguity does not have to be solved for, as in the case of the velocity measurement. In this thesis carrier-phase measurements are used in particular as carrier-phase single differences, to compute the velocity solution (Part II), and in the difference over the time the integer ambiguity cancels out.
In part IV the carrier-phase measurements are used in the CNMP (Code Noise Multipath Algorithm) and a simple but effective method using the CMC (Code Minus Carrier) measurement is applied to mitigate the effects of the unsolved integer ambiguity.

Cycle slips

Momentary loss of phase lock (see Part III) can result in a discontinuity in the integer cycle count even though the fractional part of the phase is measured continuously. A cycle slip is defined as a discontinuity in the GNSS carrier phase measurement, which means a jump of the integer number of cycles, caused by a temporary signal loss. This signal gap can be caused by some obstacle between the satellite and the receiver, such as for example a bridge under which the user passes, or buildings, or some atmospherical bad conditions. Even if the dynamics

increase too much, if the receiver moves speedily, cycle slips could occur.
Cycle slips cause an irretrievable error. Lots of techniques exist in order to detect and repair cycle slips (e. g. [23]). Some modern receivers have sophisticated implementations of the carrier tracking loops, with few instances of cycle slips, but in the early receivers of the 1980s, which were limited both processing power and memory (so had to crude approximate in carrier tracking and recovery), cycle slips were quite frequent and very problematic.
This problem is not addressed in this thesis, and the general hypothesis is to be in benign environment in which cycle slips do not occur. Anyway, some preliminary checks on the real data used have been done in order to verify the absence of cycle slips, as described in Chapter 11.

1.5.3 Code and carrier wipe-off

In Section 1.5.1 and Section 1.5.2, an overview about code and carrier-phase measurement is given and the hypothesis is to have the received signal code, in the first case, and the signal carrier in the second one.
In real life the incoming signal is the result of a modulation, where the code is modulated at RF (at the L1 or L2 frequency, in the case of GPS).
The receiver front-end hardware operates a down-conversion to an intermediate frequency, to simplify the operations of signal processing, so that the carrier frequency results to be at a lower frequency (usually around 1-30 MHz). The resulting signal is still a combination of both code and carrier, while to perform the code and carrier measurements there is the need of separate each component. The solution is to interconnect, in the receiver, the code and the carrier tracking blocks, in order to perform, respectively, the code and the carrier wipe-off on the two branches.

1.6 Receiver structure

The block diagram structure of a GNSS receiver is shown in Figure 1.7.
The front-end provides to down-convert the input signal frequency to an intermediate frequency, then it amplifies and samples the signal.
The first stage of the processing block is the acquisition stage. The acquisition gives a rough estimate of both the code and the carrier-phase, or better, the

Doppler frequency affecting the carrier, by computing correlations both in the time and in the frequency. In this way, a search space is considered, and the estimate of code and frequency corresponds to the correlation peak. The accuracy of the estimate depends on the time and frequency step used, therefore, the higher the number of correlations is, the finer the estimate is.
The output of the acquisition stage is used to initialize the tracking blocks, that refine the estimate. Respectively, a Delay Lock Loop (DLL) and a Phase Lock Loop (PLL), usually preceded by a Frequency Lock Loop (FLL), are used to refine the code and the carrier estimate. The DLL and the PLL are interconnected to perform the code and carrier wipe-off (see Section 1.5.3). Once the data are demodulated and the navigation data is available, which provides information needed as the satellite positions and several corrections, the navigation solution can be computed. The code measurements (or carrier-phase measurements) together with the information from the navigation data in fact allows to compute the pseudoranges and the geometry matrix, that are the ingredients of the algorithm to compute the PVT solution.
While the front-end has to be developed in HW, the block of signal processing can be fully designed in software. The so called software receiver is becoming nowadays a more and more important technology.

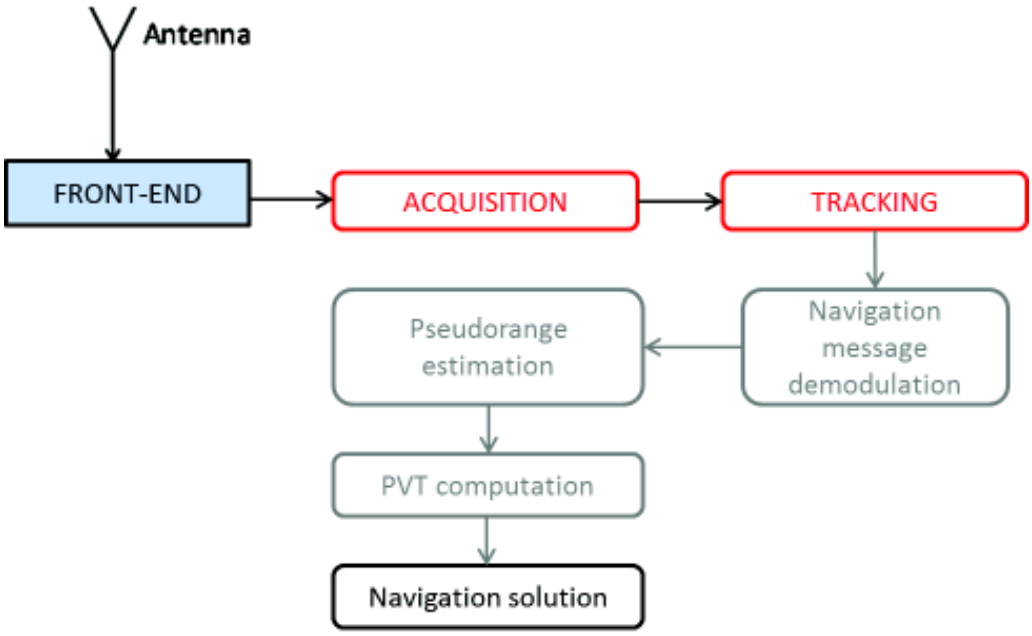

Figure 1.7: GNSS receiver, block diagram.

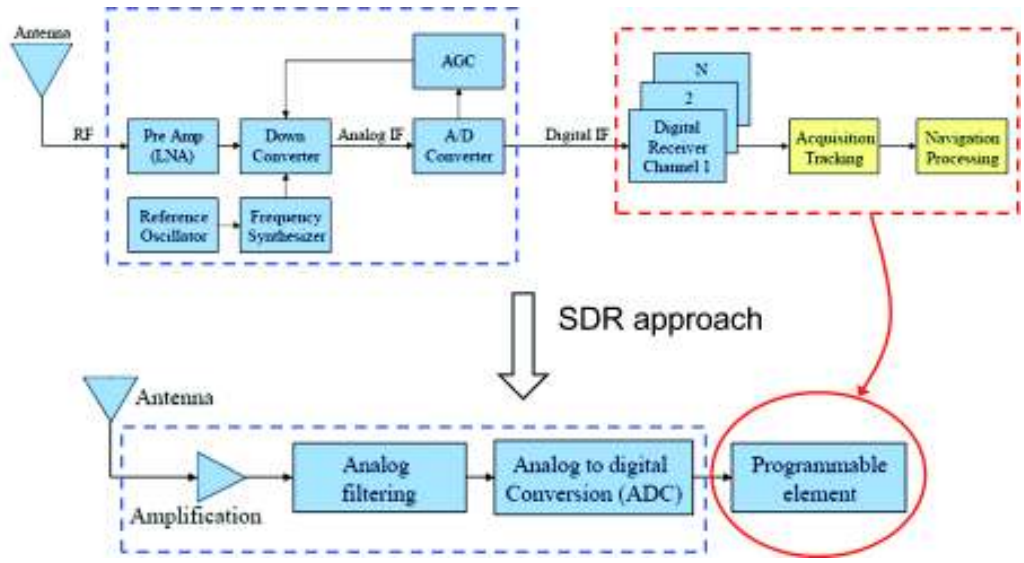

Figure 1.8: SDR architecture for GNSS receiver.

1.6.1 Software receiver

As lots of other technologies, in the last years GNSS receiver technology switched from analogical to digital architecture. Digital signal processing has many advantages in terms of performance, power, size and weight, and furthermore its software implementation has many advantages with respect to hardware implementation. Software Defined Radio (SDR) is constituted by an ensemble of hardware and software technologies that form an architecture which is easily reconfigurable. If something needs to be changed in the design it is enough to re-programme one or more blocks, without changing hardware components: SDR-enabled devices can be dynamically programmed in software in order to reconfigure the hardware characteristics. This is fundamental during the design and testing step, and in particular for research scopes. In research ambit, a fully software implementation allows to test new architectures without the need of new hardware, quickly and without additional costs. This flexibility provides an efficient and relatively inexpensive solution to the problem of building multi-functional wireless devices. In GNSS, SDR technology is very useful to implement the user receiver. As a result, the implementation of the GNSS receiver blocks in the digital domain, and their optimization, is a very actual topic [24], [25].

As shown in Figure 1.8, the parts that need to be realized in analogical are only the antenna and the front-end, then an analogical to digital converter allows to pass from the analog to digital domain.

In this thesis, a block of the sofware GNSS receiver will be analyzed and designed. A carrier-tracking block (Phase Lock Loop), which can be implemented fully in software, is described in depth in Chapter 6.

1.7 Architecture of Global Navigation Satellite System

GPS, as well as in general every GNSS, is constituted by three segments:

- space segment (satellite constellation)
- ground segment (control and monitoring network)
- user segment (receiving equipment)

[3] The interaction among the segments is shown in Figure 1.9.

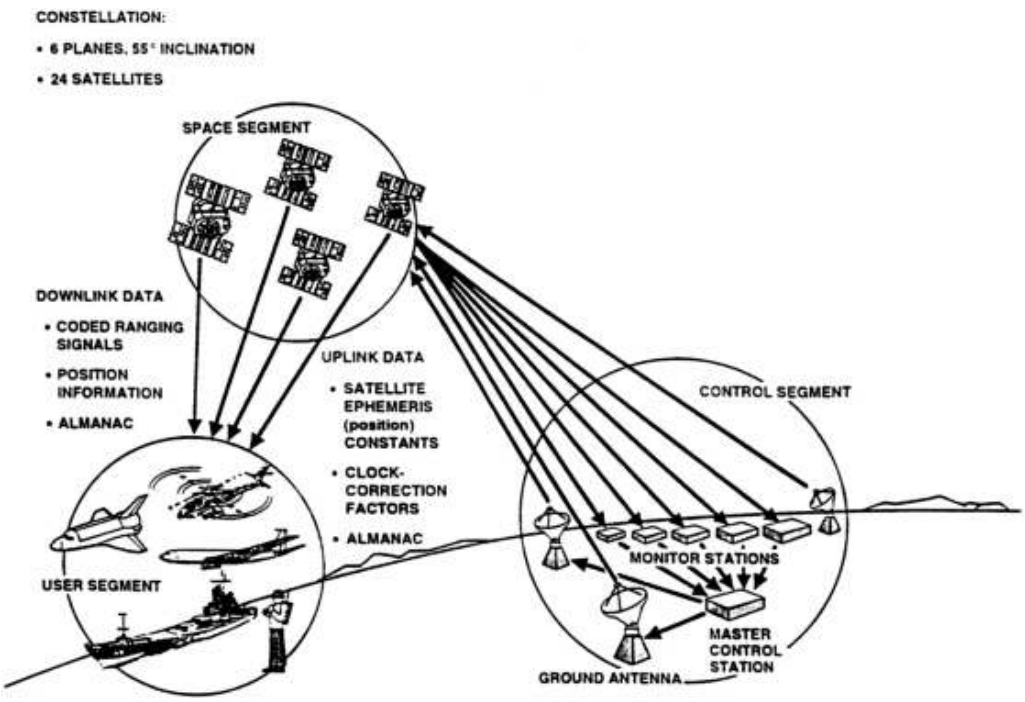

Figure 1.9: GNSS architecture: interaction between segments. Source online [4].

1.7.1 Space segment

The ensemble of satellite constituting any GNSS is the so-called *space segment*. In the case of GPS, the full constellation is made of 24 satellites but for redundancy, the complete system has 32 satellites, so that in case of outage of some of them, full coverage on the earth can be guaranteed anyway. The satellites are on six orbital planes with an inclination af about 55 degrees to the equator. The GPS satellite orbits are approximately circular, describing an ellipse with eccentricity 0.01 as maximum. The mean altitude of GPS satellite is about 20,200 km above the Earth surface and the linear velocity of the satellites is about 3.9 km/s, thus the nominal orbital period of a GPS satellite is 11 hours and 59 minutes, which means that in one day a satellite makes two complete turns around the earth. Figure 1.10 shows a representation of the GPS satellites orbits around the Earth.

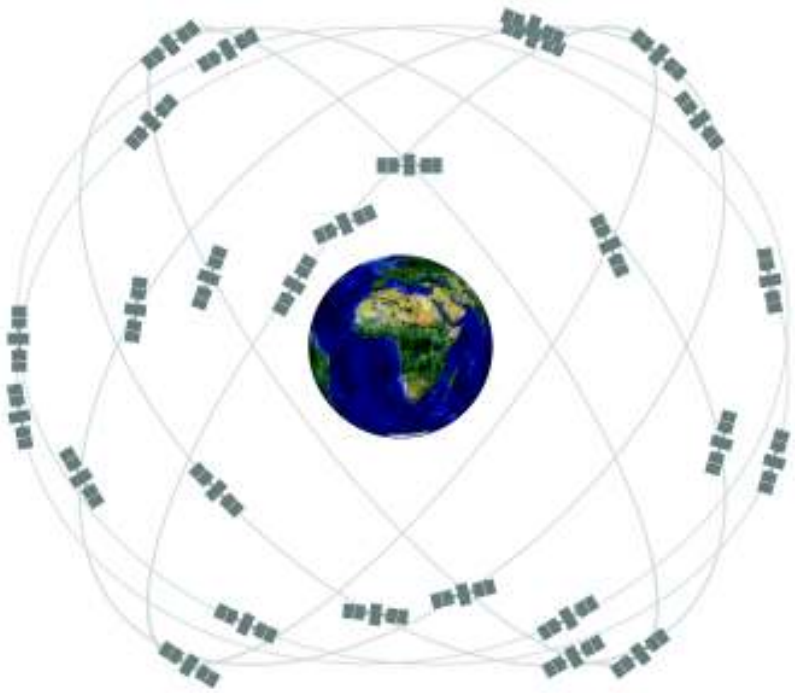

Figure 1.10: GPS constellation. Source online: [5].

As said in the introduction, the focus of this thesis is on the US GPS. Other GNSS have slightly different characteristics, for example the European Galileo system will be made of 30 satellites on three orbital planes at 29,600 km from the Earth surface. Also in this case, some satellites (3) are for backup.

The Space Vehicles (SVs) transmit signals, that all the users receive passively. Thanks to this, GNSS system can serve an ideally infinite number of users at the same time. GNSS satellites transmit PRN-coded signals together with navigation data, that include information such as the satellite ephemeris. Using the time information provided by the PRN code and the navigation data, ranging measurements can be made by the user receiver, that can compute its position applying a trilateration technique.

Notes on satellites and orbits

As outlined above, in order to determine its position each user needs at least four satellites in view. Such worldwide coverage had to be provided; clearly it had to be provided as cheaply as possible. The choice of orbits determines satellite design, number of satellites and costs of launch and maintenance.

When the first GNSS was designed, several possibilities were taken into account. Different types of satellites can be distinguished depending on the characteristics of their orbits:

- LEO: Low Earth Orbit, with altitude $<$ 2000 km, a LEO satellite is only visible for 10-20 minutes at a time, and the receiver would be acquiring new satellites constantly. Doppler rates would be high and orbital perturbations due to the atmospheric drag would be high too. It would need a constellation of at least a hundred satellites to provide global coverage. An advantage would be the lower launch cost and the lower power transmitters required.

- MEO: Medium Earth Orbit, with altitude of 5000 - 20000 km and 2-4 orbits per day, the satellites are typically visible for several hours in each pass. Launch costs are higher than those for a LEO, but a smaller satellite constellation (24 - 36 satellites) can provide global coverage. The GPS satellites are MEO.

- GEO: Geostationary orbit. Satellites, in a 36000 km orbit over the Equator appear fixed to an observer on earth. Global coverage can be provided with a small number of satellites but, besides higher launch costs, GEO satellites offer poor coverage at higher latitudes. The WAAS satellites are GEO.

1.7.2 Control segment

The Control Segment (CS) is a network of ground bases that continually monitor the satellites, to guarantee that they work properly. Messages can be sent to the SVs in order to correct their orbit if it is drifting too much, while small errors with respect to the nominal orbit are broadcasted in the navigation message so that the user can compensate for them. In case of bad working of a satellite, a flag can be raised to prevent the users from using that SV for position computing. The CS can be divided in three main parts:

- Master Control Station (MCS), located at Schriever Air Force Base (AFB) near Colorado Springs in Colorado.
- Monitor Stations (MS)
- Ground Antennas (GA), located, as well as the MS, in different locations all over the world.

A map of the GPS CS is shown in Figure 1.11.

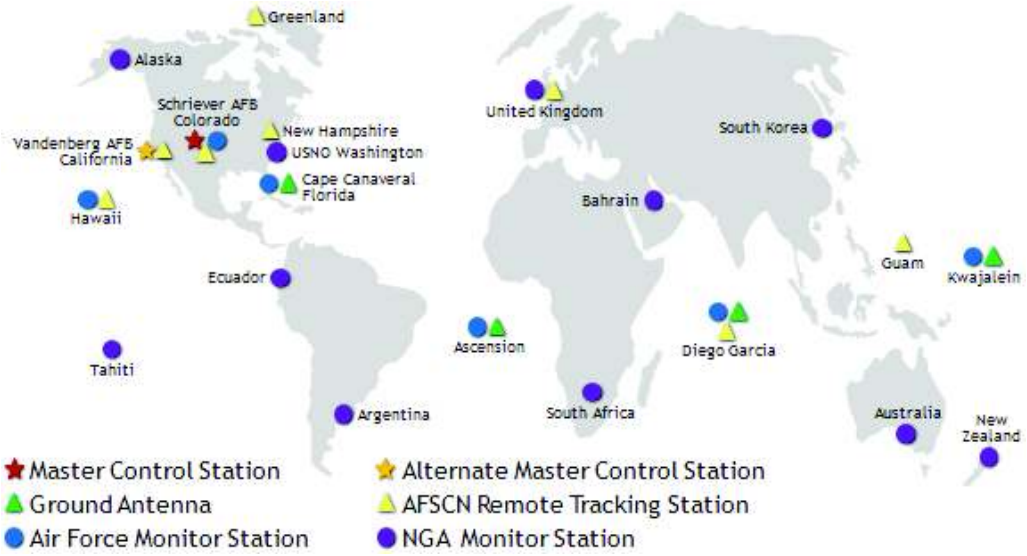

Figure 1.11: GPS control segment. Worldwide map of the ground stations. Source online: *Official U.S. Government information about the Global Positioning System (GPS) and related topics* website [5]

.

1.7.3 User segment

This thesis focus is on the user segment. It is made of the user equipment, that means a receiver, which can be of different types depending on the application. Performance can be very different, from the mass-market receivers included also in recent cellular telephones to high quality receivers used in experiments. In the cheap receivers the GNSS information is integrated with the map information, to reach performance suitable for the need, which is navigation on the roads. These receivers give as an output PVT (Position, Velocity and Time), with low accuracy and level of confidence. Professional receivers give PVT solution with higher accuracy, and can also provide the user also with other measurements such as pseudoranges or carrier phase measurements, and a measurement of the estimate's accuracy.
Some applications implying safety of life, as for example aircraft navigation, need a receiver that can satisfy also other requirements, first of all integrity (see Subsection 2.4.2).
However, each receiver has to perform the following functions:

- Identify the satellites in view,
- Measure the distance between user and the satellites,
- Estimate PVT (using a trilateration technique).

1.8 Standard method to obtain user velocity estimate

GNSS allows to solve for PVT, which includes Position and Velocity both in three dimensions, and Time. The velocity can be computed in different ways and the most common one, which is also the most intuitive, is to perform the derivative of the user position in the time. It results to be:

$$\overline{\mathbf{v}} = \frac{d\mathbf{u}}{dt} = \frac{\mathbf{u(t_2)} - \mathbf{u(t_1)}}{t_2 - t_1} \tag{1.19}$$

Notice that this measurement gives the mean velocity over the time interval dt, therefore this estimate makes sense if the velocity can be considered constant over

that time. The smaller the time interval is, the best this velocity approximates the instantaneous velocity. In fact:

$$\mathbf{v} = \lim_{\Delta t \to +0} \frac{\Delta \mathbf{u}}{\Delta t} \tag{1.20}$$

In modern GNSS receivers the velocity estimate is obtained from carrier-phase measurements, as described in Part II, where a precise velocity algorithm is applied as in [2] and other techniques are investigated to improve the velocity estimate.

1.9 Time measurement using GNSS

Last, but not least, the PVT includes the Time solution. The GPS time is referred to the UTC (Coordinated Universal Time) and it is based on the satellite atomic clock readings, statistically processed, together with references in the ground control station. In January 6th 1980, 0h, the two time scales were coincident.

A receiver has its own time reference, determined by its clock affected by an unknown bias ΔB with respect to the GPS time. The GNSS solution gives a measurement of this bias, allowing to have an estimate of the GPS time. The receiver clock time t_{rx} is

$$t_{rx} = t_{GPS} + \Delta B \tag{1.21}$$

Therefore the GPS time is

$$t_{GPS} = t_{rx} - \Delta B \tag{1.22}$$

Once the GPS time is known, the UTC can be computed as

$$t_{UTC} = t_{GPS} - \Delta t_{UTC} \tag{1.23}$$

where

$$\Delta t_{UTC} = \Delta t_{LS} + \delta t_A \tag{1.24}$$

defining Δt_{LS} as the integer number of seconds and δt_A the fractional part of the estimated difference between the GPS time and the UTC. The receiver can compute the UTC as:

$$t_{UTC} = t_{rx} - \Delta B - \Delta t_{UTC} \tag{1.25}$$

where the estimate of ΔB is provided by the GNSS PVT solution and the information on Δt_{UTC} is contained in the navigation message.

The accuracy of the estimate of ΔB directly affects the accuracy on the user UTC time definition. In particular when high-accuracy time transfer is required (better than 10 ns) (e.g. [26]), high attention must be paid to the noise and bias sources in the GNSS solution.

When the interest is on the position solution, error contributions that are in common with the satellites are usually attributed to clock effects, including these errors in the time estimate. In particular, a big error on the pseudorange measurement is introduced by the ionospheric delay, and due to the reason explained before, the time solution is the one mostly affected by the ionospheric error. The ionospheric delay can be estimated, but any residual error on this estimate must be taken into account when high accuracy is required for the time solution.

The problem of the time solution accuracy is considered in this thesis in Part IV, where the ionospheric error are studied, analyzing the errors affecting the measurement of the ionospheric Total Electron Content (TEC), a quantity which is directly proportional to the ionospheric delay.

Chapter 2

Error Budget Overview and Dilution of Precision

2.1 GNSS measurement error budget

Pseudorange code measurements and carrier-phase measurement are affected by several error sources, that arise in the transmitter, the receiver and the propagation medium. Some of them can be estimated and therefore mitigated using different techniques.

2.1.1 Satellite clock error

The time reference on-board of the satellites is much more reliable than the receiver's clock, since it is constituted by high quality atomic clocks. Anyway, a difference between these time references and the GPS time can be up to 1 ms, and the navigation data provides corrections of the satellite clock. The MCS computes satellite clock corrections and transmits them to the SV, that includes the information in the navigation message. In particular, the correction parameter is defined as:

$$\delta t_{ck} = a_{f_0} + a_{f_1}(t - t_{oc}) + a_{f_2}(t - t_{oc})^2 + \Delta t_r \tag{2.1}$$

where

a_{f_0} is the clock bias (s)

a_{f_1} is the clock drift (frequency bias) (s/s)

a_{f_2} is the frequency drift (s/s^2)

t_{oc} is the clock data reference time (s)

t is the current time epoch

Δt_r is the correction of relativistic effects.

2.1.2 Ephemeris error

The true orbit of the satellites differs from the theoretical one and an estimate of this misalignment is computed by the CS and it is transmitted in the navigation data. After this correction, a residual error still exist, which can be represented by a vector as shown in Figure 2.1, whose length is typically 1-6 meters and impact on the pseudorance (or carrier-phase) measurement with an error of about 0.8 meters [3]. Precise ephemeris are available also, but not in real-time.

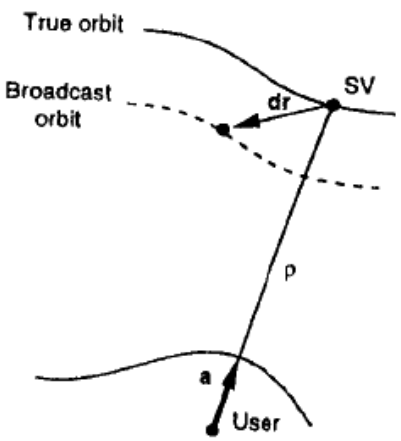

Figure 2.1: Ephemeris residual error, vector representation. Source [3].

2.1.3 Relativistic effects

Different types of relativistic corrections are needed in GNSS. In fact, if in a tx-rx system the transmitter and/or the receiver are located at different gravitational

potentials, general relativity relativistic corrections are needed.
Moreover, if transmitter and/or receiver move with respect to the isotropic light speed frame chosen as reference, which in the case of GNSS is the ECI (Earth Centered Inertial) frame, special relativity relativistic corrections are needed.
Both corrections are needed in GNSS because of the satellite clock, which is affected by both the relativistic effects described, and to compensate for it a correction is applied to the satellite clock prior to launch it. Furthermore, the user has to compute a correction to compensate for the effects of the satellite orbit eccentricity. The satellite clock has to be corrected in order to adjust the time of transmission seen by the user, and the correction is:

$$\Delta t_r = Fe\sqrt{a}sinE_k \tag{2.2}$$

where $F = -4.442807633 \cdot 10^{-10} s/m^{1/2}$, while the other two parameters, referred to the satellite orbit geometry, are the eccentricity e, the semi-major axis a and the eccentricity anomaly E_k [3].
When the satellite position is computed in the ECEF system, another relativistic effect is involved, due to the Sagnac effect caused by the Earth rotation. After the signal trasmission, during its propagation through the LoS, before it reaches the receiver, the receiver itself changes its position, since it rotates with the Earth. If the rotation is towards the satellite, the propagation time decreases, otherwise, a delay is introduced. Neglecting this effect can cause a position error up to 30 m. Details on this error source and its correction methods are available in [18].

2.1.4 Atmospheric effects

The TOA method implies the measurement of the flight time which gives the pseudorange measurement if multiplied by the propagation velocity. This is assumed to be the speed of light, which is a good approximation since most of the signal path is in the free space. In fact the satellite elevation from the Earth surface is about 20,000 km while the Earth's atmosphere is only about 1,000 km high. Even if the approximation is good, a not negligible error is due to the effects of the atmosphere, in particular to the troposphere and to the ionosphere. In the following, a brief overview of the main atmospheric effects is depicted. Further details on the ionospheric effects will be provided in Chapter 9.

Tropospheric delay

The troposphere is the lower layer of the atmosphere (up to 20 km of altitude), composed by non-ionized gases. The propagation constant for the electromagnetic waves passing through it is different from the free-space one, but it is the same for all the frequencies lower than 15 GHz, which means that at those frequencies the troposphere behave as a non-dispersive medium.
The propagation velocity is lower than the speed of light, therefore a delay is introduced with respect to the free-space propagation. The pseudorange error due to the troposphere delay can vary between 2.4 m and 25 m, depending on several parameters such as temperature, pressure, relative humidity, factors that affects the refractive index.
Furthermore, the delay depends on the satellite elevation and on the user elevation, in fact the lower the satellite is above the horizon, the longer the signal path through the troposphere is (The same concept is valid for the ionospheric effect, as explained in 10.2).
A model is used to compute estimates of the vertical tropospheric delay, given meteorological data such as the pressure and the temperature. Since the delay affecting the pseudorange depends on the satellite and user's position, a mapping function is then applied to convert the vertical correction into the one relative to the LoS.
The corrections that has been applied in the experiments shown in this thesis make use of the formula:

$$\Delta S_{tropo} = m_d d_{dry} + m_w d_{wet} \tag{2.3}$$

where d_{dry} and d_{wet} are the dry and the wet component of the delay respectively, depending on the meteorological parameters, and m_d, m_w the dry and the wet component mapping functions, related to the satellite and user's position. The details of the method are available in literature, e.g. [3].

Ionospheric delay

The ionosphere, as well as the troposphere, is a layer of the atmosphere that affects the signal propagation, and several methods exist to mitigate these errors. This topic is discussed in detail in Part IV.

2.1.5 Receiver errors

The receiver itself introduces errors on the pseudoranges measurements, in particular in the tracking blocks, which have the aim to track the signal code and carrier.
The code tracking is performed by a Delay Lock Loop (DLL), whose main error sources is represented by thermal noise (tracking jitter), after multipath (see Subsection 2.1.7) which if present can represent the biggest error source, and interference. In general, code noise and resolution error contribution are on the order of decimeters or less, negligible with respect to the multipath error [3].
The Phase Lock Loop (PLL) is in charge of the carrier-phase measurement, which resolution can be much better than the code measurement resolution (as discussed in 1.5.2). Also PLL is affected by receiver thermal noise and resolution errors, which nominally are on the order of 1-2 mm (1.2 mm for C/A code, 1.6 mm for P(Y) code). The PLL is described in depth in Part III.

2.1.6 Hardware errors

Receiver HW biases

During the path from the receiving antenna to the receiver's processing blocks that compute the pseudorange and carrier-phase measurements, GNSS signals are delayed from the HW they pass through and this delay can be quite large reaching at times a duration on the order of even 1 μs.
Even if the absolute receiver HW delay is large, it is constant among the satellites, therefore it appears as a common bias just like the receiver clock bias. This means that in the GNSS solution this bias turns into being an error that affects the clock bias estimate, while it does not affect the position estimate. This is the reason why in general the receiver bias is neglected in normal GNSS applications, but it is also the reason why the absolute delay is fundamental in timing applications, for which it must be calibrated. Notice that the problems discussed in this paragraph (Receiver HW biases) will be the crucial element in Part IV.
Furthermore hardware bias width changes with the power spectra of the incoming signal, because GPS receiver front ends do not have the same group delay through the whole passband, therefore the PRN code type influences the signal power spectra. In fact C/A code signals have measurably different spectra due to their

short PRN code, and as a consequence the C/A code measurements result to be affected by inter-satellite biases, even if quite small (on the order of a few millimeters).
Carrier at different frequencies suffer different receiver HW biases, in fact the L1 and L2 (in case of GPS) signals may follow at least in part different physical paths in the receiver. This error is included in double-frequency estimates as the TEC measurement, as described in Chapter 10, and cannot be neglected when high-accuracy TEC and time measurements are required.
Last effect that can occur in the receiver, if it is not designed properly, is the hardware induced multipath, consisting in internal signal reflection caused by impedance mismatches.

Satellite HW biases

As well as the receiver, also the satellite hardware introduces undesired effects.
A synchronization problem exists between the L1 and L1 signals, therefore a clock bias between them may exist, that results to be mitigated in the position solution if the double frequency ionospheric correction is introduced (see Chapter 10). This error has to be corrected by the single-frequency users using a broadcast information, available in the navigation data, that provides a correction with accuracy of about 0.5 ns.
In addition to the inter-frequency biases, an inter-code bias is present. Timing bias occurs in fact between different codes (P(Y), C/A in case of GPS), that causes typical errors around 3 ns, and is guaranteed to be lower than 10 ns. When high-accuracy timing is required, these errors are not negligible at all.
Corrections for these errors are available from several organizations, that estimate these biases by means of networks of receivers, as described in Part IV.

2.1.7 Multipath

Multipath occurs when reflected or diffracted replicas of the desired signals arrives to the receiver's antenna. The resulting received signal is therefore a composition of the desired one plus the delayed replicas. Because of the nature of the multipath rays, their path is longer than the one of the direct signal, therefore the multipath has a delay with respect to the desired signal. The smaller the delay of the multipath is, the higher the distortion introduced on the correlation

function is, in particular if the multipath attenuation is small. The distortion affects the pseudorange and the carrier phase measurement, causing errors in the PVT solution.

2.1.8 Interference

Interference can be of several kinds, in particular, two main types of interference exist:

- unintentional interference
- intentional interference, including
  - jamming
  - spoofing

This is not a topic analyzed in this thesis, anyway, the effects of unintentional interference are in general similar to the effects of multipath or noise, causing distortion of the correlation function. The intentional interference instead is more dangerous, in fact the jamming consists in disturbance at high power in the GNSS bandwidth [27], that aims to blind the receiver, while the spoofing aims to trick the receiver with fake GNSS signals [28].

2.2 Pseudorange error budget

The User Equivalent Range Error (UERE) is globally due to components from all the three system segments. It has to be noted that if the ionosphere error is mitigated by means of double-frequency measurements, the error budget has to be based on the two-frequencies measurements.
It is important to have a global view of the effects of the several error sources on the GNSS solution, in order to be aware of the accuracy that can be obtained if an effect is mitigated or not. Table 2.1 gives an overview of the typical UERE budget.

Table 2.1: UERE error budget [3].

Segment source	Error source	1σ error (m)
Space/control	Broadcast clock	1.1
	L1 P(Y)-L1 C/A group delay	0.3
	Broadcast ephemeris	0.8
User	Residual ionospheric delay	7.0
	Residual tropospheric delay	0.2
	Receiver noise and resolution	0.1
	Multipath	0.2
System UERE	Total (RSS)	7.1

2.3 Concept of Dilution of Precision

2.3.1 Ranging accuracy and position accuracy

Position accuracy is not the same as the measurement accuracy, nevertheless, it can be easily guessed that the larger the errors in the measurements, the larger the errors in the position estimate. It can be expected that the higher the number of satellites used for the solution computation is, the better the solution.
However, the number of satellites is not the only parameter that influences the position accuracy. The LMS navigation solution is decided by two factors, i.e. the quality of the measurements (pseudoranges) and the user-to-satellite *geometry*. In the following this topic will be discussed, both qualitatively and quantitatively.

2.3.2 Satellite geometry

The point is that the range measurements are not perfect, and their uncertainty translates into positioning uncertainty in different way depending on the relative position between the satellites and the user. The position is in fact computed as intersection between spheres, whose surfaces are not well defined, but have a thickness that depends on the range accuracy. The smaller the accuracy on the range estimate, the wider the undefined region of the sphere border. Depending on the geometry, this region of intersection among these spheres is smaller or wider. The bigger the region of intersection is, the higher the uncertainty on the PVT solution is.
Due to this it can also be said that adding to the solution a satellite with a geometry very similar to another satellite already used does not improve a lot

the solution.

The worse the geometry, the higher the Dilution of Precision (DOP). The term DOP is because different satellite geometrical configurations cause a different *dilution* in the solution precision ([3], [29]). In Figure 2.2, the geometry effects are shown qualitatively.

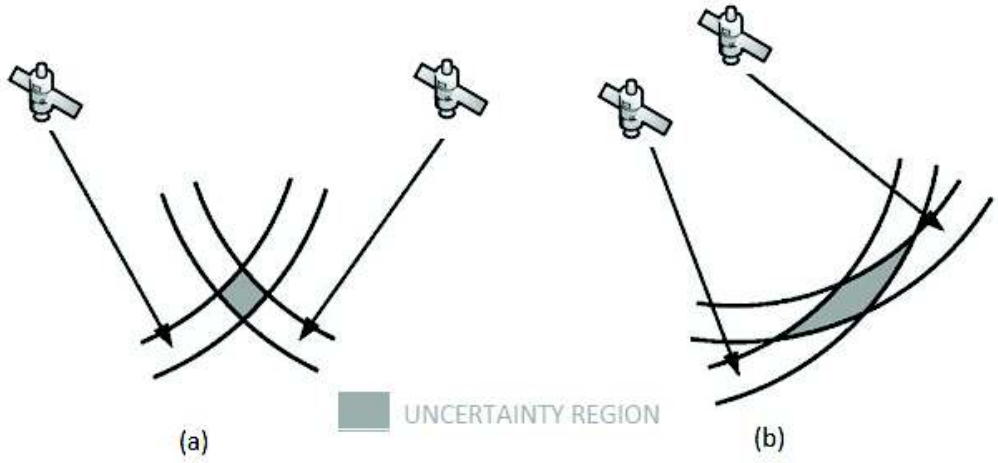

Figure 2.2: Satellite geometry and uncertainty region: basic concept.

2.3.3 DOP measurement

In section 2.3.2 a qualitative analysis of the geometry effect was depicted. A quantitative analysis can be done ([30], [3]), as shown in this section.

As discussed in Section 1.3, the ENU coordinate frame provides information easier to be understood from the user, with respect to the ECEF coordinates. In particular, talking about positioning error, it is in general more clear to talk about horizontal and vertical error components.

Defining $\delta\rho$ and $\delta\mathbf{x}$ the errors in the pseudorange measurements and in the position estimate respectively, from (1.16) it follows:

$$\delta\mathbf{x} = (\mathbf{H^T H})^{-1}\mathbf{H^T}\delta\rho \tag{2.4}$$

If the pseudorange measurements can be assumed to be independent and the error $\delta\rho$ on each pseudorange can be considered indipendent but identically distributed with Gaussian distribution, zero mean and variance σ^2_{UERE}, then the solution covariance matrix becomes:

$$\text{cov}(\delta\mathbf{x}) = (\mathbf{H^T H})^{-1}\sigma^2_{\text{UERE}} \triangleq \mathbf{G}\sigma^2_{\text{UERE}} \tag{2.5}$$

where the geometry matrix G is defined as:

$$\mathbf{G} = (\mathbf{H^T H})^{-1} \tag{2.6}$$

Details are available in [3].
As outlined before, (2.5) shows the two factors involved in the determination of the solution accuracy:

- Satellite geometry, represented by the geometry matrix **G**. A parameter used to characterize the performance of a certain satellite configuration is the so-called GDOP (Geometrical Dilution of Precision), defined as the trace of the matrix G, $GDOP = \sqrt{tr\{\mathbf{G}\}}$;
- Variance of the pseudorange measurement σ^2_{UERE}, due to all the effects causing degraded accuracy on the measurement.

Figure 2.2 showed qualitatively the effects of satellite geometry can be. In Figure 2.2, the GDOP parameter is worse in the case (a) while it is much better in the case(b). Figure 2.3 shows the GDOP meaning.

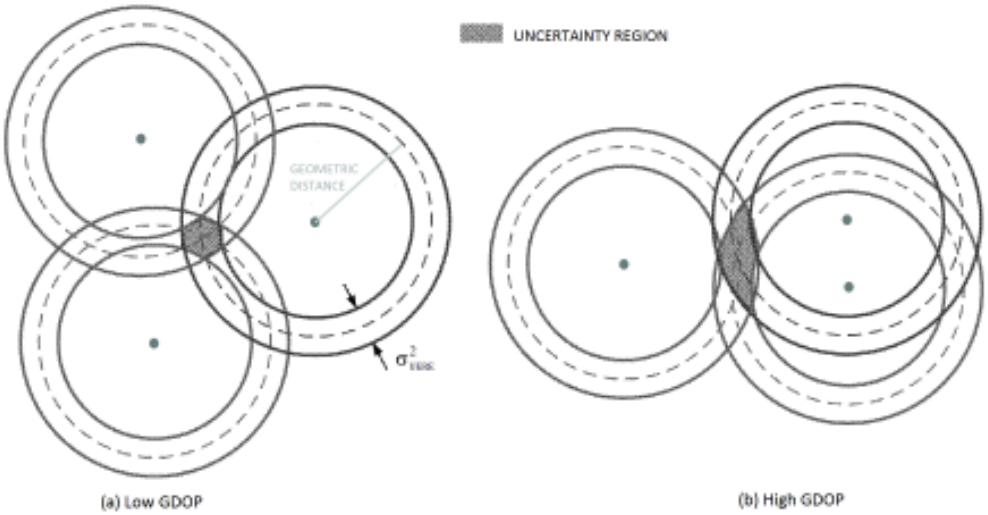

Figure 2.3: Effects of the satellite geometry on the solution accuracy: Dilution Of Precision (DOP).

The GDOP is a parameter that changes depending on the location on the Earth and with the hour of the day. The higher the latitude, the lower the number of satellites in view, therefore the poorer the geometry. In Figure 2.5 an example

of GDOP variations during one day in a specific location is shown. Nowadays, thanks to the new GNSS, more satellites are in view, therefore better performance can be achieved in terms of DOP. The result in Figure 2.5 is referred

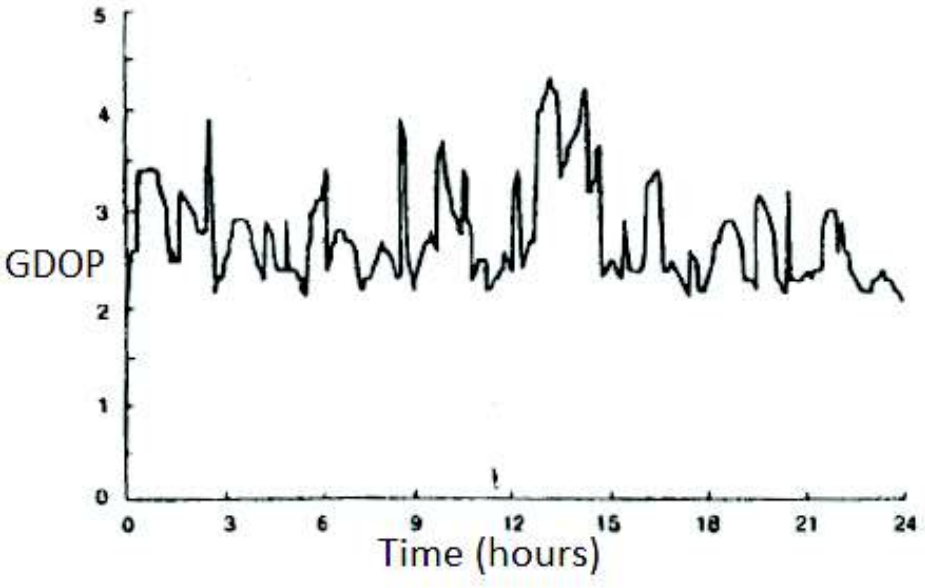

Figure 2.4: Example of GDOP variations over 1 day. Boston, MA. Source [6], *Chapter 7, p.270.*

to an open-space situation, without any obstacle along the line of sight user-satellite.

Of course, the environment affects the performance in terms of geometry. In environments such as the so called *urban canyons* (see Figure 2.5) the performance results much worse, due to obstacles that prevent a user to see many satellites above the horizon, in particular, allowing the user to see only the satellites that lay along the same direction, the only one that is not covered by obstacles. Therefore, in such a situation, even if many satellites are in view, the GDOP parameter results to be bad, because of the spatial dislocation of the in-view-satellites, that does not allow to add much information even adding more satellites.

DOP and local frequency/time reference

Due to the geometry of any GNSS system, it happens that the DOP in the vertical direction, in the user coordinate frame, results to be in general much worse than on the horizontal directions. This is because, being the earth not transparent,

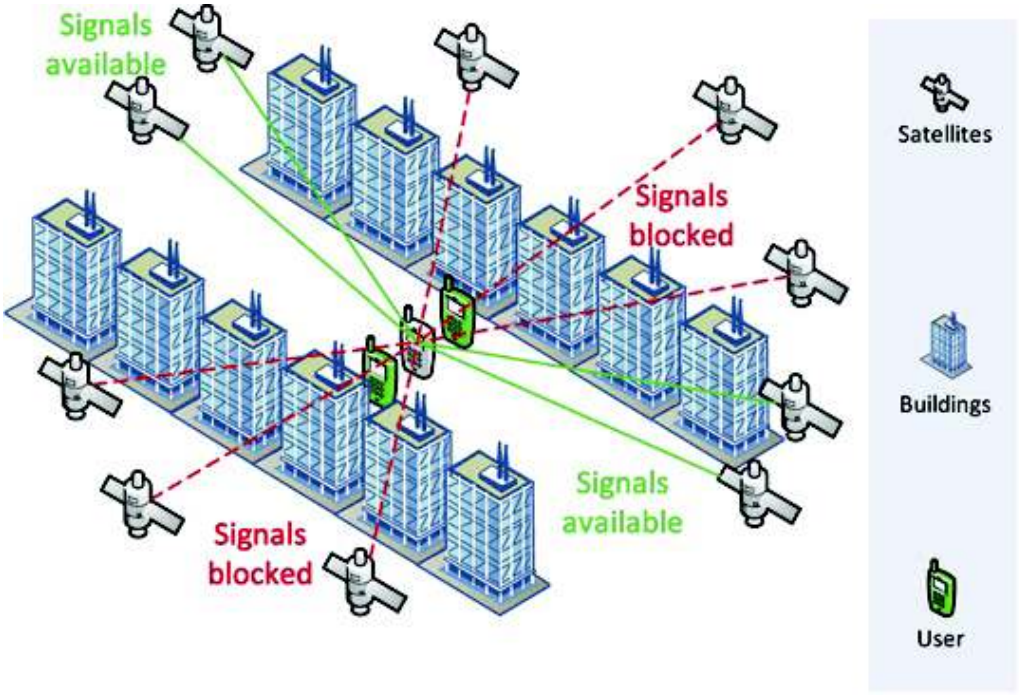

Figure 2.5: Urban canyon environment: due to the obstacle along the Line of Sight user-satellite, the geometric DOP results to be very poor. Source [7]

the user can see the satellite only above the horizon, so it can see satellites around 360Â° with respect to the horizontal directions, but only around 180Â° with respect to the vertical. As a consequence, the geometry results in general worse along the vertical direction, which causes the vertical DOP to be higher. This means that any error on the measurements (both code pseudorange and accumulated Doppler measurements) have higher impact on the vertical solution than on the other directions. In particular, the measurement error due to the local clock has higher impact on the vertical. This can be seen in the example in (2.7), where G is the matrix defined in (2.6). In (2.7) it can be noted how the cross correlation between the vertical and the clock solutions, represented by the element (3,4) or the element (4,3), is very high.

$$\mathbf{G} = \begin{bmatrix} 0.36 & 0.01 & 0.21 & -0.19 \\ 0.01 & 0.54 & 0.14 & -0.10 \\ 0.21 & 0.14 & 2.10 & -1.16 \\ -0.198 & -0.10 & -1.16 & 0.79 \end{bmatrix} \tag{2.7}$$

This effect is the main reasons why in this thesis one of the focus is on the

vertical solution improvement, analyzing and modelling the effects of the local oscillator (see Part II). Note that the focus here is on the velocity solution, which is correlated to the frequency of the local oscillator, while the position estimate is related to the local clock phase.

2.4 Other basic concepts

2.4.1 Availability

In GNSS, availability is defined as the percentage of time in which it is possible to use the system. The navigation service is defined as usable if it is satisfying a certain accuracy requirement, which depends both on the transmitting devices (satellites) and on the environment, influencing the coverage of the satellites.
For some applications, the concept of availability include also other criteria, and in those cases the system is considered to be available only if all the requirements are satisfied.

2.4.2 Integrity

Integrity is an important concept that must not be confused with availability.
Integrity is the capability of the system to provide timely warnings when the system should not be used, because there is high probability that in that moment something is going wrong in the system, and the information that is given as output is not trustable. Integrity is therefore a measure of the reliability of the correctness of the system output. A system can be available, even if according to the integrity concept it is not reliable: some output information is available, but also a warning flag is raised.
For a good working of a GNSS system, integrity anomalies must have a very low probability (order of few times per year). However integrity is a very crucial topic, in particular for critical applications implying safety of life, such as air navigation.

Part II

Frequency Aided Precise Velocity

Chapter 3

Precise Velocity Algorithm

3.1 Background overview

Standard stand-alone GNSS receivers provide PVT (Position, Velocity and Time) solution computing the position using the pseudorange code measurements and the velocity by means of the code measurement 1.8 or, sometimes, the carrier phase measurement (Ex. [31]), which allows to achieve much better performance. In literature several approaches have been proposed in order to improve accuracy and precision on the velocity solution. In particular, odometer and INS (Inertial Navigation Unit) aided solution are very used [32].
Some considerations can be done about the GNSS solution method, and a substantial improvement can be included.
As van Graas and Soloviev proposed in [2], first of all an analysis on the geometry of the velocity solution needs to be done. Since the velocity cannot be measured instantaneously, everything that changes over time, apart the user position, needs to be investigated and mitigated, in order to improve the estimate. The first step applied in this thesis is then to include the precise velocity algorithm proposed in [2], in order to correct deterministic errors and the satellite changes in geometry and Doppler.
In this chapter, the precise velocity algorithm proposed in [2] and applied in this thesis is recalled. Note that in this chapter the effects of the local oscillator are considered unknown to solve for, as in [2], while a novel technique is investigated in Chapter 5 to apply an improved frequency estimate to the GNSS velocity solution.
In this chapter, first a recall is presented of how the least squares method can

be applied to solve for the change in the user position, using as measurements the carrier-phase single differences (SDs). Then the method proposed in [2] is recalled, which allows to correct the measured SDs for the deterministic errors and to adjust them for both the satellite motions and for the changes in the geometry over the time interval considered. Finally, the LS method is applied using as measurements the carrier-phase SDs, corrected and adjusted.

3.2 Velocity estimate using carrier-phase SDs

In Subsection 1.5.2 pros and cons of the use of carrier phase measurement to compute the user position are depicted. In the velocity estimate, the con observed in using the carrier phase measurement, related to the integer ambiguity, is not involved anymore. In fact in the velocity solution the interesting information is constituted by the change in the range, over the time. As far as the lock is maintained, the integer ambiguity does not change and it cancels if single differences are performed over time of the carrier phase measurements. In this way, all the advantages of the carrier phase measurement can be obtained, without having to solve for the ambiguity. Thanks to the carrier-phase measurement, high accuracy can be reached, as discussed Subsection 1.5.2 and for velocity estimates mm/s level accuracy is enabled.
The system

$$\mathbf{H}\boldsymbol{\Delta}\hat{\mathbf{x}} = \boldsymbol{\Delta}\tilde{\boldsymbol{\varphi}} \tag{3.1}$$

is considered, where $\boldsymbol{\Delta}\hat{\mathbf{x}}$ is the vector containing the change in the user's position and also the change in the user's clock error (as explained in Chapter 1, the solution has to be found for three unknown related to the user position in three dimensions and for the receiver clock error). $\boldsymbol{\Delta\varphi}$ is the change in the carrier-phase measurement, that in (3.2) is expressed in meters. In (3.2), as in the following, the carrier-phase SD is expressed in meters, considering the phase measurement, which is $\tilde{\varphi}_{rad}$, undimensioned (degrees or radiants), to be multiplied by the propagation velocity, i.e. the speed of light c. It is

$$\tilde{\boldsymbol{\varphi}} = \tilde{\boldsymbol{\varphi}}_{\boldsymbol{rad}} \cdot c \tag{3.2}$$

H is the geometry matrix.
The unknown that we want to solve for is the vector:

$$\mathbf{\Delta\hat{x}} = \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \\ \Delta B \end{bmatrix} \tag{3.3}$$

where the Δx,Δy,Δz represents the user displacement over time along three directions, in meters, while ΔB is the receiver clock error, in meters. To obtain the receiver clock error in seconds, ΔB needs to be divided by the speed of light.

The vector of the measured carrier-phase SDs is:

$$\mathbf{\Delta\varphi} = \begin{bmatrix} \Delta\tilde{\varphi}_1 \\ \Delta\tilde{\varphi}_2 \\ \vdots \\ \Delta\tilde{\varphi}_N \end{bmatrix} \tag{3.4}$$

where, for $j = 1,2,...N$,

$$\Delta\tilde{\varphi}_j(t_2) = \tilde{\varphi}_j(t_2) - \tilde{\varphi}_j(t_1) \tag{3.5}$$

The geometry matrix is

$$\mathbf{H} = \begin{bmatrix} \mathbf{e_1}^T(t_2) & 1 \\ \mathbf{e_2}^T(t_2) & 1 \\ \vdots & \vdots \\ \mathbf{e_N}^T(t_2) & 1 \end{bmatrix} \tag{3.6}$$

where $\mathbf{e_i}$ are the line-of-sight unary vectors (as in (3.12)) and the column of ones is due to the fact that the contribution of the clock frequency error on the SD measurements is the same for all the satellites.

The solution that has to be computed is the vector $\mathbf{\Delta\hat{x}}$ containing the estimates of the changes in the user position and in the user clock. The standard way to solve for it is the use of a Least Square (LS) algorithm applied to carrier phase differences:

$$\mathbf{\Delta\hat{x}} = (\mathbf{H}^T \cdot \mathbf{H})\mathbf{H}^T\mathbf{\Delta\varphi} \tag{3.7}$$

The velocities along the three directions can be computed from the vector $\mathbf{\Delta\hat{x}}$ dividing by the time interval. The velocity components are given, in m/s, by:

$$\hat{v}_x = \frac{\Delta\hat{x}}{\Delta t}; \hat{v}_y = \frac{\Delta\hat{y}}{\Delta t}; \hat{v}_z = \frac{\Delta\hat{z}}{\Delta t}; \tag{3.8}$$

and precise frequency is, in Hz:

$$\hat{f} = \frac{\Delta\hat{B}}{\mathrm{c}\Delta t} \tag{3.9}$$

where $\Delta t = t_2 - t_1$.

The average velocity during the time interval is:

$$\bar{v} = \frac{\sqrt{(\Delta\hat{x})^2 + (\Delta\hat{y})^2 + (\Delta\hat{z})^2}}{\Delta t} \tag{3.10}$$

What needs to be measured is the change in the carrier phase caused by the user motion. What actually happens is that there are more factors that make the phase changing, so that by measuring the difference $\Delta\varphi$ lots of other effects are involved. The measured carrier phase SD at a time t_2 with respect to a previous time t_1 are, for the generic $j-th$ satellite:

$$\Delta\tilde{\varphi}_j = \tilde{\varphi_j}(t_2) - \tilde{\varphi_j}(t_1) = \Delta R_j + \Delta B + \Delta err_j + \Delta n_j \tag{3.11}$$

where ΔR_j represents the change in pseudorange that is affected by the user motion we are interested in, but also by the satellite motion; Δerr_j is the change in the deterministic errors, related to the $j-th$ satellite and Δn_j represents noise and changes in the error due to the multipath (on the $j-th$ satellite). ΔB represents the change in the receiver clock, i.e. the local oscillator frequency bias, that will be the focus of Chapter 4 and Chapter 5. Note that ΔB is a common error to all the pseudoranges (does not change with the satellite, since it only depends on the receiver oscillator). In this chapter, as in [2], ΔB is considered an unknown to solve for. In Chapter 4 and in Chapter 5, the oscillator performance will be investigated and methods to correct its error will be considered.

3.3 Precise velocity algorithm: measurement compensation and adjustment

As proposed in [2], the first step to improve the velocity estimate is to take into account some effects due to the fact that the velocity cannot be measured instantaneously. In order to estimate the user velocity, differences in the pseudorange

(and therefore in the carrier phase measurements), have to be performed over time. The velocity is a function of the change in the user position, but other components change over time and affect the measurement, which have to be mitigated in order to obtain precise velocity solution. First of all, carrier phase measurements have to be compensated for the change in the deterministic error during the time. Furthermore, it has to be taken into account that the LMS velocity solution is computed as in (3.2), where the geometry matrix is as in (3.6). While the velocity measurement is referred to a difference above times t_1 and t_2, as shown in Figure 3.1, the geometry matrix is referred to an instant (t_2). Therefore, in order to achieve better precision, an adjustment is needed on the carrier phase difference measurements, to take into account the change in the geometry, i.e. in the relative position between user and satellites. In addition, the Doppler frequency caused by the satellite motion also has to be mitigated, since the interest is on the user position change.

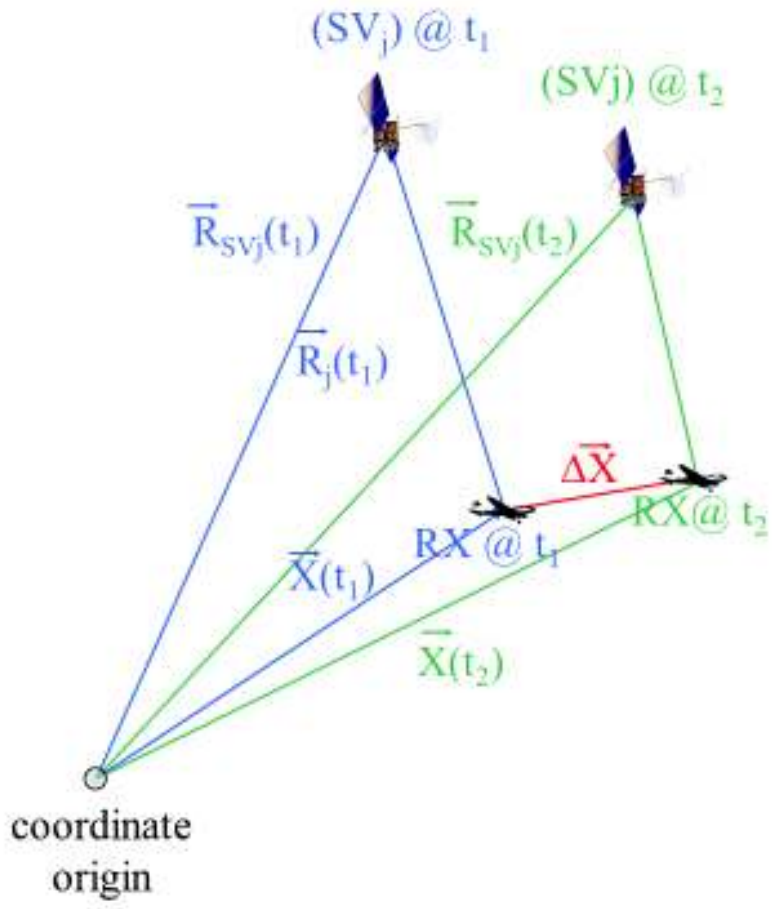

Figure 3.1: Receiver-Satellite geometry change over time

The precise velocity algorithm, proposed in [2] and applied also in this thesis, is briefly recalled in this subsection. First of all, the deterministic errors are mitigated. Lots of errors affecting the position solution are constant over a short time interval, therefore using the single difference measurements they cancel out. The errors that in this application can be considered to be deterministic are those errors such as changes in ionospheric delay and tropospheric errors, that can be estimated, respectively, with double frequency estimates (see Chapter 10) and with tropospheric models. Satellite clock frequency errors and drift components of relativistic corrections are present, that can be corrected with information from the ephemeris data.([2]). The velocity information is included in the range single difference expressed as follows:

$$\begin{aligned} \Delta R_j(t_2) &= R_j(t_2) - R_j(t_1) \\ &= (\mathbf{e_j}(t_2),\mathbf{R}_{SV_j}(t_2) - \mathbf{b}(t_2)) - (\mathbf{e_j}(t_1),\mathbf{R}_{SV_j}(t_1) - \mathbf{b}(t_1)) \\ &= (\mathbf{e_j}(t_2),\mathbf{R}_{SV_j}(t_2)) - (\mathbf{e_j}(t_1),\mathbf{R}_{SV_j}(t_1)) - (\mathbf{e_j}(t_2),\mathbf{b}(t_2)) + (\mathbf{e_j}(t_1),\mathbf{b}(t_1)) \end{aligned}$$

where the vector $\mathbf{b}(t_i)$ represents the receiver position at the time (t_i) and the line-of-sight unit vector is:

$$\mathbf{e_j}(t_i) = \frac{\mathbf{R}_{SV_j}(t_i) - \mathbf{b}(t_i)}{|\mathbf{R}_{SV_j}(t_i) - \mathbf{b}(t_i)|} \tag{3.12}$$

and the user/satellite range, at the time t_i, with $i = 1,2$, is the inner product:

$$R_j(t_i) = (\mathbf{e_j}(t_i),\mathbf{R}_{SV_j}(t_i) - \mathbf{b}(t_i)) \tag{3.13}$$

The details of the algorithm are explained in [2].
After the mitigation of the deterministic errors [2], the single difference measurements are

$$\Delta\tilde{\varphi}_j{}^c = \tilde{\varphi}_j(t_2) - \tilde{\varphi}_j(t_1) = \Delta R_j + \Delta B + \Delta n_j \tag{3.14}$$

that still contain the effects of the change in the geometry. The change in the receiver position $\mathbf{b}$ is the vector

$$\mathbf{\Delta b} = \mathbf{b}(t_2) - \mathbf{b}(t_1) = (t_2 - t_1)\bar{\mathbf{v}} \tag{3.15}$$

which is a function of the average velocity $\bar{v}$ that needs to be evaluated. The measurement available is the pseudorange measurement affected by Doppler errors induced by the satellite motion.
Furthermore, the solution is computed with the geometry matrix $\mathbf{H}$ that does

not take into account the change in the geometry between time t_1 and and time t_2. A geometrical correction is introduced to take into account this effect.
The corrections for the satellite Doppler and the geometry changes are, respectively:

$$\Delta f_{d_j} = (\mathbf{e_j}(t_2),\mathbf{R_{SV_j}}(t_2)) - (\mathbf{e_j}(t_1),\mathbf{R_{SV_j}}(t_1)) \tag{3.16}$$

$$\Delta g_i = (\mathbf{e_j}(t_2),\mathbf{b}(t_2)) - (\mathbf{e_j}(t_1),\mathbf{b}(t_1)) \tag{3.17}$$

In figure 3.2, the block diagram of the precise velocity algorithm is shown.

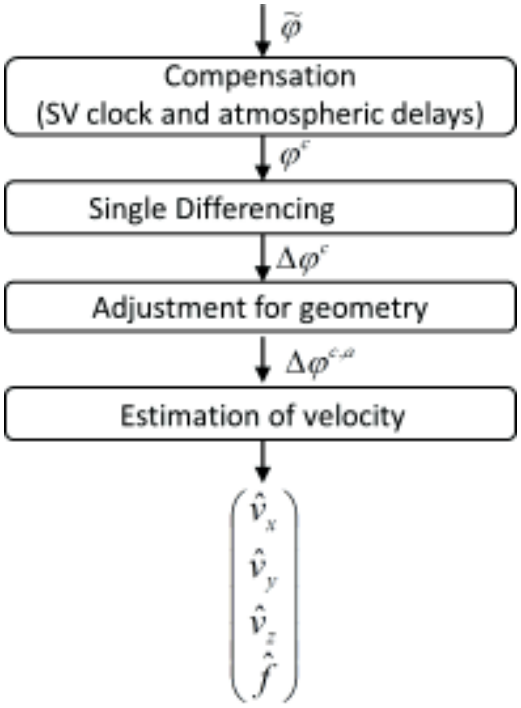

Figure 3.2: Precise velocity algorithm: block diagram

The carrier-phase single differences compensated and adjusted can be computed as:

$$\Delta\tilde{\varphi}_j{}^{c,a} = \Delta\tilde{\varphi}_j{}^{c} + \Delta f_{d_j} - \Delta g_j \tag{3.18}$$

where $\Delta\tilde{\varphi}_j{}^{c}$, Δf_{d_j} and Δg_j are expressed by (3.16), (3.17) and (3.18) respectively.
Then, using the corrected and adjusted SDs in (3.18), the LS solution can be computed as in the theoretical (3.7) substituting the true SDs with the measured ones, corrected and adjusted. The solution found using the corrected and adjusted measurements is therefore:

$$\mathbf{\Delta\hat{x}^{c,a}} = (\mathbf{H}^T \cdot \mathbf{H})\mathbf{H}^T\mathbf{\Delta\tilde{\varphi}^{c,a}} \tag{3.19}$$

Therefore, like in (3.8) the average precise velocities along the three directions, are given, in m/s, by:

$$\hat{v}_x^{c,a} = \frac{\Delta\hat{x}^{c,a}}{\Delta t}; \hat{v}_y^{c,a} = \frac{\Delta\hat{y}^{c,a}}{\Delta t}; \hat{v}_z^{c,a} = \frac{\Delta\hat{z}^{c,a}}{\Delta t}; \tag{3.20}$$

while, as in (3.9), the precise frequency is, in Hz:

$$\hat{f}^{c,a} = \frac{\Delta\hat{B}^{c,a}}{\mathrm{c}\Delta t} \tag{3.21}$$

where $\Delta t = t_2 - t_1$.

Chapter 4

Rubidium Oscillator Analysis

4.1 Motivations

As described in Section 1.2.2, if the receiver clock correction is known, three satellites are sufficient to obtain a navigation solution, increasing in this way the system availability and eventually also the integrity. This clock-free solution can be obtained for instance by incorporating a stable frequency reference, enough stable to be modelled, such as a Rubidium (Rb) oscillator.
A standard GNSS receiver includes, as time reference, a quartz oscillator. Such inexpensive reference is inaccurate, but considering the local clock bias as an unknown and solving for it using the GNSS measurements, it can be continually corrected in the position and time solution.
However, better performance can be reached, as explained in the following, using a higher stable oscillator. Different approaches can be used in this case. If an approach without the need of continuously include clock corrections is desired, very stable oscillators must be used which are very expensive, such as hydrogen masers. The alternative solution is the use of a less accurate reference, as a Rubidium clock, which errors are larger but can be modeled and corrected. In this way, the clock drift is estimated and compensated for, reaching good performance with moderate cost.
In this thesis a novel approach is proposed, to use an accurate frequency reference in order to improve not the position but the velocity solution accuracy. This means that what needs to be modeled is not the local clock phase error, but the changes in the clock error over time, i.e. the frequency error of the local oscillator. Therefore in this part of the thesis the interest is not on the local time

reference, the *clock*, but it is on the local frequency reference, the local *oscillator* and its frequency drift.
The objective is to investigate frequency-aided GPS velocity solution and the first step of the study is the analysis of the local frequency reference. In particular, in this thesis the focus is on Rubidium cell oscillators. This type of oscillators is very common in GNSS, in particular it is used as external reference for professional receivers, as well as onboard of satellites. Since the Rb and its characteristics are so crucial in this thesis, this chapter is entirely dedicated to the Rb characterization. Some theoretical background is recalled, then some laboratory and in-flight tests are described (see [33], (I)).

4.2 Frequency reference stability: analysis criterions

4.2.1 Frequency deviations

Before starting the analysis of the Rb frequency reference and the study of how to integrate it in the GNSS velocity solution, it is opportune to consider the general definitions and characterization methods of the frequency standards.
If an ideal oscillator is considered, the time interval is exactly in accord with the theoretical definition [34] and the output of the device is perfectly synchronous to the theoretical one.
In the real life, an oscillator frequency deviates from the nominal frequency and the reasons of this are essentially two, as explained in [34].
The first category of deviation is called *non-random*, since it is due to deterministic biases, such as a calibration error in the time interval or in the oscillator frequency, which accumulates over time and causes a drift.
The initial clock error instead is not an issue for the frequency solution, since it cancels out.
The second category of deviation is called *random*, because it is caused by all the different random noise processes in the oscillator. Noise affects oscillator the behavior, creating disturbances and modulations, both in amplitude and in phase. The amplitude noise is the cause of a continuous spectrum disturbance, anyway this effect is in general not significant in amplitude, also because the resonator

bandwidth is generally narrower. An issue is represented by the phase perturbations, which causes the frequency to have the same shape of the amplitude of the phase disturbances, but with narrower spectrum and much higher power. As a consequence, the frequency assumes a different value from the nominal one, even if it is distributed around it.
The shot noise caused by the photo detector in an atomic beam frequency standard, as the Rb cell (see Section 4.4), causes a time deviation which can be represented by a random walk noise process and represents one of the main frequency deviations.

4.2.2 Stability, reproducibility, accuracy

The performance of a frequency reference can be evaluated, as well as the time reference performance, in terms of stability, reproducibility and accuracy. According to Allan [34], the definition of these characteristics are as in the following.

Stability

Stability is related to precision, in fact even if stability depends only on the device that is considered, during a measurement also the non-ideality of the measurement device has to be taken into account. Precision is affected both by the frequency reference stability and by the measurement instrument deviations. In general the available information is precision, because it is the accessible information from a measurement process. It has to be taken into account that the precision of a frequency standard measurement can be limited by the instability of the worse between the standard frequency and the measurement instrument, and this is an important aspect in the analysis described in this thesis, especially in the measurement campaign conducted to characterize a frequency reference.

Reproducibility

Reproducibility, which will be investigated in depth in the next sections by means of laboratory and flight tests (see Section 4.7, 4.7.3), can be referred to a set of devices of same design, or to a single device.
If referred to a set of devices, reproducibility is defined as the degree of agreement across the devices, while if referred to a single device what is considered is the

intrinsic reproducibility, which means the degree of agreement across measurements taken in different times, in same conditions, of the same object. Several situations are considered and described in the following, for both the cases, to characterize a Rb frequency reference.
A measure of the reproducibility can be done taking some sets of measurements, for different devices or for the same device at different times. For each set of measurements the mean can be computed, and the standard deviation of those means can be considered a good measure of the reproducibility.

Accuracy

Accuracy is defined as *the degree of conformity of a measured and/or calculated value to some specified figure or definition* [34]. The main causes of inaccuracy are non-random deviations, however also random deviations affect the accuracy. Note that the accuracy can never be better than the precision. Also, accuracy can never be better than reproducibility, even if these two expressions are sometimes wrongly used in place of each others. Moreover, accuracy is an absolute measure, differently from reproducibility which is a relative one.

4.3 Stability measurement and sample variance

4.3.1 N-sample variance

In order to characterize the stability over time of an oscillator, the sample variance can be used. Different definitions are possible ([35]), which include different pros and cons. The sample variance gives an estimate of the theoretical time variance using a number N of samples. The N-sample variance can be defined as:

$$\sigma_x^2(N,T,\tau) = \frac{1}{N-1}\sum_{i=1}^{N}(\bar{x}_i - \frac{1}{N}\sum_{j=1}^{N}\bar{x}_j) \tag{4.1}$$

where N is the number of samples used in computing the variance, T is the time interval between each sample and τ is the time duration of each estimate x_i.
The con of this estimate is that it is tied up by the estimate time duration τ [36]. To overcome this issue, the two-sample variance is used (see Section 4.3.2).

4.3.2 Allan variance

The two-sample variance, known as Allan variance, was proposed by David Allan in 1966 [37]. Using the average of the variance with a defined value of samples $N = 2$ and adjacent samples (which means to take $T = \tau$), the estimate represents a unambiguous measure in the time-domain.
The Allan variance is defined as:

$$\sigma_x^2(\tau) = \frac{1}{N} < (\bar{x_2} - \bar{x_1})^2 > \tag{4.2}$$

(4.2) represents the theoretical definition of the Allan variance, that in real life can just be approximated with measurements, since the mean cannot be taken on an infinite interval as the ideal case would require. Due to the fact that a finite value m of samples has to be used for the estimate, a statistical uncertainty is present on the estimate.
Among the estimators of Allan deviation, a common one used is:

$$\sigma_x^2(\tau,m) = \frac{1}{2(m-1)} \sum_{i=1}^{m-1} (\bar{x}_{i+1} - \bar{x}_i)^2 \tag{4.3}$$

Allan variance in atomic oscillators

The Allan deviation $\sigma_x(\tau) = \sqrt{\sigma_x^2(\tau)}$, the square root of the Allan variance, is used to characterize phase stability of atomic oscillators, as in [34]. The Allan variance in this case can be defined theoretically as the following average over an infinite time:

$$\sigma_\phi^2(\tau) =< \frac{[\phi(t-2\tau) - 2\phi(t+\tau) + \phi(t)]^2}{2\tau\omega_0} > \tag{4.4}$$

where τ is the averaging time, $\phi(t)$ is the clock phase, ω_0 is the natural frequency of the source and t is the time. The Allan variance can be computed using a finite number m of samples as in (4.3). Table 4.1 shows the Allan variance of common oscillators, which are also the most common oscillators used in GNSS. In Figure 4.1, the Allan deviation of some oscillators is depicted. While quartz oscillator are widely used in mass market receivers, Rubidium and Cesium oscillators are used as clock reference on board of the satellites (e.g. GPS satellites ([38]), [39]), as well as the more performing but also more expensive masers (used on board of Galileo satellites [40]). Most professional receivers implement the option to use external clocks, such ar Cs or in particular Rb references. As it can be seen

	τ = 1 second	τ = 1 day	τ = 1 month
Quartz	10^{-12}	10^{-9}	10^{-8}
Rubidium	10^{-11}	$10^{-12}/10^{-13}$	$10^{-11}/10^{-12}$
Cesium Beam	$10^{-10}/11^{-12}$	$10^{-13}/10^{-14}$	$10^{-13}/10^{-14}$
Hydrogen Maser	10^{-13}	$10^{-14}/10^{-15}$	10^{-13}

Table 4.1: Allan variance of common oscillators

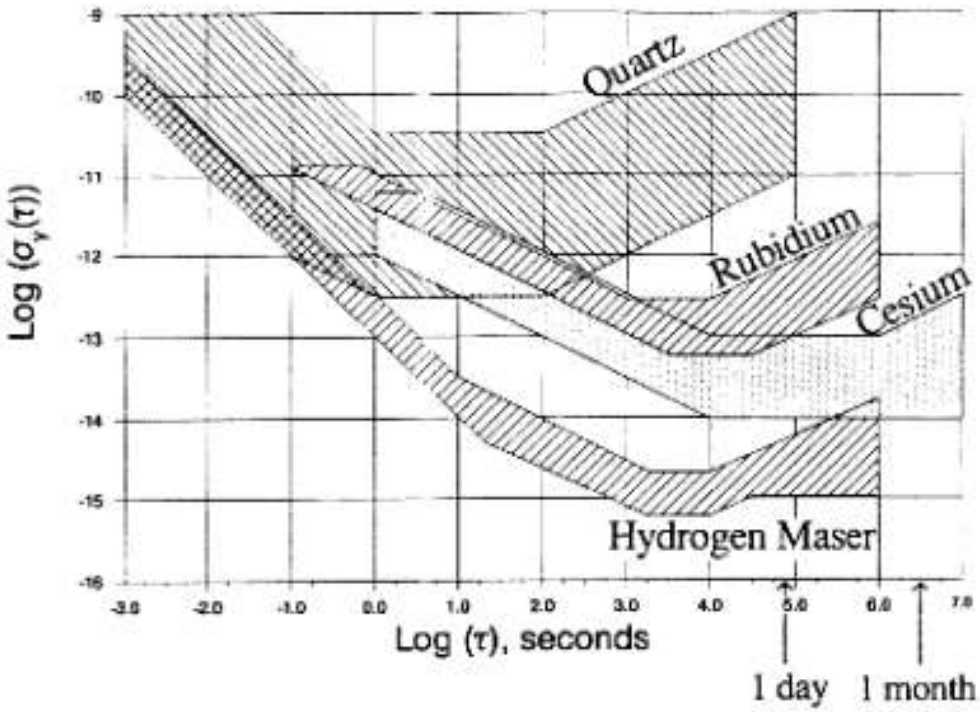

Figure 4.1: Rb Allan deviation. Source online: IEEE *Ultrasonics, Ferroelectrics, and Frequency Control Society, Frequency Control*, [8].

from Figure 4.1, the performance of the oscillators is not constant with the time interval τ. The analysis of Figure 4.1 is important since from it the proper way to make measurements of a certain oscillator can be evaluated. This means that since the Allan variance is higher for longer averaging time, it is not possible to average over time longer than a certain threshold, if the aim is to decrease the noise effects. Any average longer than that would lead to misleading results.
After this preliminary analysis, the study can be focused on the Rubidium oscillator, which is the oscillator considered in this thesis.

4.4 Structure of a Rb cell frequency reference

As firstly studied in the 1950s, hyperfine structure transition in alkali atoms as the Rubidium (Rb), can be optically detected. The first atomic frequency standards made with alkali vapour were built in the 1950s, then in the 1960s the first compact atomic frequency standards were developed, and their commercial versions became available. Quickly, the Rb cell oscillator revealed to be suitable for lots of applications, and when the GPS was developed, it became a fundamental time reference to be used and now its main uses include navigation and telecommunications systems.
Rb cell clock is a secondary atomic frequency standard, which means that its frequency has to be calibrated using a primary frequency standard as a reference. The structure of a Rb cell oscillator is shown in Figure 4.2. The Rb cell clock is

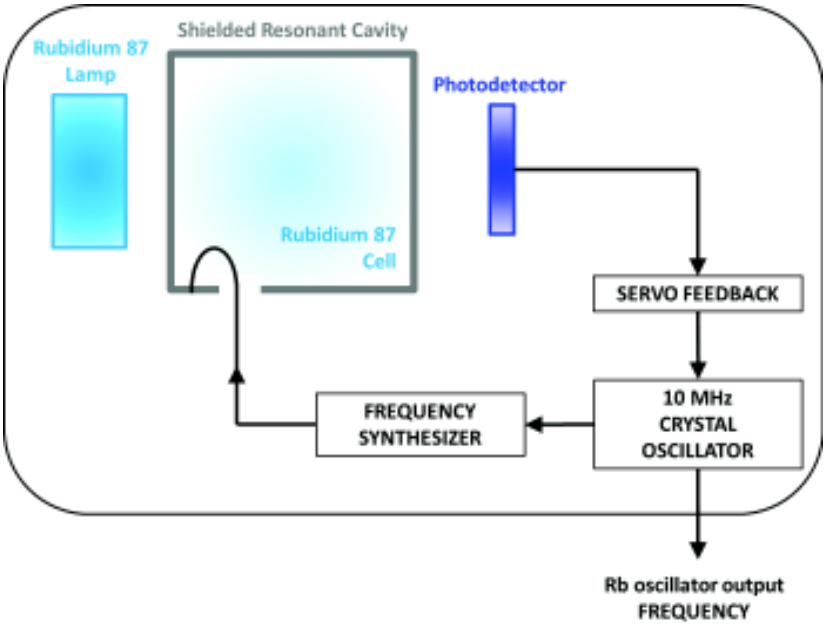

Figure 4.2: Rb structure

a passive standard and it is based on the double resonance method, because it uses both optical and microwave resonance. As depicted in Figure 4.2, a small lamp is included, which contains Rb 87 (or in some cases both the isotopes Rb 85 and Rb 87) mixed with a noble gas (Kripton). Through coupling with a generator at radio frequency (RF), from the lamp a discharge is produced and the generated light passes through an hyperfine filter containing Rb 85 and a noble gas (Argon). In the resonance cell, Rb 87 is present together with a low pressure

buffer gas, made of nitrogen and rare gases.
Some more compact Rb cells have the hyperfine filter built-in together with the resonant cavity, and nowadays, miniaturized Rb oscillators are available [41].
The resonant cavity results to be tuned to the hyperfine transition frequency of the Rb 87, which is 6,834,682,610.904324 Hz. A microwave (MW) signal is sent into the cell, and depending on this signal frequency, a different amount of light from lamp passes thought the Rb cell and reaches the photo detector (a silicon photo-electric detector). If the MW signal frequency changes from the Rb 87 hyperfine transition frequency, the photo detector output, which is a current, varies.
In this way a feedback loop is built. The MW signal frequency is controlled and kept stable using as feedback the photo-detector output, and the desired frequency can be then obtained from the controlled frequency by means of a frequency divider. In general, the controlled frequency is generated by a simple quartz oscillator at 10 MHz.
In this way, shown in Figure 4.2, a feedback is present that allows to calibrate continously the oscillator frequency.
More details about the Rb cell clock's operating principle can be found in literature, as e.g. in [42].

4.5 Rubidium cell metrological properties

The main interest of this thesis is on the Rb cell frequency instabilities, that, as in [42], can be divided in three main categories:

- medium term frequency instability
- shifts in the atomic transition frequency
- long term frequency instability and effect of environmental conditions

4.5.1 Medium term frequency instability

In a Rb cell the medium term frequency instabilities (sampling period τ smaller than $10^3 s$) are mainly due to the shot noise on the photodetector output and, as it can be seen from Figure 4.1, are within $5 \cdot 10^{-12}\tau^{-1/2}$ and $10^{-10}\tau^{-1/2}$. For

$10^3 < \tau < 10^4$, the instabilities do not vary with τ, due to systematic effects wich becomes dominant.

4.5.2 Shifts in the transition frequency

Shift in the transition frequency are the cause of biases in the frequency oscillator, that due to aging effects can also change in time.
Two effects cause these shifts in the Rb cell. The first one is associated to the optical pumping and can never be totally cancelled, causing effects of a few herts. The second one is due to the buffer gas, whose particles collides with the atoms of Rb, therefore this effect is proportional to the buffer gas pressure. Typically it is about 1 kHz, but since it is not possible to control so accurately the gas composition and pressure, this value changes from one device to another one.
The Rb cells produces as output a nominal frequency of 10 MHz, thanks to the calibration applied, as explained in Section 4.4. The Rb cell is a secondary standard because this calibration needs to be done with respect to a primary reference.
It has to be noted that these uncontrollable characteristics change in time also after the calibration and cause aging effects in the oscillator.

4.5.3 Long term frequency instability and effect of environmental conditions

Over the long term, which means for $\tau > 10^3$,$10^4 s$, the Rb cell oscillator presents frequency drifts of about $1 - 5 \cdot 10^{-11}$ Hz/month. The reasons of this effect are not perfectly known yet, even if the idea is that some chemical interaction occurs between the Rb and the glass containing it, causing changes in composition and pressure of the buffer gas.
Among all the atomic frequency standards the Rubidium is the most sensitive to the environmental changes, even if recently less sensitive Rb oscillators have been produced, thanks to technical arrangements such as the choice of the gases used and the working temperature. For example, this kind of low sensitivity Rb oscillator are mounted on board of the GPS satellites together with Cs standards. The temperature changes are the most significant ones (about 10^{-10} Hz/Â°C), however this problem can be mitigated by means of ovenized cells, keeping the

temperature constant.
The oscillators considered in this thesis are ovenized Rb, therefore they are not affected by changes in the environmental temperature. Other environmental changes will be considered in this chapter, as the laboratory and in-flight tests described in the following sections will show.

4.6 Rubidium oscillator models

In order to improve the GPS velocity solution by means of integration with the local oscillator measurement, the local oscillator must provide an accurate frequency reference. To do that, if a Rb cell is used, the oscillator frequency error must be mitigated. In order to do that a possible approach, considered here, is to use a model of the Rb frequency error.
Since the aim is to use the frequency estimate to improve the velocity solution, the interest is on dynamic cases. Therefore, the behavior of the Rb in non-static condition needs to be characterized. With this aim, some dynamics experiments are studied, but in this thesis, where a preliminary analysis is presented, only static models are considered.
The performance of Rubidium frequency standards is well-documented for stationary operation (e.g. [43]) and it looks repeatable and possible to be modeled. In dynamic environments that include changes in orientation, acceleration, temperature, pressure and magnetic field, the performance of frequency references for position, velocity and timing applications is not well characterized in the literature. Therefore, the first step is to study the feasibility of the work, i.e., to study the behavior of the Rubidium oscillator in non-static conditions, in order to eventually model its response, if possible.
Laboratory and flight test experiments were conduct at the Ohio University, as described in [33], and further analysis based on those experiments is performed as described in this thesis (I).

4.6.1 Standard model (phase and frequency)

The standard model used in literature (e.g. [44]) to represent the frequency and phase errors for oscillators such as Rb beam oscillators is shown in Figure 4.3, which considers the oscillator frequency being a first order random walk process,

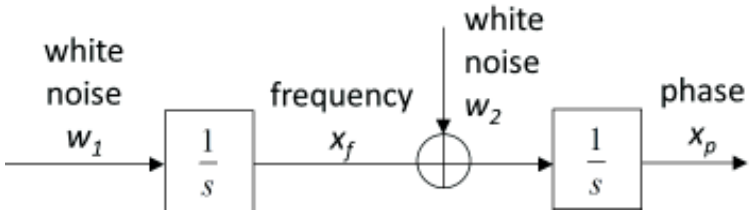

Figure 4.3: Standard model for Rubidium frequency and phase error

while the phase a second order random walk process.
In Figure 4.3 the random process w_1 is white noise and it is measured in units of s^{-2}. The frequency error x_f and the white noise w_2 are measured in s^{-1}, while the phase error x_p is in rad. The block $1/s$ in Figure 4.3 indicates an integrator in the continuous time (the variable s is the variable in the Laplace domain). Note that both the frequency and the phase errors x_f and x_p are here considered in the discrete time, but the errors are the result of continuous-time noise processes in the Rb cell, as explained in the following in the explanation of (4.7), and this is the reason why the integrators are represented in the continuous time.
Experimental values were found and are commonly used in literature [44] for the power spectral densities of the noise processes w_1 and w_2 in input to the system, which can be expressed, respectively, as

$$S_f \simeq \frac{h_0}{2} \tag{4.5}$$

and

$$S_g \simeq 2\pi h_{-2} \tag{4.6}$$

where the parameters h_0, h_2 depend on the oscillator type as indicated in Table 4.2

Oscillator type	h_0	h_{-2}
Crystal (temperature compensated)	$2 \cdot 10^{-19}$	$2 \cdot 10^{-20}$
Ovenized crystal	$8 \cdot 10^{-20}$	$4 \cdot 10^{-23}$
Rubidium	$2 \cdot 10^{-20}$	$4 \cdot 10^{-29}$

Table 4.2: h_0 and h_{-2} parameters for common oscillators.

Q matrix analysis

According to the model in Figure 4.3, defining the system noise as the vector:

$$\mathbf{w_{k,k+1}} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \tag{4.7}$$

where the subscript k,$k+1$ indicates that the system noise, which contributes to the system error in the discrete time, is the result of the integration of the continuous time noise (w_1 and w_2 respectively), over the time interval [k,$k+1$]. the Q matrix of the system is

$$Q_k = E\mathbf{w_{k,k+1}}\mathbf{w^T_{k,k+1}} \tag{4.8}$$

that becomes:

$$Q = \begin{bmatrix} \frac{S_p(\Delta t)^3}{3} & \frac{S_p(\Delta t)^2}{2} \\ \frac{S_p(\Delta t)^2}{2} & S_p(\Delta t) \end{bmatrix} \tag{4.9}$$

where Q is the system noise matrix, fundamental element in the design of a Kalman filter (see [44]).

4.6.2 Frequency models

In this thesis the focus is on the velocity solution, therefore on the oscillator frequency and not on the clock phase. The frequency model for the Rubidium can be considered, according to 4.3, as a first order random walk, but in order to properly model the oscillator behavior, also a frequency bias should be taken into account. In the following, different frequency models are considered. Different models allow to reach different degrees of accuracy in the modelling of a Rb oscillators. More accurate models, which take into account more aspects of the oscillator, are more accurate but also more difficult to be designed, because more parameters need to be evaluated.

Standard model for Rb frequency

The Rb frequency model which is considered to be the standard one, models the oscillator frequency error as random walk, as in 4.3. Therefore the frequency error is

$$x_{f_{k+1}} = x_{f_k} + w_{2k,k+1} \tag{4.10}$$

and the scheme is as in Figure 4.4. In this way, any eventual additive frequency bias is not modeled.

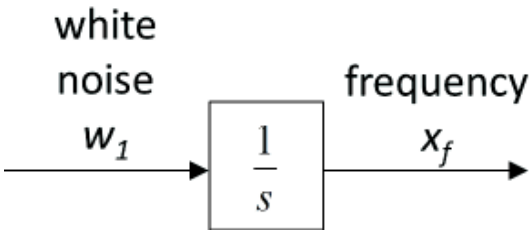

Figure 4.4: Standard model for Rubidium frequency error

Model including constant bias

To improve the frequency model above, a bias should be added to the random walk process to take into account the frequency shift effects in the Rb cell described in Section 4.5.2. In first approximation that bias could be considered constant. The resulting frequency is therefore:

$$x_{f_{k+1}} = x_{f_k} + w_{2k,k+1} + f_B \tag{4.11}$$

and the block scheme is like in Figure 4.5

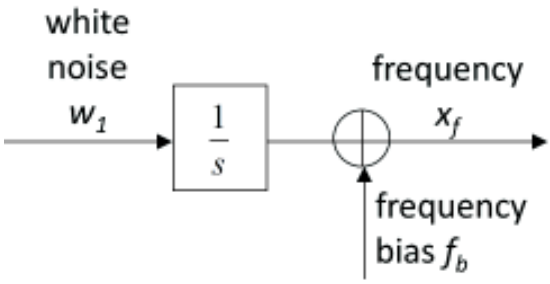

Figure 4.5: Model for Rubidium frequency error including constant bias

Model including Gauss-Markov process

What actually happens in a real device as a Rubidium oscillator is that the frequency bias is not constant over time, but it is a slowly changing process, due to the aging effects (see Section 4.5.2 and [42]). A way to model a slow process like that one is a first order Gauss Markov process. The frequency error results then to be modelled as

$$x_{f_{k+1}} = x_{f_k} + w_{2k,k+1} + f_{B_{GM(k+1)}} \tag{4.12}$$

therefore the block diagram rapresentation becomes as in Figure 4.6.

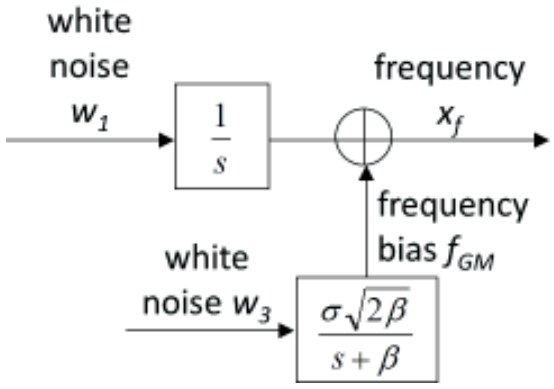

Figure 4.6: Model for Rubidium frequency error including slowly changing bias modelled as first order Gauss Markov process

4.7 Laboratory and in-flight tests

With the aim of modelling the behavior of Rubidium oscillators, the environment where they need to be used has to be considered, which is a dynamic environment. Some preliminary tests were conducted to evaluate the feasibility of the work [33], i.e. to analyze the Rb behavior under stress, to verify if it is possible to model it.

The aim is to characterize the Rb oscillator performance in non-static conditions. The oscillator accuracy defined in Section 4.2.2 has to be tested since the aim is to enhance the Rb frequency estimate to improve the accuracy on the velocity estimate. In particular it is fundamental to test the performance reproducibility (defined in Section 4.2.2), which is required in dynamics environment.

In fact, only if the Rb performance results repeatable, then it could be possible to design models to forecast the oscillator behavior.

Tests were conducted in flight, but before to do them, a preliminary analysis was done in controlled environment (laboratory test). The measurements in laboratory and flight analyzed in this thesis were conducted at the Ohio University, Avionics Center, by Dr. Frank van Graas, Dr Wouter Pelgrum and Samantha Craig. The first part of the analysis was presented in [33]. Further results and analysis were then included in (I).

Rubidium frequency performance is well-documented in stationary environment, but under dynamic conditions, including changes in orientation, acceleration, temperature, pressure and magnetic field, the performance of frequency references is not well characterized in the literature. The laboratory tests conducted

at the Ohio University aimed to investigate this topic, characterizing the performance of two Rubidium oscillator under varying gravity and magnetic field conditions, using as measurement device a Cs referenced frequency counter. Then, as described in Section 4.7.3, three Rubidium oscillators were tested also in flight.

Error Source	Frequency Error	Notes
g-sensitivity	$200/m/s^2$ to	Includes both
temperature	<200 over the	Residual is not
magnetic field	X, Y < 20/Gauss	-
pressure	0.1/mbar (20.1 for 0 to 6,000 ft)	-
retrace	50	after 24 hours

Table 4.3: Main factors affecting error sources in Rubidium oscillators.

From Table 4.3 it can be observed that, for a temperature-controlled environment, the g-sensitivity and magnetic field susceptibility dominate the frequency performance of a Rb oscillator in flight.

4.7.1 Laboratory test

Rubidium static performance

The phase and the frequency errors of a Rb oscillator can be modeled as in 4.3.

A simulation was run using the model in Figure 4.3. Theoretical performance according to the model was obtained, and Figure 4.7 shows the phase and the frequency error, expressed in units of m and in units of mm/s and pHz/Hz, respectively, as a function of time for a duration of 2.4 hours. Figure 4.7 shows that the phase error reaches a value of -15.6 m (or -52 ns), while the frequency error reaches -1.1 mm/s or -3.5×10^{-12}.

This simulation was done in order to compare the measured performance with the ideal performance according to the model in Figure 4.3: the ideal performance of the Rubidium oscillator will be compared with the measured performance in static conditions first, and then compared with the performance in dynamic conditions.

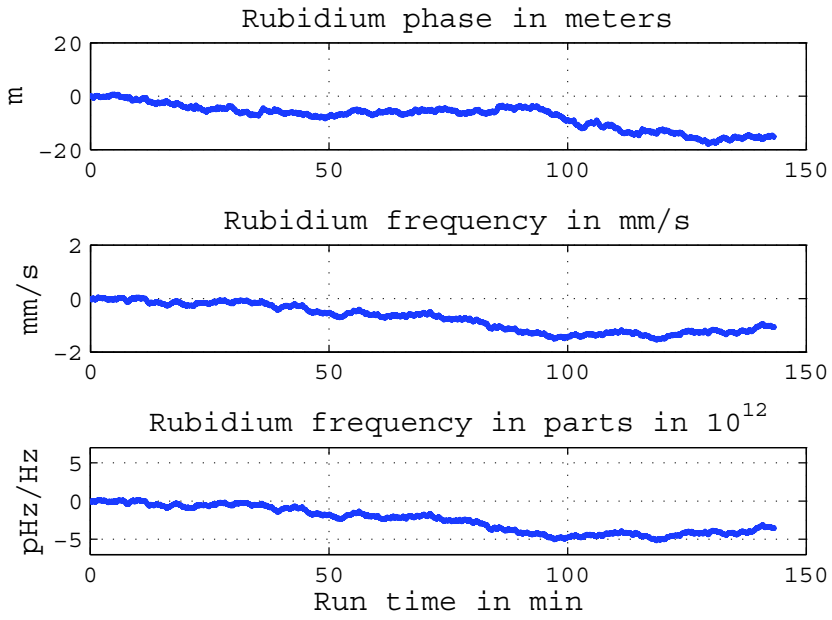

Figure 4.7: Rubidium phase and frequency error (Ideal simulation result)

In the measured real case, the precision is a relevant factor, which means that both the performance of the oscillator and of the measurement equipment are involved and have to be taken into account in the evaluation of the results. Considerations will be done about this issue.
A note needs to be done about the measurement units: the frequency error will be expressed is parts per 10^{12}, or pHz/Hz, which corresponds to a velocity error of 0.3 mm/s.

Laboratory setup

Given the Rb error source effects as in Table 4.3, the interest is on gravity and magnetic fields. Thus, to evaluate the Rb frequency error performance due to gravity and magnetic fields, a laboratory experiment was designed, ad hoc to measure the gravity error independent of magnetic fields, and vice versa. The experiment consists of three main steps. First, a Helmholtz coil is aligned with the local earth magnetic field vector and a current is applied, to cancel the earth magnetic field vector. Second, the Rb g-sensitivity is measured, orienting the Rb in six different orientations, along three orthogonal directions. Last, different magnetic fields are applied, for each of the three orthogonal directions, using

a single large magnetic coil. The tests were done on two Rb oscillators from different manufacturers. It is important to notice that neither oscillator was optimized for a dynamic environment, such as the aircraft environment is, in fact the goal is exactly to characterize the major error sources as a function of aircraft orientation and dynamics for a standard ovenized Rb oscillator not specifically designed for flight.

The block diagram of the laboratory instrumentation is shown in Figure 4.8.

A 3-axis magnetometer is used to verify that the earth magnetic field has been cancelled at the Rb test location at the centers of the Helmholtz and single magnetic coils.

The Cesium reference has a frequency stability on the order of 10^{-13} and is approximately an order of magnitude better than the Rb oscillator under test, which have frequency stabilities on the order of 10^{-12}, while the frequency counter has a measurement resolution of 5×10^{-12}, which represents the limiting factor

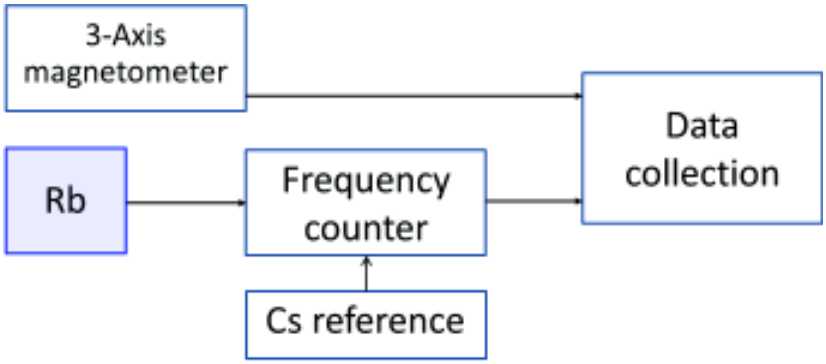

Figure 4.8: Laboratory instrumentation block diagram

in the measurement accuracy.

The instrument configuration used for the laboratory test is shown in Figure 4.9, where the Helmholtz coil aims to cancel the earth magnetic field and the single large magnetic coil to apply magnetic fields in two directions. As specified in [33], both coils were verified against the theory and were found to agree to within 0.55% and 1.94% in the areas of interest for the Helmholtz coil and the single large coil, respectively. In order to measure the residual field inside the Helmholtz coil, the biases and scale factors of a 3-axis magnetometer were calibrated outside the building based on the known value of the earth magnetic field obtained from the World Magnetic Model (WMM) 2010 [45]. The approximate value of the magnitude of the earth magnetic field at the Ohio University test location is 48,687 nT or 0.48687 gauss with an inclination angle of approx. 67°. The

inclination angle is reflected in the orientation of the Helmholtz coil in Figure 4.9.

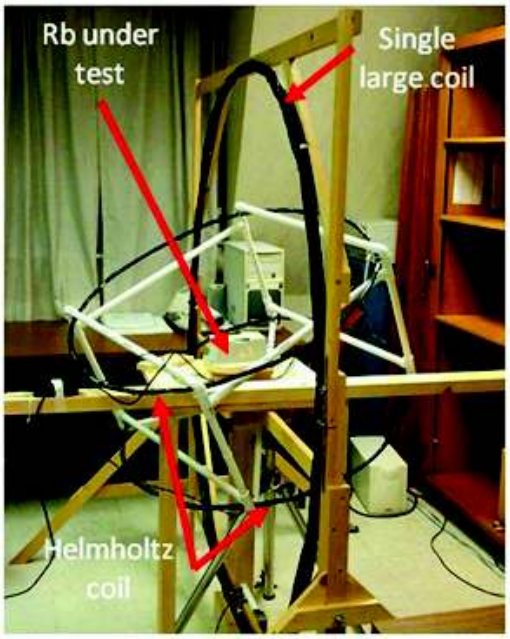

Figure 4.9: Laboratory setup including Helmholtz coil and magnetic coil.

Gravity field sensitivity

Two Rubidium oscillators were evaluated in terms of sensitivity to the gravity field. Note that the Rb oscillator is considered as a *black box*, since the manufacturer does not provide details on the internal orientation of the constitutive components of the oscillator, with respect to the external case.
In this thesis some results of the test conducted at the Ohio University are described [33], (I).
Rb-A was rotated around its Z-axis with the earth magnetic field cancelled (within 5% of its value), reducing the magnetic field-induced error to below ±2.5 pHz/Hz (see Table 1). Multiple 90° rotations in the counter clock wise (CCW) and clock wise (CW) directions were done every 5 minutes around the Z-axis, completing a full 360° after 25 minutes. This was followed by the CW rotation that lasts 45 minutes, after which the CCW rotation is repeated.

The result gives a mean frequency error of -2.13 $\times 10^{-10}$, while the standard deviation is 1.14$\times 10^{-11}$. The mean error is within the specification for Rb-A, while

the standard deviation is primarily induced by the frequency counter output resolution of 1 part in 10^{11}. From this experiment, it was seen that a rotation about the Z-axis induces a change of less than 10 pHz/Hz in the frequency stability for Rb-A.

While the rotation around the Z-axis induces a very small error, bigger error is visible with rotations around the other axes. In Figure 4.10 the upper plot shows the frequency error for Rb-A for a rotation around the X-axis. From Figure 4.10, g-sensitivity errors of up to 100 pHz/Hz are observed in addition to much larger transients immediately following each 90° turn. The rest time after each 90° turn is 10 minutes to allow the frequency error to settle. It is also noted that the settling time of the frequency error is a function of the orientation before and after each turn.

For a rotation around the Y-Axis as shown in the lower plot in Figure 4.10, the frequency errors are larger than for rotations on the other axes, being the error up to 160 pHz/Hz. To further investigate the transient frequency error following each 90° turn, as shown in Figure 4.11 the turn angle rate was varied between 30, 3, and 1.5 °/s. For smaller turn angle rates, the transient response lasts longer, but the peak amplitude is reduced as shown in Figure 4.11. Note that the transient turn response if not identical between different 90° turns.
For comparison purposes, a second Rubidium oscillator from a different manufacturer was tested too. The resulting frequency g-sensitive errors for Rb-B do not exceed 50 pHz/Hz. Note that for Rb-B no transient error behavior was observed.
For both oscillators, the g-sensitivity error is repeatable and could be corrected based on knowledge of the clock orientation with respect to to the gravity vector.

Electromagnetic field sensitivity

Laboratory tests were conducted to characterize the Rb behavior also under different magnetic field, which constitutes one of the main factors affecting the Rb performance (see Table 4.3). The focus here is on the response to the variations. The tests were done varying the magnetic field using the single large coil, while the Helmholtz coil cancelled the earth magnetic field.
Figure 4.12 shows the response of Rb-A to magnetic fields varying between 0, ±0.5, and ±1 gauss. For Rb-A, the most sensitive axis for magnetic field-induced

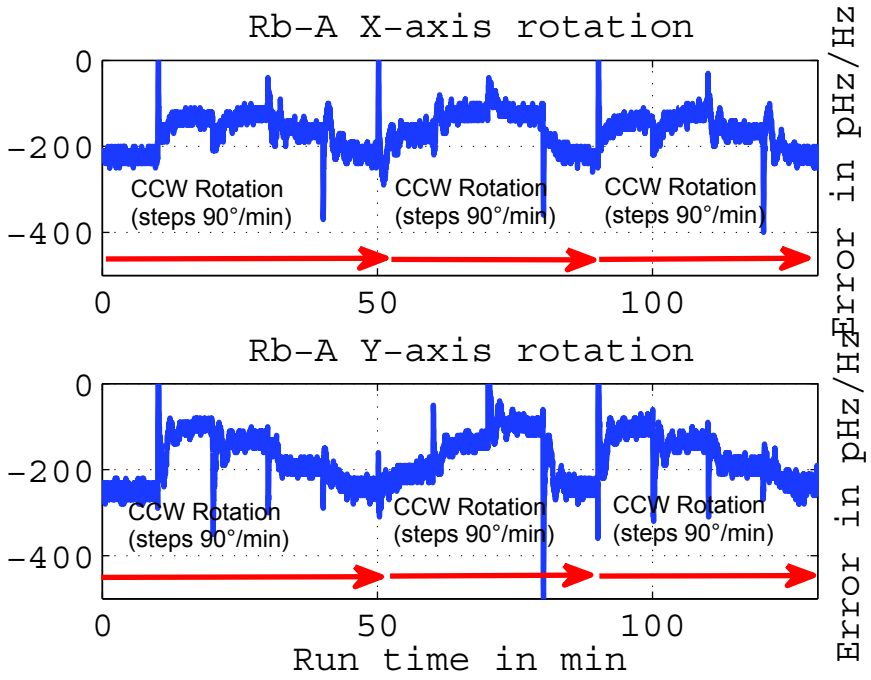

Figure 4.10: Rb-A: G-Sensitivity for Rotation around X-Axis and around Y-Axis, with Earth Magnetic Field Cancelled

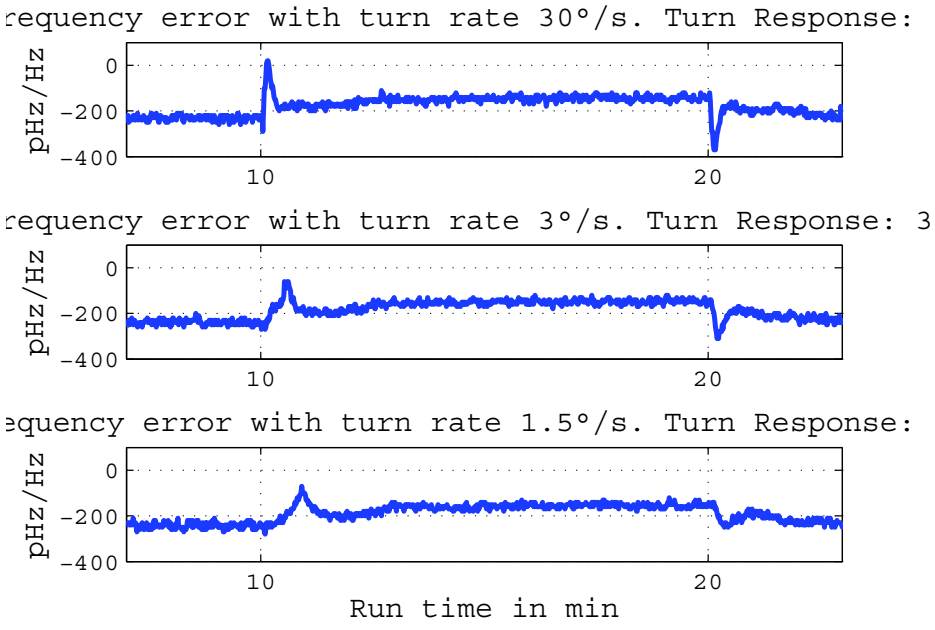

Figure 4.11: Rb-A: Freq. Error in Response to 90° turns around X-Axis at Different Angular Rates, Earth Magnetic Field Cancelled.

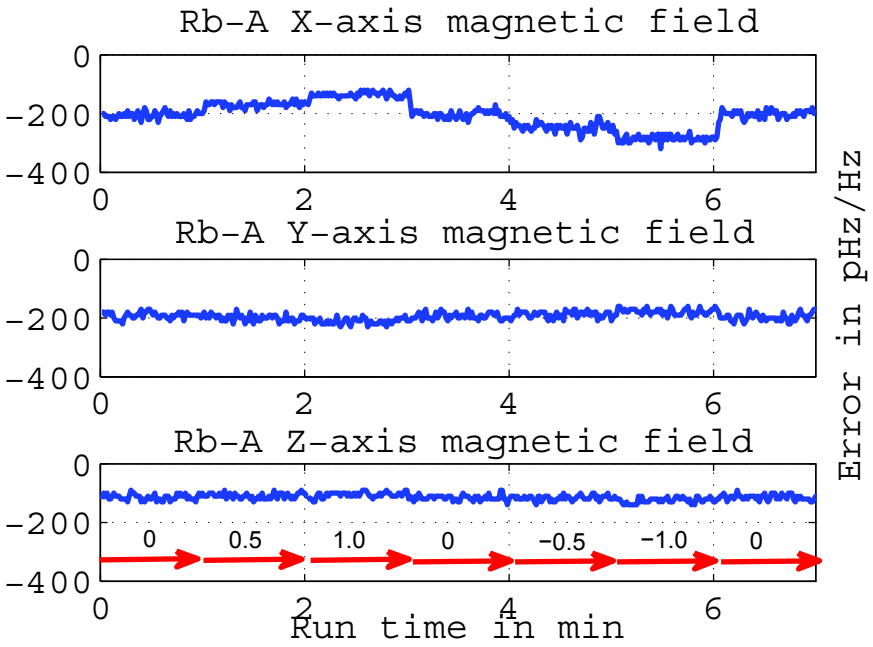

Figure 4.12: Rb-A: Magnetic Field Sensitivity in the X, Y and Z Direction.

frequency error is the X-axis, as it can be seen in the upper plot in Figure 4.12, with variations of up to 50 pHz/Hz for 0.5 gauss. The error in the Y direction, in the middle plot of Figure 4.12, is less than 10 pHz/Hz for 0.5 gauss and even smaller in the Z direction, in the lower plot, Figure 4.12.
The test on the magnetic field sensitivity was also applied to Rb-B along its most sensitive axis, showing sensitivities of less than 15 pHz/Hz for 0.5 gauss.

Comments on the laboratory tests

From the laboratory tests it can be said, in summary, that both the Rb oscillators tested have repeatable responses for both g-sensitive and magnetic field susceptibility. Worst case g-sensitive frequency errors due to rotation in the gravity field for Rb-A reach 160 pHz/Hz, while Rb-B frequency errors are less than 50 pHz/Hz. Worst case magnetic field induced frequency errors due to the earth magnetic field for Rb-A reach 50 pHz/Hz, while Rb-B frequency errors are less than 15 pHz/Hz.
Since the response to the gravity and magnetic fields for both oscillators is repeatable in the time the Rb behavior under the effects of those fields can be modeled.

4.7.2 Comparision between Cesium and GPS measurements

While in laboratory Rubidium frequency measurement could be conducted using a Cesium referenced frequency counter, in flight the measurement is done by means of GPS solution. In particular, the GPS precise frequency algorithm is used as described in Chapter 3.
To verify the implementation of the GPS precise frequency algorithm, a static test was performed. The Rb oscillator output was divided and connected to the external frequency input of a NovAtel OEM-V GPS receiver as well as a frequency counter. The frequency counter was referenced to the Cs oscillator as shown in Figure 4.8. The GPS receiver used an antenna on the roof of Ohio University Stocker Center, which represents a more severe multipath environment than what would be experienced for an aircraft in flight. Figure 4.13 shows the GPS precise frequency estimate for 1 hour of data, filtered by a 0.1-Hz bandwidth (BW) low-pass filter (upper plot) and by a 0.01-Hz BW low-pass filter (lower plot).
The filtering is introduced to remove the high frequency components due to GPS noise and multipath errors. The mean frequency error is also removed to enable a direct comparison with the Rb model shown in Figure 4.3. From the model, the frequency error change after one hour should be on the order of 2-3 pHz/Hz. The GPS estimates vary between -10 and +10 pHz/Hz, mostly due to GPS noise and multipath errors.

Figure 4.14 shows the Rb frequency errors measured by the frequency counter using the Cs oscillator as the reference. In the upper plot of Figure 4.14, the frequency is filtered by a 0.1-Hz BW low-pass filter, while the lower plot shows the 0.01 Hz filtered frequency. In both cases, the mean frequency error has been removed. The higher frequency errors in Figure 4.14 are dominated by the counter measurement resolution of 5 pHz/Hz. After filtering, the counter estimates vary between -5 and +5 pHz/Hz and support the model that predicts a frequency error change on the order of 2-3 pHz/Hz after one hour of operation.

In Figure 4.15 the estimates provided by the GPS LMS solution and by the Cesium-referenced frequency counter are compared. From Figure 4.15 considerations can be drawn about the Rb frequency bias. The GPS and the Cs-referenced measurements in fact agree in estimating the Rb frequency bias, with a error of less than 3.5 pHz/Hz. In fact the mean computed using the filtered LMS solution

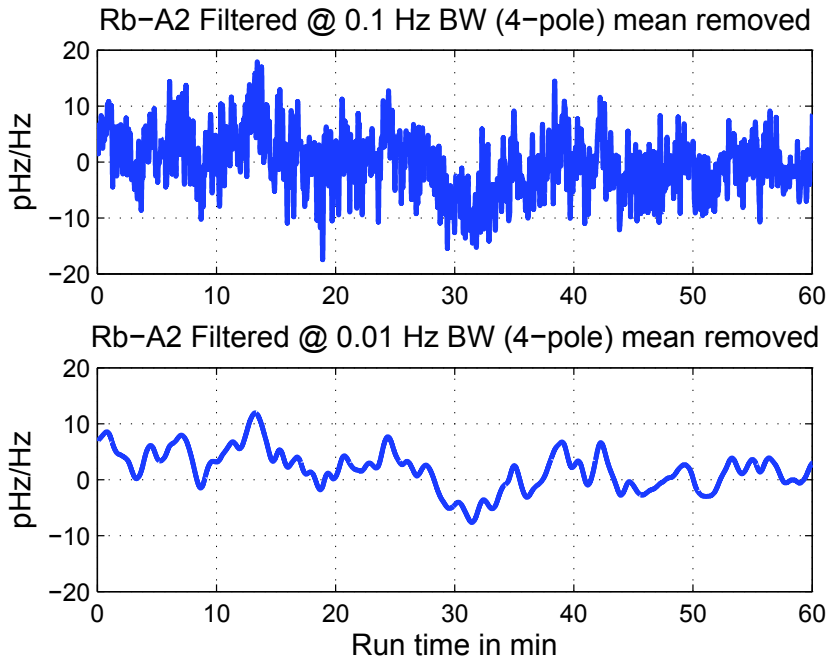

Figure 4.13: Rb Frequency Error in Static Condition: GPS Precise Frequency Estimate, mean removed.

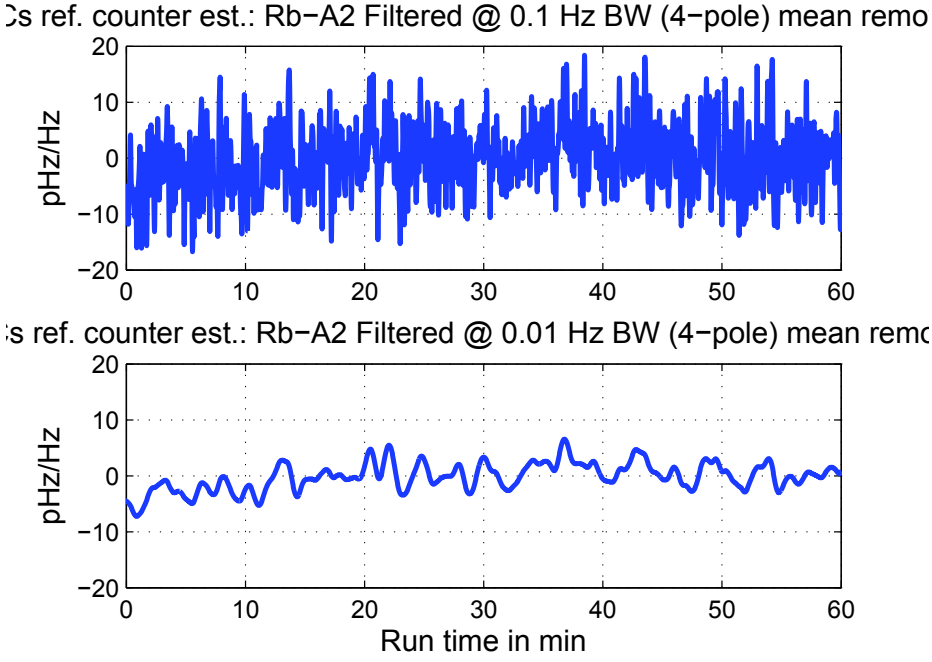

Figure 4.14: Rb Frequency Error in Static Condition: Cesium Referenced Counter Estimate, mean removed.

is 11.5 pHz/Hz while the frequency counter filtered mean value is 15 pHz/Hz. The difference in the frequency bias estimates is acceptable because of the counter resolution and GPS noise and multipath errors. To further investigate the GPS frequency error, Figure 4.16 shows the GPS frequency estimate and the estimated change in the vertical position, Δz, in the top graph. The lower graph in 4.16 shows the correlation coefficient between the frequency, mean removed, and the vertical delta position solutions. Since the receiver is stationary, the true vertical delta position is zero. Since the correlation is large, especially during the first 35 minutes, the vertical delta position errors are a good statistical approximation of the frequency errors, which confirms that the GPS LMS frequency solution errors are dominated by noise and multipath.

From the static experiment, it is found that the multipath errors dominate the precise GPS frequency estimates, but are generally within 10 pHz/Hz after filtering the frequency errors with a 0.01 Hz BW low-pass filter. In-flight performance is anticipated to be better due to the lower multipath environment.

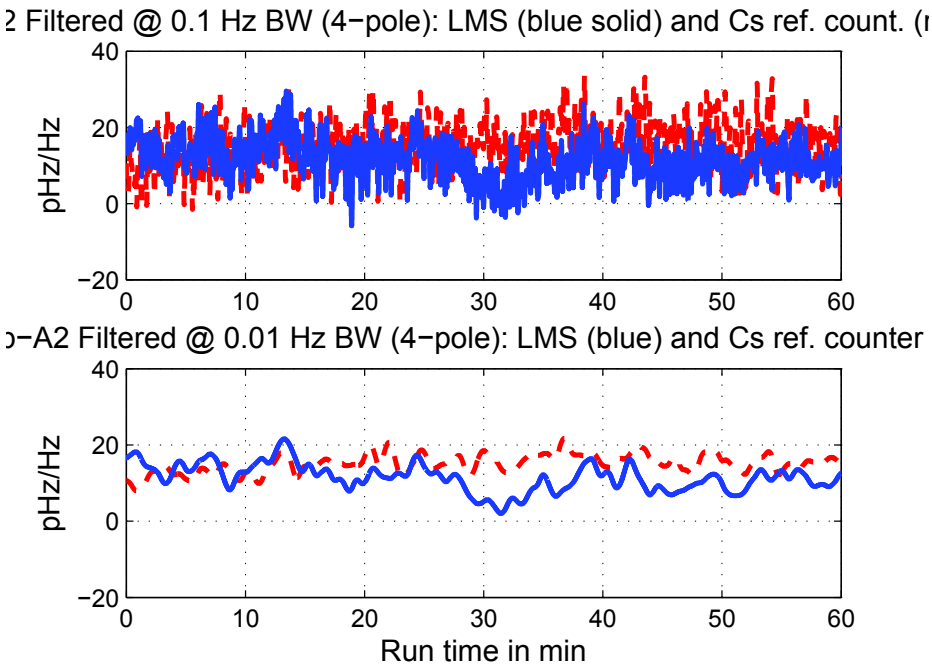

Figure 4.15: Rb Frequency Error in Static Condition: comparison between the bias estimate of the LMS solution (blue) and the Cesium Referenced Counter.

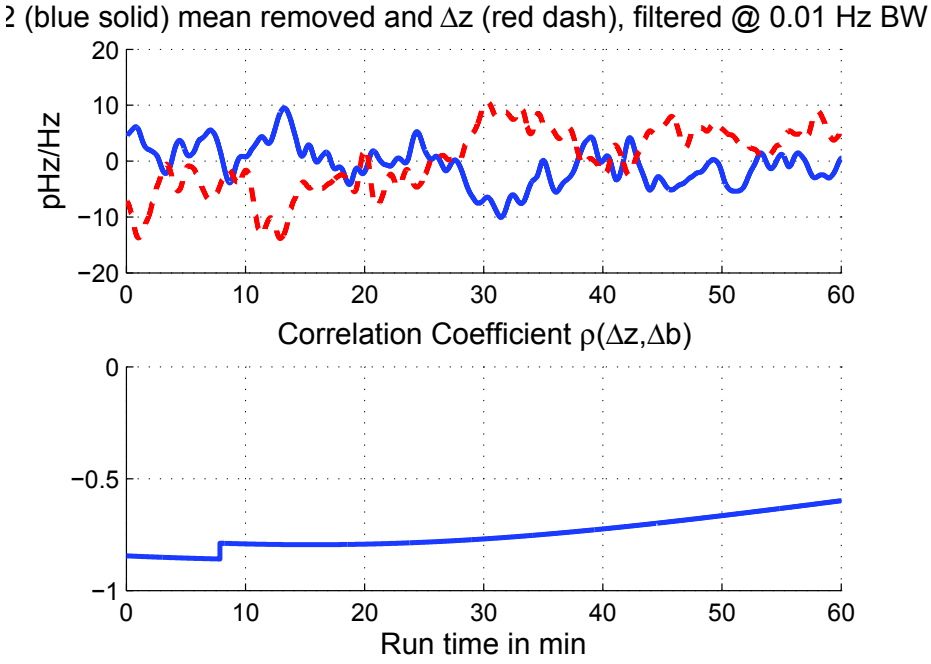

Figure 4.16: LMS solution: Comparison between Rb Frequency and Delta z estimates (upper plot) and Correlation coefficient in the time (lower plot)

4.7.3 In-flight test

After the laboratory test, also an in-flight test was performed. Three Rb oscillators of type Rb-A, which in the following will be indicates as Rb-A1, Rb-A2 and Rb-A3 respectively, were tested in flight, using as measurement device a GPS receiver. The scope was characterize the error to profile frequency stability performance against the flight dynamics, based on a navigation-grade inertial navigation system.
For the flight test the Ohio University DC-3 research aircraft (N7AP) shown in Figure 4.17 was used.
The three Rb oscillators were the external frequency reference inputs to three separate NovAtel OEM-V GPS receivers installed in two locations.
Rb-A1 was installed inside Ohio University Rubidium Inertial GPS Receiver (RIGR), on top of a Honeywell HG-1150 navigation-grade inertial navigation system (INS) mid-ship on the starboard side (see Figure 4.18, upper part).
Rb-A2 and Rb-A3 were mounted on a second RIGR, installed on a plate directly mounted to the seat rails in a starboard aft location (see Figure 4.18, lower part).
Rb-A2 was connected to the second RIGR, while a separate NovAtel OEM-V was used to interface with Rb-A3. The GPS pseudorange and carrier phase measurements were collected at a rate of 20 per second in the MatlabÂ® environment.
The flight test experiment was conducted on June 11, 2010 (17:56-20:28 local time) and consisted of maneuvers in the vicinity of Ohio University, Athens, Ohio. Note that the aircraft was stationary during the first five minutes of the experiment, followed by a five-minute taxi and take-off.
The flight dynamics, as measured by the HG-1150 INS, are shown in Figure 4.19 and include roll angles up to 48Â° and pitch angles between -10Â° and 10Â°.

Figure 4.17: Ohio University DC-3.

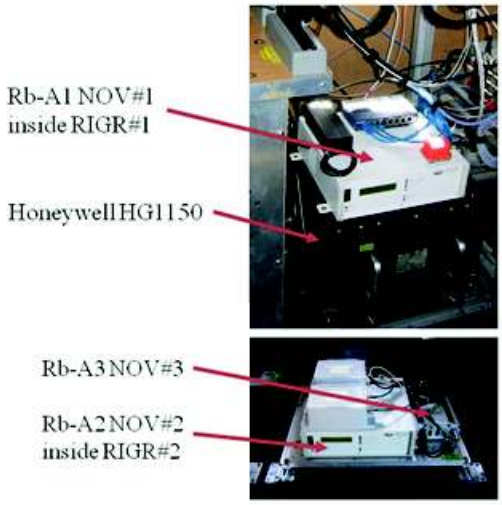

Figure 4.18: Rb-A1, Rb-A2 and Rb-A3 Aircraft Installation.

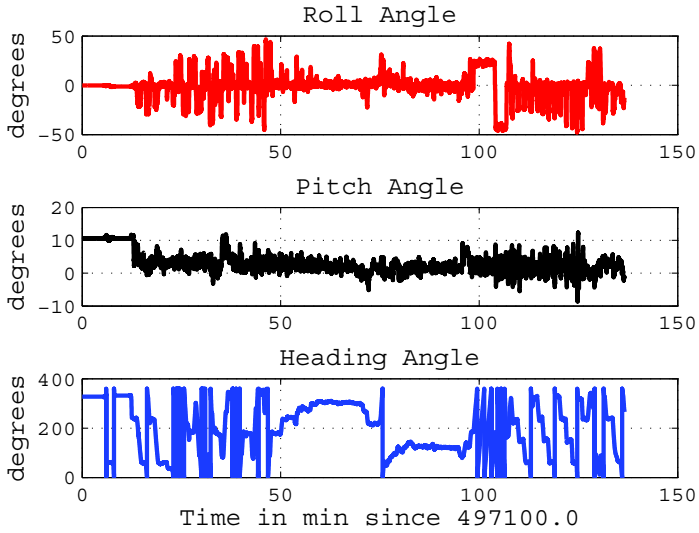

Figure 4.19: DC-3 Flight Dynamics

Figure 4.20 shows the frequency errors derived from the precise GPS frequency solution using the NovAtel OEM-V GPS receiver measurements. The three subplots in Figure 4.20 are referred to Rb-A1, after filtering with a low-pass filter with bandwidths of 1 Hz, 0.1 Hz and 0.01 Hz, respectively.

The mean frequency errors is removed from the middle and lower subplots. The middle sections from 50 to 100 minutes of Figure 4.20 represents straight and level flight towards Dayton, Ohio, followed by a turn and straight and level flight back to Athens, Ohio. Figure 4.21 shows the comparison of the three Rb

oscillators for a filter bandwidth of 0.01 Hz. Although all three oscillators were mounted in the same orientation, Rb-A1 shows a frequency error response that is out-of-phase with the other two Rb oscillators. This difference is likely related to the magnetic field differences due to nearby equipment and/or differences in magnetic shielding of the oscillators. All three Rb oscillators showed a frequency error that is correlated with the change in the direction of the magnetic field. The figures also show the effects of the sustained turns in the flight dynamics that result in oscillatory frequency errors. From Figure 4.21, it is found that frequency errors vary between -50 and +50 pHz/Hz during dynamics and between -30 and +30 pHz/Hz during straight and level flight. These frequency variations are one to two orders of magnitude larger than those observed during the laboratory experiment. For Rb-A1 in Figure 4.20, the standard deviation of the frequency error for the 1-Hz bandwidth filter is 26.6 pHz/Hz during the straight and level section of the flight test. For Rb-A2 and Rb-A3, the standard deviations of the frequency error for the 1-Hz BW filter are 42.8 and 31.3 pHz/Hz, respectively. The level of frequency error noise for Rb-A1 is smaller compared to Rb-A2 and Rb-A3, which is most likely due to the mounting of Rb-A1 on top of the HG-1150, which provides some level of vibration isolation through its avionics tray installation.

4.8 Conclusions about Rb tests

Models of the frequency error in Rb oscillators exist (as example [46], [47] or [48]) that reveal to be good models to be applied to benign environments, but care must be taken when these models are used in the flight test environment. In fact the g-sensitivity and magnetic field susceptibility dominate the nominal frequency errors.

In flight, Rb oscillator frequency errors degraded by one to two orders of magnitude compared to the laboratory environment. Different oscillators from the same manufacturer exhibited similar, but not identical frequency errors (three oscillators of same manufacturer were tested in flight).

Laboratory tests on the g-sensitivity and magnetic field susceptibility show that the errors due to these effects are repeatable. In quantitative terms, worst case g-sensitive frequency errors due to rotation in the gravity field for Rb-A reach 160 pHz/Hz, while Rb-B frequency errors are below 50 pHz/Hz. Worst case

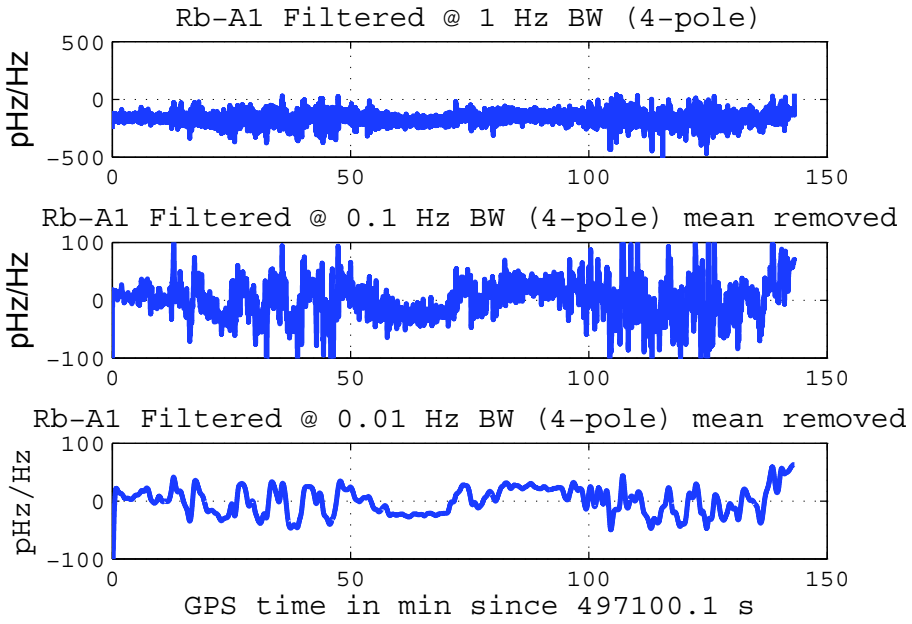

Figure 4.20: Rb A1 In-Flight Frequency Errors.

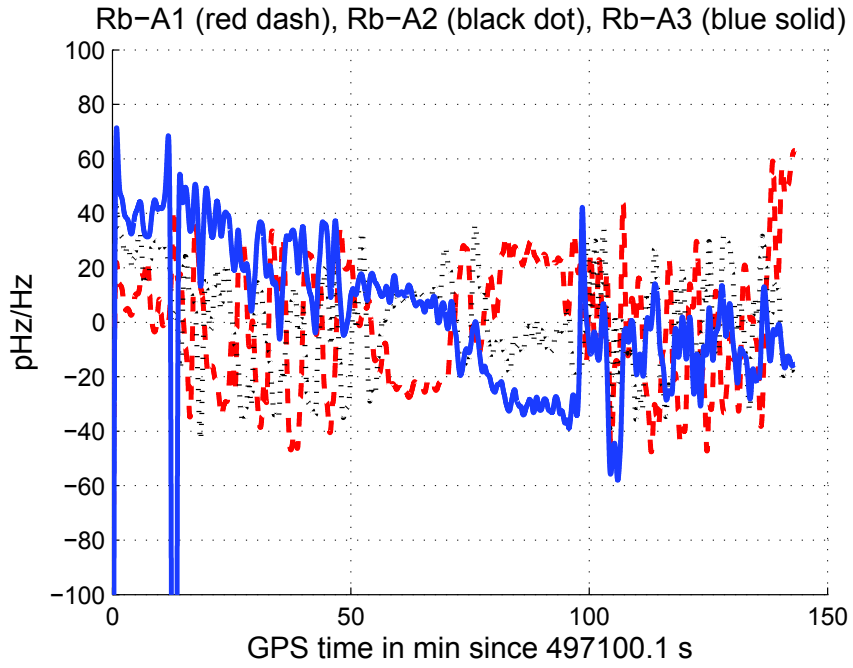

Figure 4.21: Comparison of Three Rb Oscillators In-Flight.

magnetic field induced frequency errors due to the earth magnetic field for Rb-A reach 50 pHz/Hz, while Rb-B frequency errors are less than 15 pHz/Hz.
Since the response to gravity and magnetic fields, for both the oscillators tested in laboratory, is repeatable, models of those response could be found, on the basis of more laboratory and in-flight experiments.

Chapter 5

Frequency Aided Precise Velocity

5.1 Objective

In order to obtain an accurate frequency reference, the first requirement is to have a stable oscillator, where the remaining drifts can be at least modeled. In Chapter 4, the Rubidium oscillator reference is analyzed, with the aim of evaluating the possibility and eventually the method to integrate Rb frequency measurements with the GPS velocity solution. In this chapter a novel approach is investigated, and the possible performance achievable integrating accurate frequency estimates with GPS velocity solution is evaluated.

In literature, several examples are available of combined use of atomic clock and GPS measurements, as instance in [49], [50], [51], [52], [53], with the aim to discipline the oscillator using GPS measurements, or to improve GPS availability and accuracy of position solution thanks to the clock phase aiding. In literature there are not presents applications involving the oscillator frequency to improve the velocity performance. Studies have been done in order to improve the velocity measurement, but the methods implied did not make use of the oscillator estimate. Here a novel approach is presented.

In this chapter, some results are shown, which represent preliminary results. The novel field of the frequency-aided velocity still needs to be investigated and many steps are on the to-do-list for the future.

5.2 Frequency aiding method

The basic idea is to improve the GPS local frequency estimate by means of a filter that takes into account the local oscillator characteristics. The improved frequency estimate can be then used to correct the carrier phase single differences (compensated and adjusted as described in 3.3), which are then used to compute the velocity solution using a frequency-free method. In this way the unknowns result to be only the three components of the user velocity.
Thanks to the frequency aiding, the velocity accuracy can be improved, in particular on the vertical direction, given the correlation existing between the frequency solution and the velocity, which is in general much higher for the vertical velocity component (as explained in Chapter 2.
The general structure of the frequency aided precise velocity algorithm considered in this thesis is represented in Figure 5.1. A standard precise velocity algorithm (see Chapter 3) uses a LMS method to give an estimate of the change in the position $\Delta\hat{x}$, $\Delta\hat{y}$, $\Delta\hat{z}$ and the change in the local clock bias, i.e. the local oscillator frequency error $\Delta\hat{B}$. Note that in Figure 5.1 the carrier-phase single differences $\Delta\varphi$ are the SD corrected and adjusted according to [2].
The frequency estimate $\Delta\hat{B}$ is then filtered, using a proper filter that takes into account the characteristics of the oscillator, and the filtered frequency is used to correct the carrier-phase SDs(corrected and adjusted). Then the LMS method is applied to the frequency-error-free SD, solving this time only for three unknown, the user velocities along three directions. The output is the frequency-aided velocity, which can be obtained from $\Delta\hat{x}_{FA}$,$\Delta\hat{y}_{FA}$,$\Delta\hat{z}_{FA}$ dividing by the time-interval Δt on which the SD are computed. Note that the estimated frequency is the mean frequency over the time interval Δt
The crucial element of the architecture in Figure 5.1 is constituted by the filter, which has to be properly designed taking into account the type of the local oscillator used as frequency reference in the GPS receiver. The challange of the filter design is strictly connected to the problem of properly modelling the oscillator errors. Mismodeled characteristics constitute contributions to the residual error bias on the frequency and velocity solution.

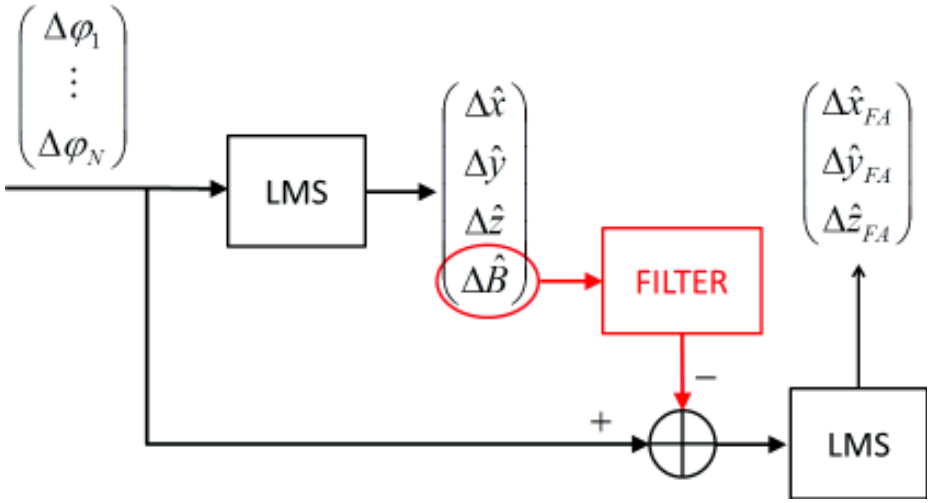

Figure 5.1: Frequency aided precise velocity algorithm: general structure.

5.3 Simulation results

In the following, some simulations will be shown. In all the cases the geometry condition is kept constant, as well as the carrier phase noise. The simulation conditions are summarized in Table 5.1.

σ_φ	0.04 m
Num Satellites	4
XDOP	0.84
YDOP	1.14
ZDOP	1.61
ZDOP/XDOP	1.92
ZDOP/YDOP	1.41
XDOP/YDOP	0.74

Table 5.1: Simulation conditions.

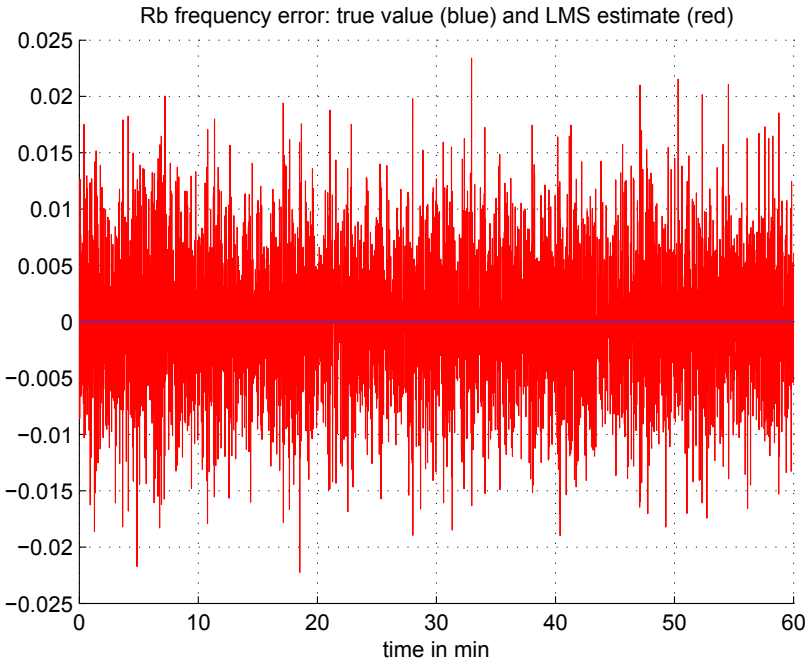

Figure 5.2: LMS estimate, when the frequency reference is perfect.

5.3.1 Ideal case: frequency error perfectly known (zero frequency error)

A simulation to test the performance can be done placing the hypothesis to perfectly know the local frequency error, which is equivalent to have a perfect reference, and be aware of that. Note that if the local frequency reference is perfect but the user does not know that, the local frequency is considered as unknown and the GPS solution is used to estimate it. As a result the frequency and the velocity solution are affected by an uncertainty which depends on the carrier phase noise, as well as on the DOP, without taking many advantages from the goodness of the local oscillator.

In the ideal case of perfect frequency correction, the filter is simply substituted by a block that introduce the known true value of frequency. In Figure 5.2 the LMS estimate is shown, in the case of null frequency error. Figures 5.3 and Figure 5.4 show the velocity estimate using a standard LMS, where the local frequency is considered unknown, and a frequency corrected LMS, respectively. In Figure 5.5 a comparison is done between the vertical velocity computed with the LMS method and with the frequency aided method, while Figure 5.5 shows the ideal improvement that could be reached in the vertical velocity by totally correcting

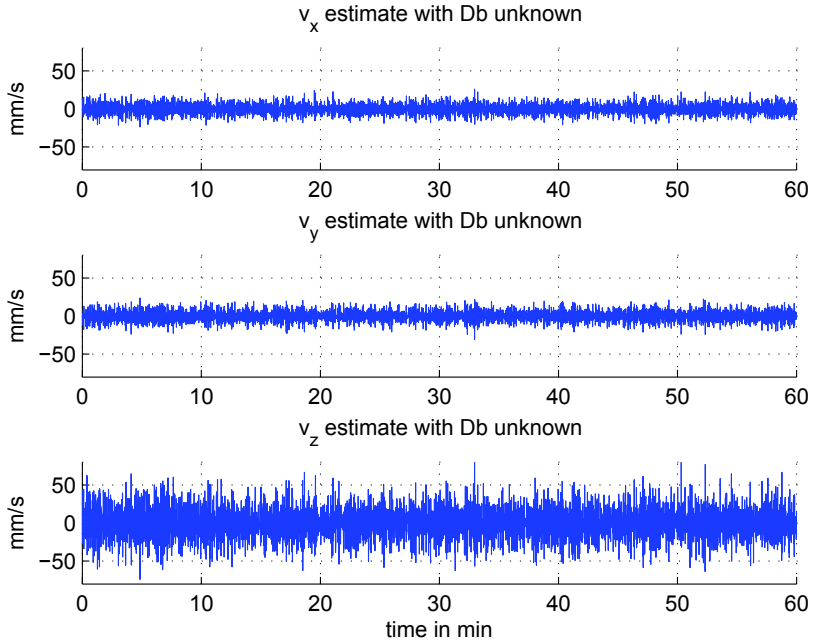

Figure 5.3: LMS velocity solution (perfect frequency)

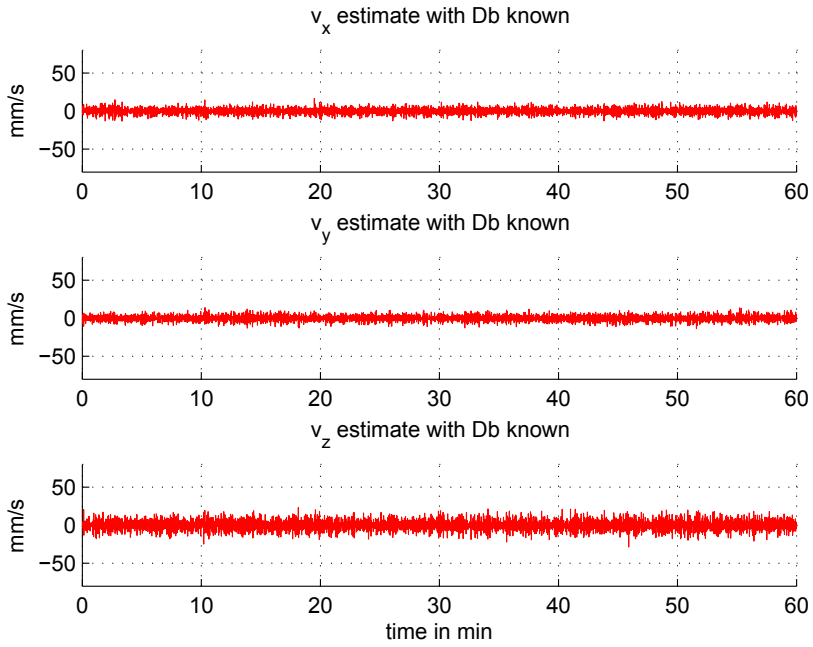

Figure 5.4: Ideally frequency corrected LMS velocity solution.

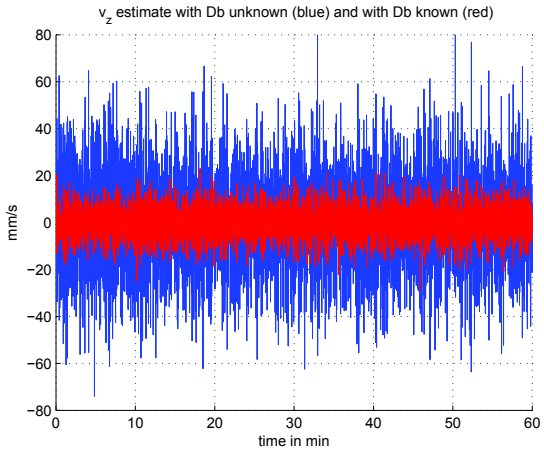

Figure 5.5: Vertical velocity: LMS and frequency removed LMS.

for the frequency error. Numerical values about the improvement produced by the frequency aiding are summarized in Table 5.2.

5.3.2 White noise frequency error

A preliminary analysis is done to test the possible performance of the frequency aided precise velocity solution. The easier case that can be imagined is to have a local oscillator whose frequency is affected only by white noise. Figures 5.7 and Figure 5.8 report respectively the LMS and the frequency aided velocity solution, while Figure 5.9 zooms on the vertical velocity comparing the two solutions. This hypothesis of white frequency may appear completely unrealistic. On the contrary, a good approximation of this case is possible in real life, when a bench of oscillators of same type is considered. According to the central limit theorem in fact, the mean of independent random variables identically distributed, with finite mean and variance, tends to be normally distributed at the growing up of the number of considered variables. Therefore, if a sufficiently large bench of same type oscillator is considered, their mean can be approximated as white gaussian noise process.
If, as in this case, the frequency error is simply white gaussian noise, a frequency estimate can be done simply averaging over time the measurement. This means that the filter to be used in this case is just an averaging filter. The result of the estimate is shown in Figure 5.6. Also in this case, a high improvement is present thanks to the frequency aiding. Numerical results are shown in Table 5.2.

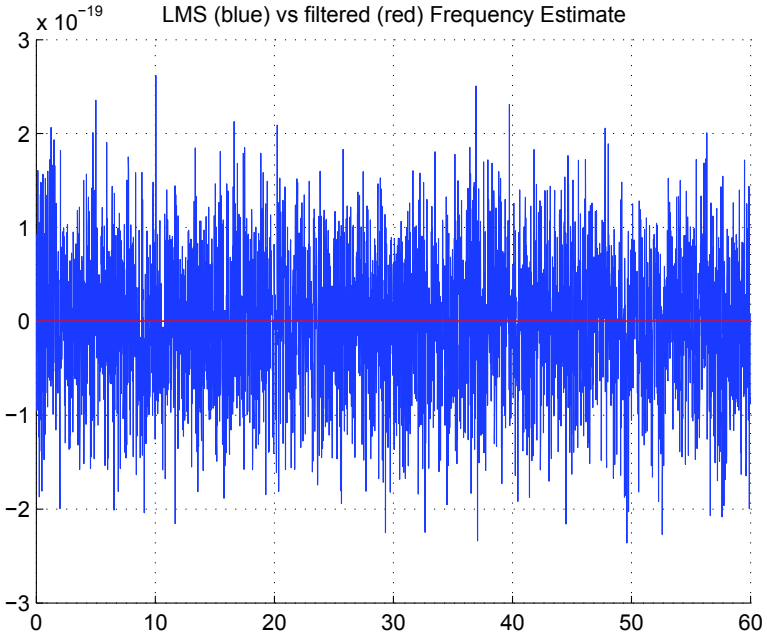

Figure 5.6: True frequency error (white noise) and estimate (averaging).

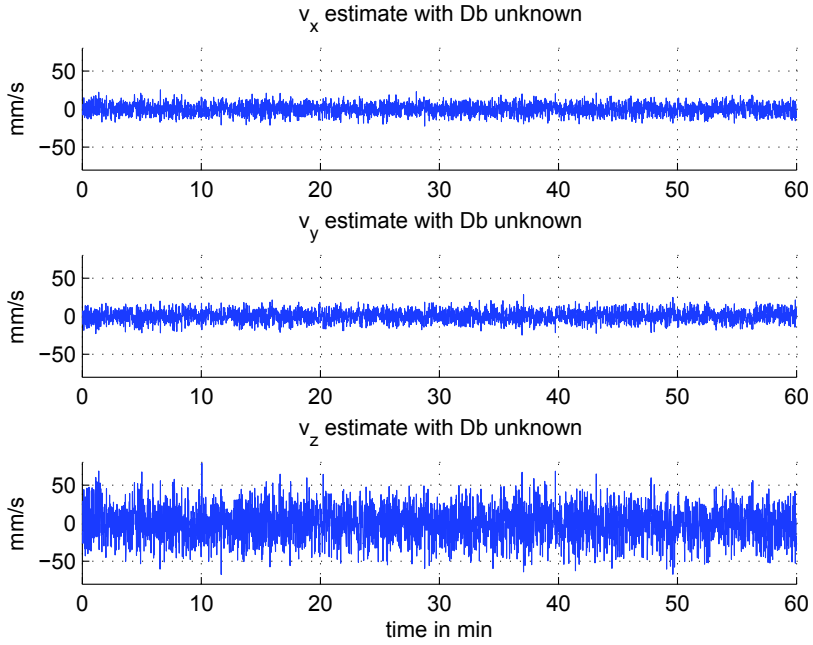

Figure 5.7: LMS velocity solution (white noise frequency).

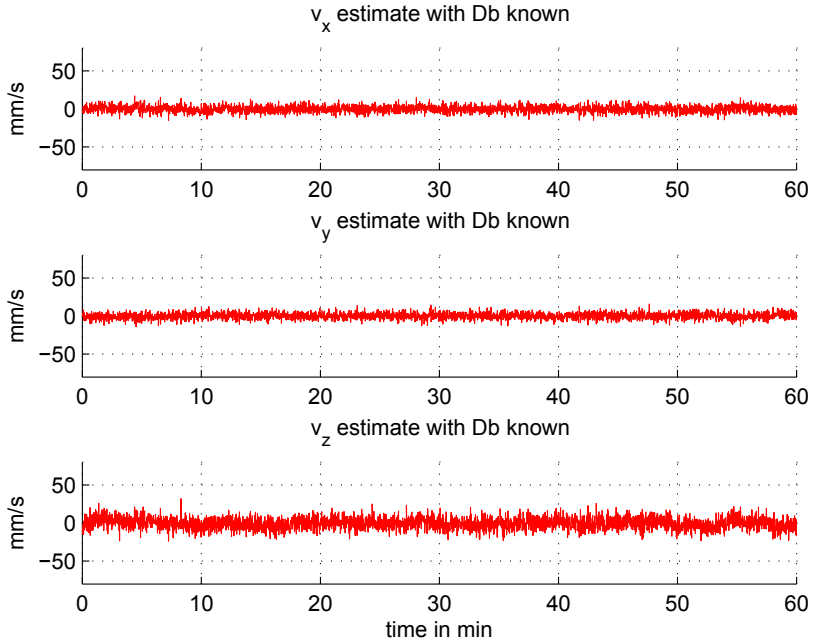

Figure 5.8: Frequency aided LMS velocity solution (white noise).

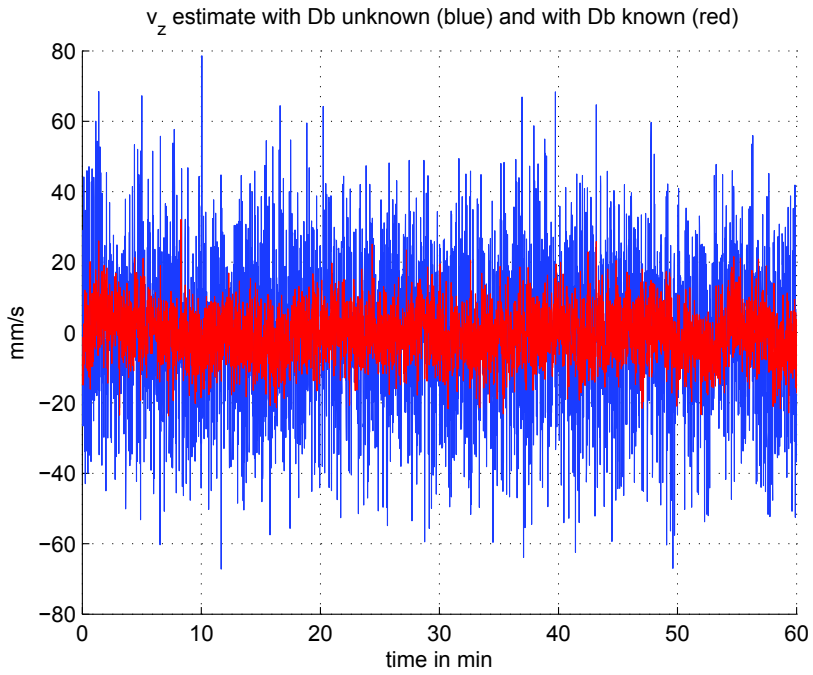

Figure 5.9: Vertical velocity: LMS and frequency aided LMS (white noise).

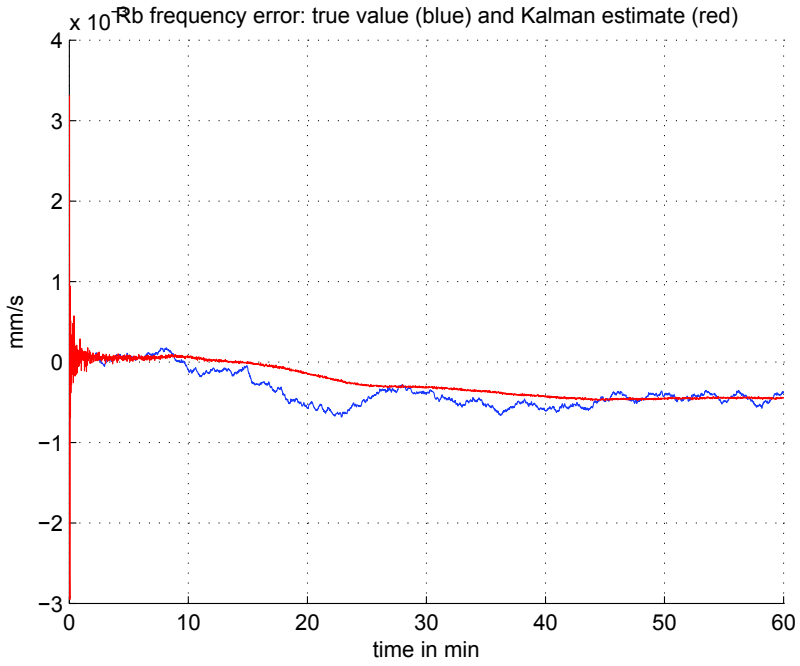

Figure 5.10: True frequency error (RW) and Kalman estimate.

5.3.3 Random Walk frequency error

If the frequency drift is modeled as a random walk, as in (4.10), the frequency error at a generic instant k is:

$$x_{f_k} = x_{f_{k-1}} + w_{2k-1,k} \tag{5.1}$$

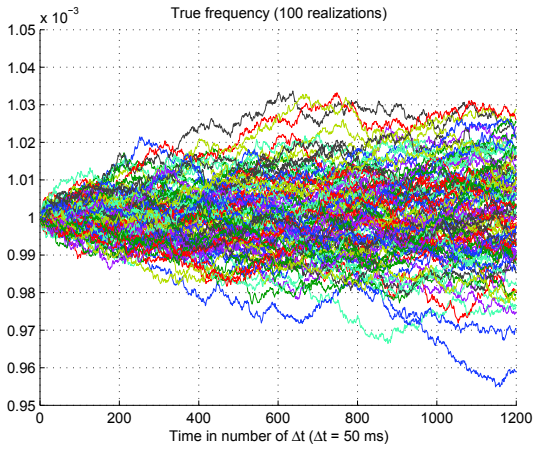

Figure 5.11: Frequency random walk (100 realizations).

Figure 5.11 displays as example a number of realizations ($N = 100$) of random walk frequency. In this case the filter used in the scheme in Figure 5.1 is a Kalman filter (see [44]).

One of the basic hypothesis in the Kalman filter theory is that both the measurement noise and the system noise must be white and uncorrelated.

Actually, in the frequency-aided velocity solution, the measurements do not satisfy this requirement. In fact the measurement in input to the Kalman filter is constituted by the LMS frequency estimate, which is affected by the noise present on the carrier-phase single differences. The hypothesis of white uncorrelated noise affecting the carrier-phase measurements is fair, but the operation of differentiation applied to obtain the SD measurements causes the carrier-phase SDs to be affected by correlated noise. As a consequence, the noise correlation reflects on the noise affecting the LMS frequency estimate. Therefore, the measurement noise in input to the Kalman filter is correlated.

In these simulations, uncorrelated noise was added to the carrier-phase SDs, so that the Kalman filter assumptions are satisfied, even if in the real life this is not the case.

The true frequency and its Kalman estimate are compared in Figure 5.10, while in Figure 5.12 the LMS frequency estimate is shown. A comparison is done of the velocity solution performance in the case of standard LMS solution in Figure 5.13, and in the case with frequency aid, in Figure 5.14. Figure 5.15 shows a comparison of the two solutions, for the vertical velocity.

Also in this case, for the numerical results see Table 5.2.

Theoretical performance has been computed and compared with simulation results obtained with Monte Carlo simulations, finding good agreement between the expected results and the obtained ones. Theoretical performance has been computed and compared with simulation results obtained with Monte Carlo simulations, finding good agreement between the expected results and the obtained ones.

As desired in a Kalman filter, the residuals are uncorrelated, thanks to the characteristics of the simulation, where the measurement noise is white and uncorrelated. The uncorrelated residuals in Figure 5.16 reveal that the Kalman filter design is optimal, which means that both the system and the measurements are properly characterized. If the residuals are correlated means on the contrary that some system characteristics have been mismodeld.

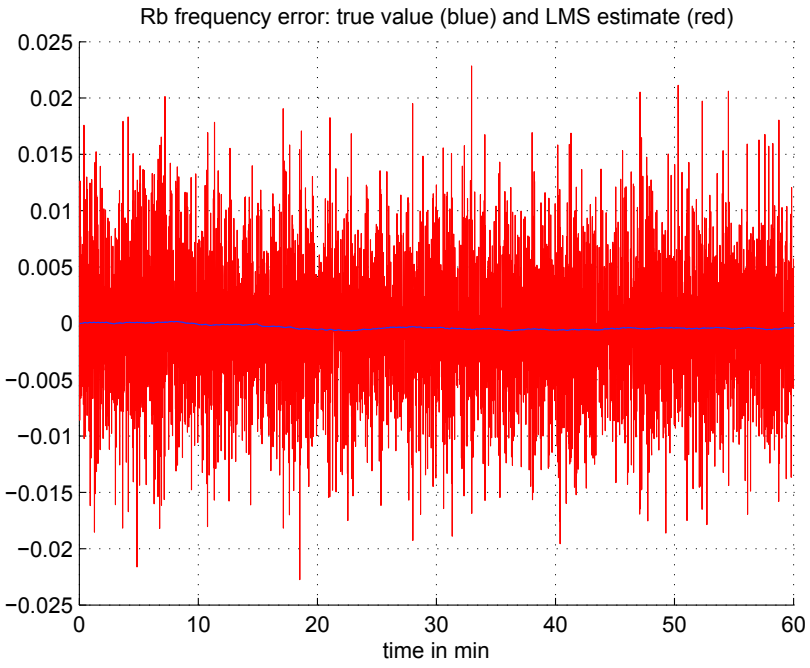

Figure 5.12: True frequency error (RW) and LMS estimate.

Additive frequency bias (Gauss-Markov model)

Due to aging effects, as described in Section 4.5.2), in a Rb oscillator a not constant frequency bias is present, which varies slowly, i.e. with very long time constant. An approximation of such variations can be a Gauss-Markov process, as in (5.1). Analysis have been done on this topic, but the results obtained are

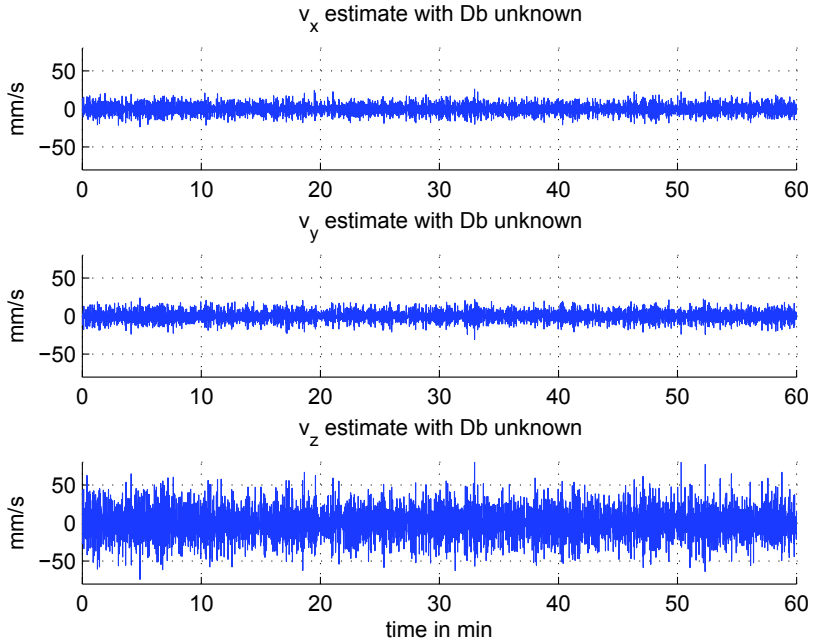

Figure 5.13: LMS velocity solution (random walk).

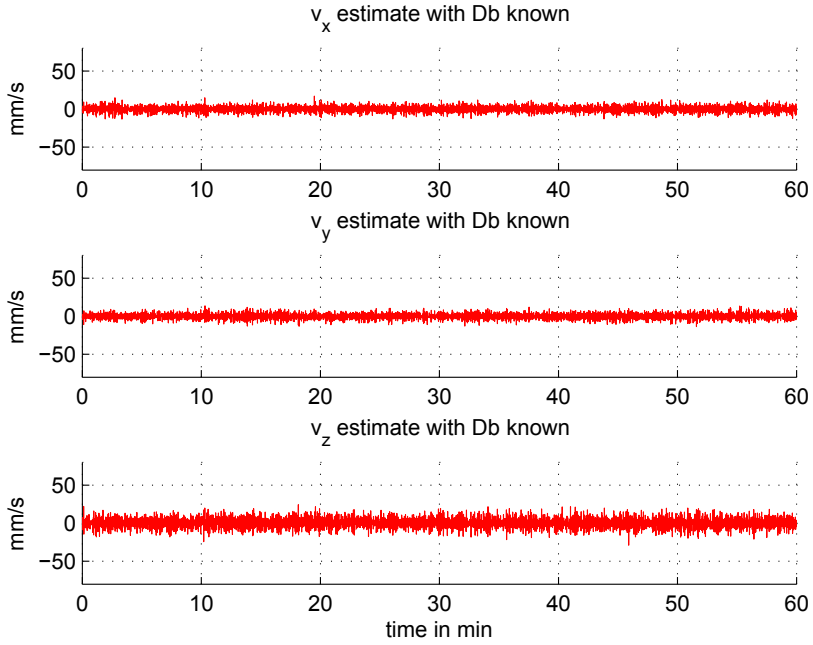

Figure 5.14: Frequency aided LMS velocity solution (random walk).

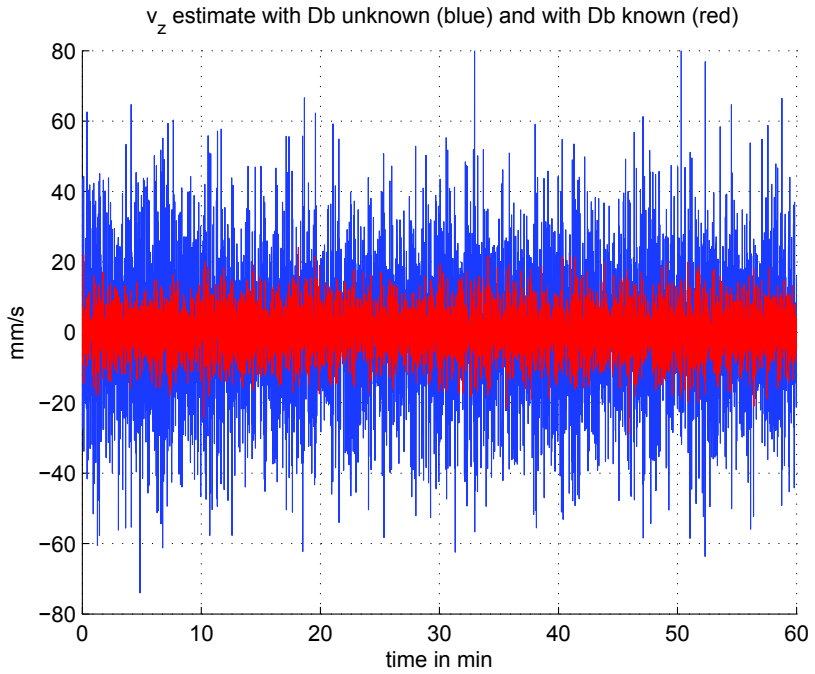

Figure 5.15: Vertical velocity: LMS and frequency aided LMS (random walk).

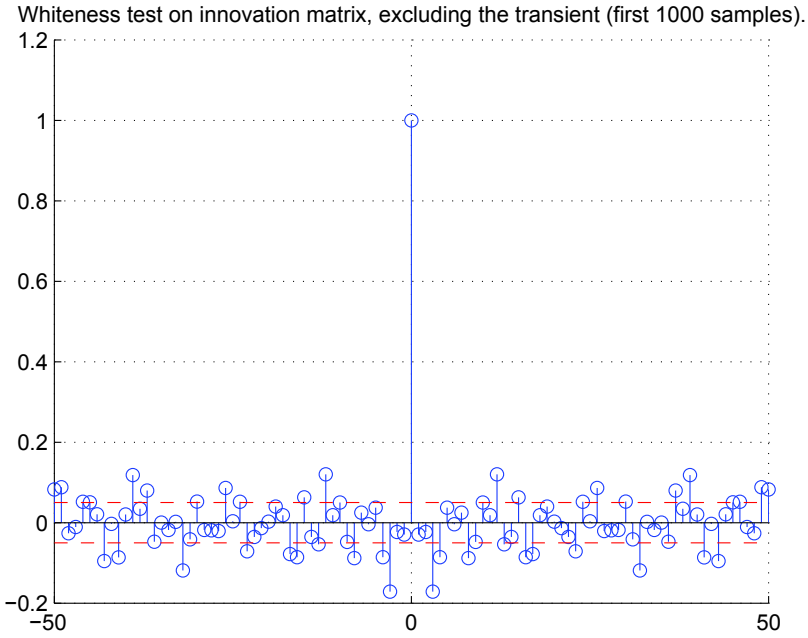

Figure 5.16: Whiteness test result on Kalman filter residuals.

very preliminary and are not shown in this thesis. This issue will be object of future work.

One of the main problem of Gauss Markov processes is that, due to the long time constant of the process, a long observation time is needed in order to find a proper model, which means to correctly estimate the parameters of the system.

5.4 Test with real oscillator

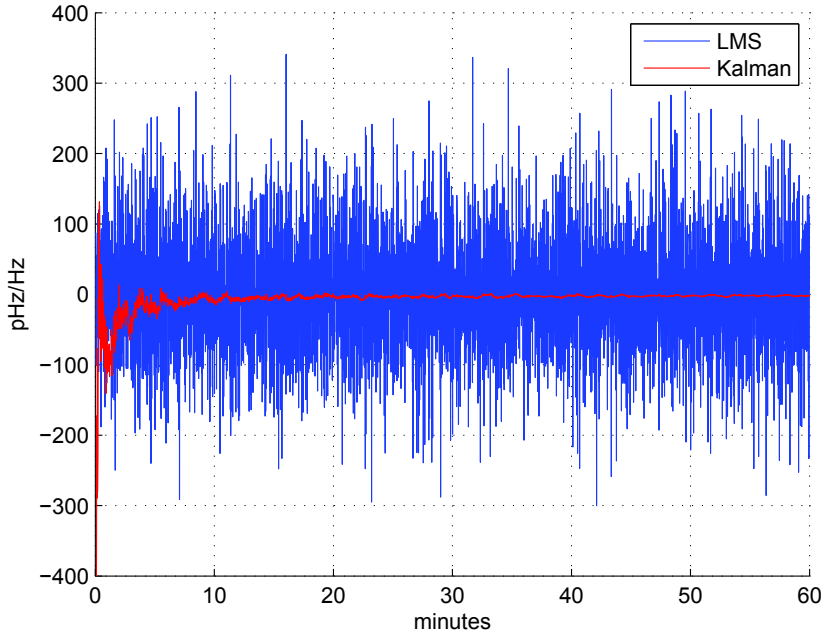

Figure 5.17: LMS (blue) and Kalman (red) estimate of the local frequency (Rb). Novatel receiver, simulated GPS.

In this section, some results obtained with a real Rb oscillator are presented. Tests have been performed using a Novatel receiver using the Rb as external frequency and time reference, while the GPS signal come from a GPS simulator. The use of simulated GPS data allows to analyze the effects of the Rb frequency error, without any disturbance from the GPS system, in particular, without any multipath.

Figure 5.17 compares the local frequency estimate from the standard LMS with the Kalman estimate. Figure 5.18 gives a zoom on the Kalman estimate.

Figure 5.19 shows a comparison between the velocity solution computed with a

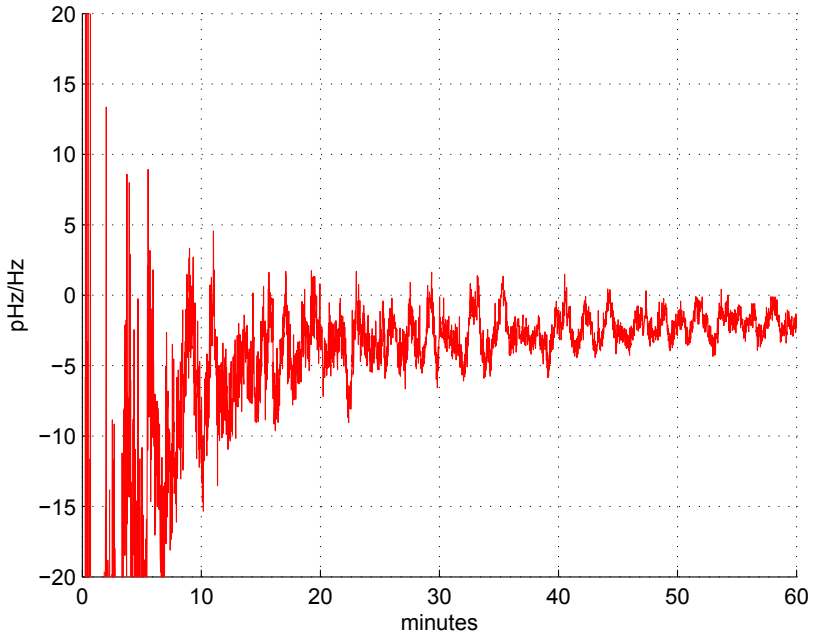

Figure 5.18: Kalman estimate of the local frequency (Rb). Novatel receiver, simulated GPS.

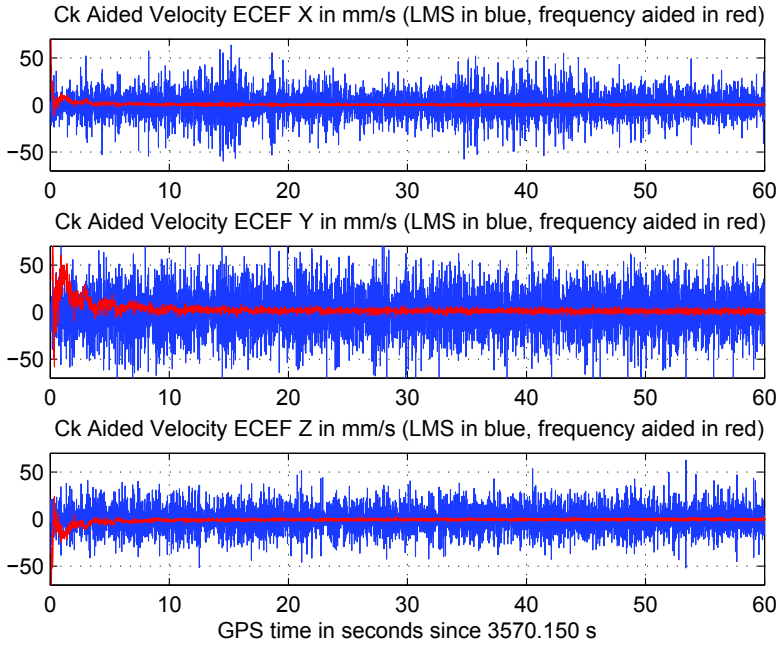

Figure 5.19: Velocity (ECEF) solution: LMS (blue) and Kalman frequency aided (red) estimates. Novatel receiver, simulated GPS.

standard LMS with four unknowns and the frequency-aided velocity solution. It can be seen how the frequency aid can improve the solution accuracy.
Note that here the velocity solution is in ECEF coordinates. Then, only 4 satellites are used, and the DOP is not very worse on one direction with respect to the other ones, as indicated in Table 5.2. These are the reasons why, as shown in Table 5.2, there is not high difference between the DOP on the three axes, and as a consequence there is not a higher improvement on one axis with respect to other ones, using the frequency aided solution, as expected from the discussion in Section 2.3.3, in particular in the sub-section 2.3.3.
Even if the case is ideal using a simulator (no multipath, no interference, good SNR), the carrier-phase noise is present due to the effects of the receiver. Therefore, also the noise on the carrier-phase single differences is present, which is correlated. However in this case, the neglection of the noise correlation, using a standard Kalman filter, does not have high impact on the solution performance, thanks to the very small amount of noise.
In real life, where the noise effects are higher, the noise correlation cannot be neglected anymore.

5.5 Correlated noise and Kalman filters

As explained in Section 5.3, in real life the frequency measurement is taken as single difference of the phase over a time interval, and this causes to have correlated noise on the measured SDs. If this correlation in the measurement noise is not taken into account, one important hypothesis is not valid anymore in the Kalman filter design.
In a nut-shell, what a Kalman filter must do is to include in a model what in the system can be modeled, while the only components that cannot be forecasted are the random noise variables, which can be characterized by the white noise power spectral density, giving the expected deviation caused by the noise. If a noise component is considered to be white but in reality it is correlated, it means that in the Kalman filter some characteristics that could be modeled are not modeled. As a result, the filter is not optimal and if the un-modelled correlations are significant, the filter does not work anymore, since the model results to be completely wrong.
The mismodelling appears in the whiteness test on the Kalman residuals, as

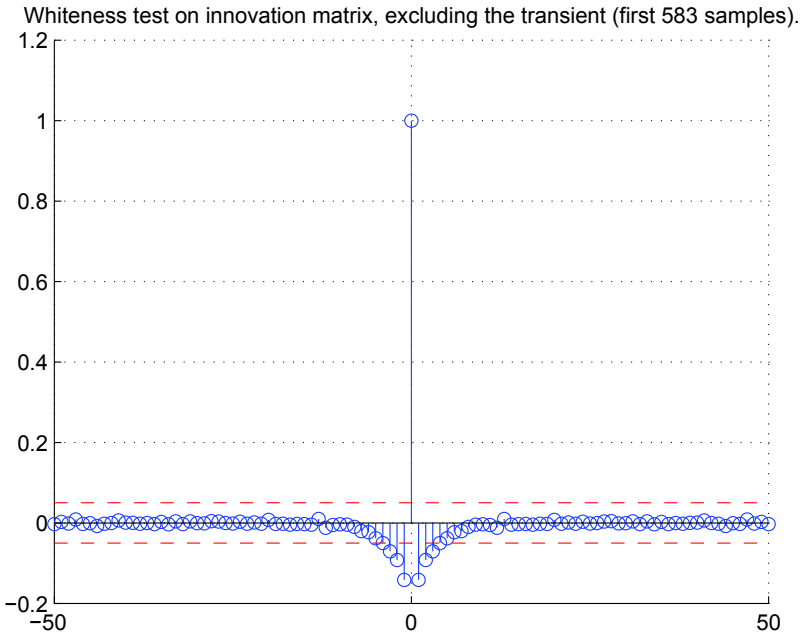

Figure 5.20: Whiteness test result on Kalman filter residuals.

explained in Section 5.3.3. Figure 5.20 shows the test applied to a simulation where the noise on the carrier-phase measurements is correlated, as it happens in the real case, and applying those measurements to a standard Kalman filter. In the real life, the problem of the noise correlation cannot be neglected. This problem has been firstly addressed in [54], then for different applications, as in [55], [56], [57], [58], [59], [60]. Some analysis have been done already regarding the frequency aided velocity solution, but work still needs to be done. A solution is for example to consider a modified Kalman filter to take into account this correlations.

5.6 Final remarks on frequency aided precise velocity

Table 5.2 summarizes some simulation result providing a comparison between some tested methods, in the same conditions as in Table 5.1.

From Table 5.2 it can be seen how the results obtained with filtered frequency estimates are not far from the ideal case where the frequency is perfectly known.

	LMS	freq aided		freq aided		LMS	freq aided		
mod	Rb	Rb		Rb		awgn	awgn		
est	none	ideal		Kalman		none	averaging		
	σ_{LMS}	σ_{LMSB}	ratio	σ_{LMS}	ratio	σ_{LMSB}	σ_{LMSB}	ratio	DOP
v_x	2.398	2.383	1.006	2.383	1.006	2.400	2.366	1.014	0.84
v_y	3.237	3.135	1.033	3.135	1.033	3.193	2.087	1.530	1.14
v_z	4.650	2.584	1.800	2.588	1.797	4.632	2.593	1.786	1.61

Table 5.2: Performance summary

The analysis done so far outlines how the frequency-aided velocity can be a good way to improve the GPS (and, in general, the GNSS) velocity performance.
One of the main issues to be solved is to test the accuracy of the frequency models, using real data. As a final remark it can be said that the main troubles that have been found and needs to be addressed are the following:

- distinguish between oscillator errors and measurement system errors, in case of both Cesium referenced counter or GPS;
- model the residual frequency bias, which means to find the parameters of a Gauss Markov process with long time constant;
- address the problem of the time correlation in the measurement noise (for Kalman filters).

Only the static case is analyzed here, but also dynamic models are needed, since the aim is to use the Rb models to improve the velocity estimate. Dynamic tests done in Chapter 4 are the starting point for the next work involving dynamic models.

Part III

Adaptive
Phase Lock Loop

Chapter 6

Digital Phase Lock Loops

6.1 Introduction to Phase Lock Loops

A Phase Lock Loop (PLL) is a closed loop architecture that, given a sinusoidal signal as input, gives as output the estimated phase error between the input and its local estimate. The aim of a PLL is to recover the error, minimizing the tracking error.

In order to get phase synchronization, several techniques can be used. Both a open and a closed loop architecture are possible, even if a close loop architecture, a PLL, is generally used in GNSS, in order to allow the phase detector to work in linear region (small error region, see Section 6.6.1). The loop closing in fact means that the local estimate is corrected according to the estimated error, while in the open loop architecture the local signal is kept constant, so the device only estimates the error without minimizing it.

A lot of work exists in literature about PLLs, referred to different applications. Entire books also exist about PLLs, as for e.g. [61], [62], [63], [64], [65].

A PLL can be designed both in the analogical and in the digital domain, and in both cases it is composed by three essential elements: a Phase Detector (PD), a loop filter and a controlled oscillator, that is a VCO (Voltage Controlled Oscillator) in the case of analogical PLL and a NCO (Numerical Controlled Oscillator) in a Digital PLL (DPLL).

A PD compares the phase of the input signal with the phase of the local signal and outputs a measure of the phase error between them. The estimated error is filtered by the loop filter, and the filtered error controls the oscillator, which means that the local signal is changed according to the estimated error, in order

to track the input signal.
In the ideal world, if there are not disturbances on the input signal and if the PD is perfect, there is not any need of filtering the error. In order to correct the local signal and track the input phase, in this ideal world it is enough to simply apply the estimated error to the oscillator.
In the real world, the PD is not perfect, therefore it can just give an approximated measurement of the phase error. However, the main un-ideality is represented by the disturbances affecting the input signal, in particular the noise. In the real world, if the PD output is directly used to pilot the local oscillator, the effects of the disturbance may cause instabilities and even lost of lock. Therefore the use of a loop filter is mandatory, which has to be properly designed depending on the application, taking into account the expected characteristics of the input signal, in particular the dynamics that it can assume and the noise that could affect it.
Furthermore, even if the PLL measures a phase error, the local oscillator is tuned changing its frequency, and not its phase and this is done for several reasons. In the analogical domain, abrupt phase changes are not practical and a continuity in the signal must be guaranteed. In the digital domain this change in the phase could be possible, but it is not used, on one side because the local signal continuity is desired as well, but the main reason is another one.
Let us imagine to represent the input and the local signal phase as phasors, laying in the phasors plane. The phasors are rotating at a certain frequency, one at the input signal frequency and the other one at the local frequency. The phase difference between the phasors is the phase difference between the signals.
Since both the phase and the frequency of the input signal are unknown, the local signal has both the phase and the frequency that are different from the input. Furthermore, the phase error changes over time due to the frequency error between the signals.
If the correction on the local signal is applied on the signal phase, i.e., applying to the signal a phase jump without changing its frequency, the lock cannot be maintained, because of the frequency misalignment. Therefore, a frequency lock is needed first of all. In this way, the local phasor can be in track of the input one, rotating at its same velocity in the phasors plane.

6.2 Motivations

In GNSS the PLL is used to lock the local carrier-phase to the incoming signal. This is a crucial operation in a GNSS receiver, fundamental to perform range estimates. In fact, if a code-measurement is performed, the carrier removal is needed, as explained in Section 1.5.3, which implies the phase-lock between the local and the incoming signal. Furthermore, accurate range estimates can be performed using the phase information as described in Section 1.5.2 (after solving for the integer ambiguity), and in order to do that the phase must be locked and measured accurately. .

In this thesis the interest on the PLLs is related to the single differences of the carrier-phase measurements to be used in the velocity estimate, as described in Chapter 3. Thanks to the differentiation over a time interval, the integer ambiguity does not represent a problem for the velocity solution, if the absence of cycle slips is verified. In this thesis, the importance of accurate phase measurements is very high, since the objective is to find high-accurate velocity solution.

Figures of merit for a PLL can be defined such as tracking error, in terms of residual bias an tracking jitter, convergence time and robustness, in particular in case of changes in the scenario.

The goal that this thesis wants to address is the possibility to adapt the PLL parameters according to the input signal characteristics, in order to optimize the performance in terms of the figures of merit cited above.

Nowadays there is an increasing interest towards SDR receivers (described in Section 1.6.1), therefore the choice has been to analyze Digital PLLs (DPLLs), with the aim to include the designed loop in a GNSS fully software receiver [24]. Thanks to the software design, a flexible architecture is developed, that can easily changed just modifying some parameters.

6.3 Scenario overview

Before to analyze the PLL characteristics, an overview of the scenario where it has to operate is fundamental to outline the objectives of the work.

In general, and in particular in GNSS, the phase of the signal that has to be tracked is not constant, but it varies due to several factors, in particular due to frequency changes. The frequency can be expressed as in (6.1) and the total

phase is a function of the frequency, following the relation in (6.2).

$$f(t) = \frac{1}{2\pi}\frac{d}{dt}\phi(t) \tag{6.1}$$

$$\phi(t) = 2\pi f(t)t + \theta(t) \tag{6.2}$$

A phase shift can occur, due for example to a delay τ (unknown) introduced by the channel. In this case if the transmitted signal is

$$x\left(t\right) = A \cdot \cos\left(\omega_c(t)t + \theta(t)\right) \tag{6.3}$$

where $\omega_c(t) = 2\pi \cdot f_c(t)$, being f_c the signal frequency, and $\theta(t)$ the initial phase, the, neglecting both noise and other effects, the received signal is:

$$\begin{aligned} x\left(t\right) &= A \cdot \cos\left(\omega_c(t)\left(t-\tau\right) + \theta(t)\right) \\ &= A \cdot \cos\left(\omega_c(t)t - \omega_c(t)\tau + \theta(t)\right) \\ &= A \cdot \cos\left(\omega_c(t)t + \psi(t)\right) \end{aligned} \tag{6.4}$$

where $\psi(t) = \theta(t) - \omega_c(t)\tau$ is the new unknown initial phase, which depends on the channel delay τ. Even if the time delay τ is very small, the constant phase shift in Equation 6.4 can be very large in GNSS applications, since $\omega_c(t)$ is large because of the high carrier frequency (in the order of more than 1 GHz).
It is important to notice that with the term *phase* it can be indicated both the entire argument of the sinusoidal function $\phi(t) = \omega_c(t)\left(t-\tau\right) + \theta(t)$ and the initial phase $\theta(t)$. In this thesis the term phase will be generally used to indicate the entire argument $\phi(t)$ of the sinusoidal signal, because the phase-lock is reached when the total phase is properly tracked.
The phase $\phi(t)$ can change due to several causes and in GNSS the main effect is due to the frequency variations caused by Doppler effects, which in GNSS are always present since the transmitters (satellites) continuously move, and also the user can move. The relationship between the relative velocity transmitter-receiver and the observed frequency is:

$$f = (1 + \frac{\Delta v}{c_0})f_0 \tag{6.5}$$

where $\Delta v = v_r - v_t$ is the velocity difference between the receiver and the transmitter, c_0 is the propagation velocity of the signal in the medium and f_0 is the

frequency emitted by the transmitter. The change in frequency is:

$$\Delta f = \frac{\Delta v}{c_0} f_0 \tag{6.6}$$

Similarly, the Doppler rate is related to $\Delta a = a_r - a_t$, the difference between the receiver and the transmitter accelerations, following the equation:

$$\Delta^2 f = \frac{\Delta a}{c_0} f_0 \tag{6.7}$$

Constant velocity or constant acceleration mean, respectively, to have constant Doppler frequency and linear phase, or constant doppler rate and linear frequency and quadratic phase.
If the user is in static condition, the Doppler frequency due to the satellite motion changes during the day depending on the satellite position over the horizon. Due to the satellite trajectory, a user on the earth sees a Doppler and a Doppler rate that change with the position of the satellite above the horizon.
Considering the satellite motion (see Section 1.7.1), the maximum Doppler is present when satellite is at the lower point on the horizon, where the Doppler rate is minimum, and viceversa the Doppler rate is maximum if the satellite is at the zenith, while at that point the Doppler is minimum. Due to the satellite, the maximum Doppler that the user can experience is about 5 kH, while the maximum Doppler rate is about 1 Hz/s.
A user moving with high accelerations experiments values of Doppler and Doppler rate much higher: this is the case of high-dynamic users, which is the case that this thesis aim to address.

6.4 PLLs

6.4.1 PLL basic elements

PLLs provide filtering, mitigating the effects of input noise, and thanks to this important capability they are very widespread systems. A PLL can be designed in order to follow more or less quickly the input or to remove more or less the noise, and global performance depend on these choices.
As introduced in Section 6.1, a PLL contains three main elements:

1. phase detector (PD)

2. a loop filter

3. a controlled oscillator (VCO in analog PLL or NCO in digital PLL)

The PD compares the phase of a periodic input signal against the phase of the local controlled oscillator signal, then the output is a measure of the phase error between the two signals. The error is then given as an input to the loop filter, that is a low pass filter and allows to decrease the noise effects, then the output of the loop filter is applied as control signal to the oscillator. When the loop is locked, the controlled oscillator sets its average frequency to the average frequency of the input signal. In the following, the PLL structure is described, with the focus on the digital PLL.

6.4.2 PLL phase model

A PLL can be represented using a phase model as in Figure 6.1.

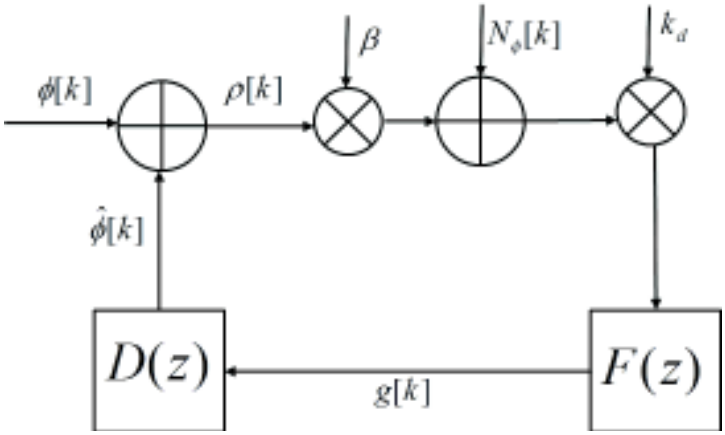

Figure 6.1: Phase model of a PLL.

where $\phi[k]$ is the input phase, $\hat{\phi}[k]$ is its local estimate, $\rho[k]$ is the phase error. The model in Figure 6.1 refers to digital PLLs and the time k is the digital time, which in the continuous time t represents the instant kT, calling T the updating interval of the phase estimate.
In real life, the transmitted signal is affected by noise and disturbances, that have effects on the signal phase. A phase noise $N_\phi[k]$ can be modeled that corresponds to a certain amplitude noise on the input signal, as in Figure 6.1. It is realistic to consider $N_\phi[k]$ as white and gaussian noise.

Notice that the power spectral density of $N_\phi[k]$ depends on the updating time T, which in the phase model represents the sampling time. In fact, the larger T is, the longer the phase noise results to be integrated, therefore the smaller the variance of $N_\phi[k]$ is, and the smaller the impact of the noise on the phase estimate is. Further details will be given in Section 6.7.

PD model

In Figure 6.1 the Phase Detector is modeled as a sum block that compute the phase error computing the phase difference between the input and the local phase, followed by the block that multiplies the result by β. This parameter is introduced to take into account the fact that in the phase model a linearization is operated to model the PD. In fact in real life the phases are not directly available but the phase error information has to be extracted using signal processing techniques. As explained in Section 6.6.1, the PD used in this thesis leads to $\beta = 1$.

NCO model

In the phase model there is not a controlled oscillator, since only the phase is considered. The block that simply model it is a numerical integrator, whose transfer function is

$$D(z) = \frac{z^{-1}}{1 - z^{-1}} \tag{6.8}$$

What this block does is simply to update the phase estimate to be used in the next time step, by summing to the current phase estimate the phase correction computed at the current step.

Loop filter

The loop filter $F(z)$ in the phase model is exactly the same as in the real DPLL. The aim of the loop filter is to mitigate the effects of the disturbances on the estimated phase error, taking into account the trend of the input, by means of memory elements (in the case of filter with at least one pole). The filtering can work just as a attenuator (filter of order zero), or it can have memory elements, that allow to take into account the trend of the signal in the time. Higher order of the filter means more memory blocks and more poles in the transfer function,

and also ability to track phases behaving as functions of higher order (e.g. linear phase or quadratic phase).

The general expression of the transfer function of a DPLL loop filter is:

$$F(z) = \frac{az^{N-1} + bz^{N-2} + (N-2)\,c}{(z-1)^{N-1}} \tag{6.9}$$

where $N = 1,2,3$ is the order of the PLL (see Section 6.7.1).

If N = 1, it is simply $F(z) = constant$. Values of $N > 3$ are not practical.

This thesis is focused on the analysis of second and third order loops. The design and the meaning of $F(z)$ is discussed in detail in Appendix A and in Appendix B.

In this thesis, the loop gain (shown in Figure 6.1) is considered to be unary: $k_d = 1$.

6.5 Closed loop: transfer function and phase error

The phase error can be written in the Z-transform domain [66]. Considering the ideal case without noise, it is:

$$\begin{aligned} \rho(z) &= \phi(z) - \hat{\phi}(z) \\ &= \phi(z) - \phi(z) \cdot H(z) \\ &= \phi(z) \cdot (1 - H(z)) \end{aligned} \tag{6.10}$$

where $H(z)$ stands for the closed loop transfer function. The closed loop transfer function is expressed by (6.11).

$$\begin{aligned} H(z) &= \frac{\hat{\phi}(z)}{\phi(z)} \\ &= \frac{\beta \cdot F(z) \cdot D(z)}{1 + \beta \cdot F(z) \cdot D(z)} \\ &= \frac{\beta \cdot F(z)}{(z-1) + \beta \cdot F(z)} \end{aligned} \tag{6.11}$$

where $D(z)$ represents the NCO, as in (6.8).

Substituting (6.11) in (6.10), a general expression for the phase error is found:

$$\begin{aligned} \rho(z) &= \phi(z) \cdot \left(1 - \frac{\beta \cdot F(z)}{(z-1) + \beta \cdot F(z)}\right) \\ &= \frac{z-1}{(z-1) + \beta \cdot F(z)} \cdot \phi(z) \end{aligned} \tag{6.12}$$

Making use of the Z-transform property in accordance with

$$\mathcal{Z}\{x(k+1)\} = z \cdot \mathcal{Z}\{x(k)\} \tag{6.13}$$

it can be noted that the anti Z-trasform of the numerator in Equation 6.12 can be expressed as

$$\begin{aligned} \mathcal{Z}^{-1}\{(z-1)\cdot\phi(z)\} &= \mathcal{Z}^{-1}\{z\cdot\phi(z)-\phi(z)\} \\ &= \mathcal{Z}^{-1}\{z\cdot\phi(z)\}-\mathcal{Z}^{-1}\{\phi(z)\} \\ &= \phi[k+1]-\phi[k] \end{aligned} \tag{6.14}$$

So, dividing by z, it is:

$$\begin{aligned} \mathcal{Z}^{-1}\left\{\left(1-z^{-1}\right)\cdot\phi(z)\right\} &= \mathcal{Z}^{-1}\left\{\phi(z)-z^{-1}\phi(z)\right\} \\ &= \mathcal{Z}^{-1}\{\phi(z)\}-\mathcal{Z}^{-1}\left\{z^{-1}\phi(z)\right\} \\ &= \phi[k]-\phi[k-1] \end{aligned} \tag{6.15}$$

Calling $\Delta_\phi[k]$ the change in the input phase over a time interval

$$\phi[k]-\phi[k-1] = \Delta_\phi[k] \tag{6.16}$$

the phase error for a generic PLL of order N can be expressed as

$$\begin{aligned} \rho(z) &= \frac{z\cdot(1-z^{-1})}{(z-1)+\beta\cdot F(z)}\cdot\phi(z) \\ &= \frac{z}{(z-1)+\beta\cdot F(z)}\cdot\Delta_\phi[k] \end{aligned} \tag{6.17}$$

The stability condition has to be satisfied in the PLL design, which implies that all the poles of the closed loop transfer function $H(z)$ lie within the unary circle on the complex plane, i.e. it has to be:

$$|p_i| < 1 \text{ for } i = 01,2,...N \tag{6.18}$$

6.6 DPLL architecture and parameters

The objective of this work is to design an adaptive PLL and implement it in N-Gene [24], the software receiver developed by the NavSAS group, therefore a digital PLL must be designed.
Digital signal processing operates under several conditions, which of course are valid also in this specific case of digital PLLs:

- Signals are sequences of discrete samples.
- A sample contains as information a number (stored as digital number)
- Quantization is indispensable, since digital numbers necessarily have finite precision they have to be quantized.
- Operation within the digital PLL are computed. Note that a digital PLL is sometimes called software PLL, but that term can be referred both to a software and an hardware implementation, since in some cases, like this one, the underlying algorithms can be used equally well in either software or hardware.

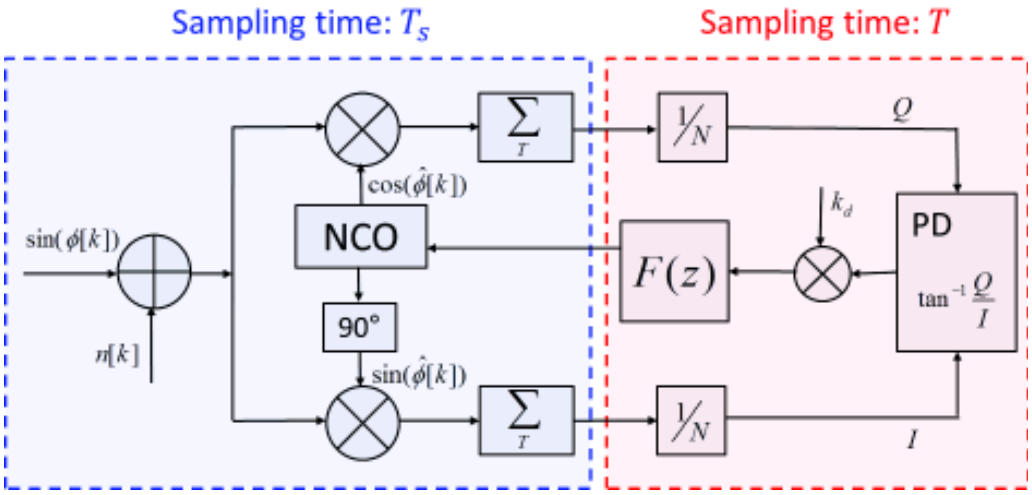

Figure 6.2: DPLL architecture.

The DPLL architecture, described in depth in [61], is shown in Figure 6.2. The main blocks are described in the following.

6.6.1 Phase Detector

The phase detection of a real signals is not a trivial operation.
At the input of the DPLL there is a sampled signal, which can be represented by a sinusoidal signal whose frequency changes and that is affected by noise. affected by noise, and the objective of a PD is to detect the phase error between that noisy signal and a local signal. The idea is to divide the In-phase and in-Quadrature components of the phase error, in order to describe it error as a

phasor, i.e. by means of its In-phase I and its in-Quadrature Q components. Since noise is present, the estimate cannot be performed considering just one sample of the signal, which would be affected too much by any disturbance, but it is performed taking into account a certain time interval during which the signal is integrated. In this way the effects of the noise, which is supposed to be zero mean, are mitigated. So, the phase detection is performed through several operations:

- Splitter
- Mixers
- Integrate and dump blocks
- Phase Detector, that estimates the phase error

A splitter spreads the input signal on two branches, where the signal is multiplied respectively by two orthogonal signals generated locally (one signal is the copy of the other one, with a $\pi/2$ phase shift).
After that, the signals are integrated over an integration time T, so that the IF component cancels, since the period of that component is much smaller than T and over that longer time an integration operation gives as a result almost nil. The output of the integrator is taken after a time T and divided by the number of samples summed over that time. Thanks to this averaging operation, the noise component, which can be considered to be WGN, with zero mean, results to be mitigated. What is obtained on the two branches respectively are the In-phase and in-Quadrature component of the phase error:

$$I[k] \simeq cos(\rho[k]) \tag{6.19}$$

and

$$Q[k] \simeq sin(\rho[k]) \tag{6.20}$$

where k is the sampling time at the sampling frequency $1/T$.
What the phase detector gives as output is the estimated phase error $\rho[k]$, which is the average error over the interval T. In this thesis, an arcotangent type PD is considered, which has as input the I and Q components, and can easily compute the estimated phase error. By definition it is:

$$\rho[k] \simeq atan\frac{sin(\rho[k])}{cos(\rho[k])} \tag{6.21}$$

Linearization

In Section 6.4.2, the PD is modelled considering a linearization parameter β, to take into account the characteristics of a real phase detector and its linearization. Even if the PLL is nonlinear (a real PD, as the arcotangent type for example, is not linear), the tracking and phase error theory is based on the assumption that the phase error is small enough to allow a linear approximation around the zero-error point. This means to neglect the non-linear effects and to consider the loop to be linear. The more the error increases, the less this hypothesis is plausible. Nonlinearity involves effects that are mostly unpredictable with theory, but that simulations can show. If the error becomes too large, this assumption turns completely wrong and the loop losses the phase lock.
The residual estimation error affecting $\rho[k]$ depends on the type of PD implemented, since the estimated $\rho[k]$ depends on the true phase error through a non-linear function, the so called discrimination function *S-curve*. For small errors, the *S-curve* can be linearized as:

$$e[n] \simeq \beta\left(\phi - \hat{\phi}[n]\right) \tag{6.22}$$

$$\simeq \beta \cdot \rho[n] \tag{6.23}$$

where

$$\beta = \frac{\delta}{\delta\rho} S(\rho)\,|_{\rho=0} \tag{6.24}$$

For an arctangent type PD, as considered in this thesis, it is $\beta = 1$ since in the origin (i.e. for error equal to 0) the tangent of the S-curve has unary slope.

6.6.2 Loop filter

The loop filter block is as for the phase model (see Section 6.4.2). See Appendix A for details.

6.6.3 NCO

The Numerical Controlled Oscillator produces a periodic signal and another one shifted by 90°, as explained in Section 6.6.1. The oscillator is controlled by the loop filter output $g[k]$, i.e. by a filtered version of the estimated phase error. As discussed in Section 6.1, the phase is tracked by controlling the local

frequency. Therefore the phase correction $g[k]$ estimated by $F(z)$ is transformed in an equivalent correction in frequency, which in the digital domain is, using the incremental ratio in the sampled time:

$$\Delta\hat{f} = \frac{g[k] - g[k-1]}{2\pi T} \tag{6.25}$$

The correction $\Delta\hat{f}$ is applied (added) to the current local frequency, keeping in this way also the continuity in the local signal.

6.7 Equivalent Noise Bandwidth

The loop Equivalent Noise Bandwidth (ENB), indicated in the following as B_L, can be expressed by the integral relation

$$B_L = \frac{1}{2\pi T} \oint H(z)H(z^{-1})z^{-1}dz \tag{6.26}$$

where $H(z)$ is the closed loop transfer function of the loop.
The ENB can be defined as the bandwidth of the ideal rectangular filter that allows passing the same quantity of noise that actually the considered filter allows to pass through. This parameter is one of the most important characteristics of a PLL, because it is directly proportional to the tracking jitter and it is inversely proportional to the transient (time and energy) and to the Steady State Error (SSE), if a SSE is present.
The ENB can be seen as the *window* through which the loop can look at the frequency components of the input signal, which means that the wider the bandwidth, the more the loop has the possibility to see changes in the input frequency. On the contrary, if ENB is very tight, dynamics effects can easily cause the frequency to go out of the window, small, causing the PLL to loose the lock on the input frequency.
On the other side, the wider the bandwidth, the more disturbances affects the loop, since more power of noise can enter in, and eventual interference component on the signal spectra can be involved too. A small bandwidth has better performance in term of noise and disturbances rejection.
Equation 6.26 shows that the ENB depends on both $H(z)$ and the integration time T. This is equivalent to say that $H(z)$ is a function of the parameter $B_L \cdot T$. The understanding of these parameters is fundamental for the correct design of a tracking loop.

Let us consider the PLL phase model in Figure 6.1. In this model the integrators working at integration time T are not involved, but the integration time T represents here the sampling time of both the input ϕ, the estimated $\hat{\phi}$ and the additive phase noise component N_ϕ. The closed loop transfer function $H(z)$ does not depend on T but only on the loop filter $F(z)$, keeping constant the other blocks. Therefore, under these hypothesis, the product $B_L \cdot T$ only depends on $F(z)$. The global performance of the PLL depends both on $F(z)$, or in general, $H(z)$, and the integration time T.

In terms of noise, the larger T the longer the noise is averaged by the integrators, therefore the more the noise is filtered out and the less it affects the estimate. Then, in order to filter out more the noise, the parameter B_L needs to be as small as possible.

This analysis that considers separately the effects of the integration time on one side and the loop filter on the other side can be done looking at Figure 6.2, where the part of the loop that works at the signal sampling time is well distinct from the part of the loop after the integration, where the working frequency is equal to $1/T$ (updating frequency of the local oscillator). The global performance of the loop depends of course on the combination of the two parts.

Limits on the loop design exist as rule-of-thumbs on the parameter $B_L \cdot T$, but due to the reasons explained above, loops with same $B_L \cdot T$ are not equivalent. For example, considering $B_L \cdot T = 20 \cdot 10^{-3}$, it can be taken, for example, $B_L = 20Hz$ and $T = 1ms$ or another loop having $B_L = 10Hz$ and $T = 2ms$. The second loop filters out the noise much better, thanks to a smaller bandwidth B_L and thanks to a longer integration time. On the opposite, this good performance towards the noise are paid with less capability in tracking the signal dynamics.

A good design rule says that the product $B_L \cdot T$ should be in the order of $1 \cdot 10^{-2}$. In this thesis the analysis is done keeping constant the integration time T to 1 ms, so that the focus can be posed on B_L, thus on the design of the loop filter $F(z)$, while all the other blocks are kept constant.

Therefore, fixing everything except only for the parameter B_L, it can be stated that the smaller the bandwidth B_L is, the more the noise is filtered out but the smaller the loop capability to track the signal dynamics is. On the other side, a wider bandwidth makes the loop faster (faster transient time, shorter transient), but it allows more noise to enter the system and to affect the estimate. In Figure 6.3 and Figure 6.4, the behavior of the loop respect to the input noise and the dynamics is shown, for different values of ENB.

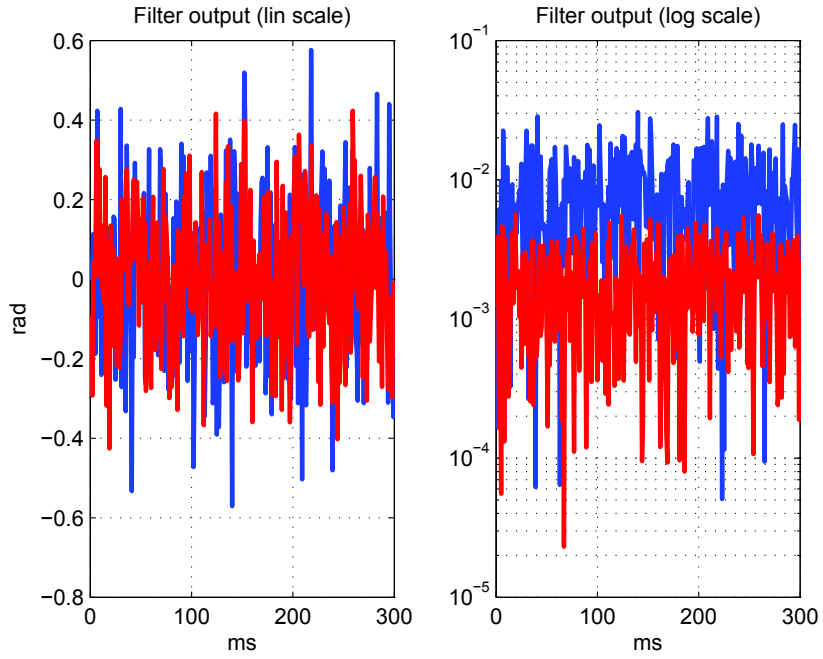

Figure 6.3: Behavior of PLL with different ENB, in noise condition.

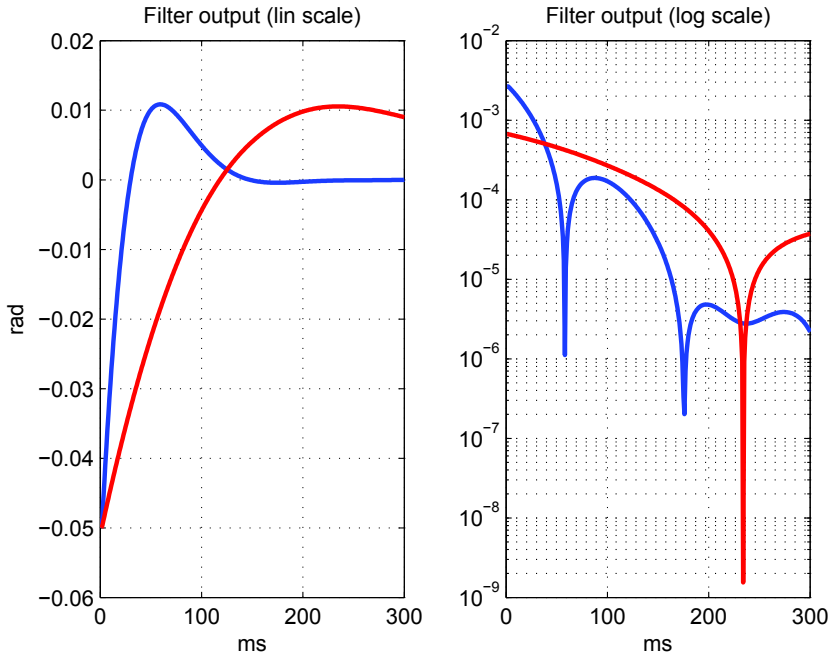

Figure 6.4: Behavior of PLL with different ENB, in dynamics condition.

From the considerations done, and from Figure 6.3 and Figure 6.4, it is evident that a trade-off is needed in the choice of B_L, which means to accept a trade-off in terms of performance towards the dynamics and the noise. The aim of the analysis described in Chapter 8 is to investigate how this trade-off can be reached, adapting the value of B_L.

6.7.1 PLL order

Depending on the number of poles in the closed loop transfer function, the PLL order is defined. The loop order is a crucial aspect of a PLL, since the ability to track or not a signal with a certain dynamics depends on it.
After a transient time, a steady state condition should be reached. If the loop order is suitable to track the input dynamics, the mean SSE is zero, otherwise a SSE is be present. In the case when it is present, the SSE can be decreased by increasing the loop bandwidth B_L. Details about this are available in Appendix B.
Note that, at times, a SSE can be acceptable, if the residual error is below a desired threshold. If the SSE is smaller than the tracking jitter, it results to be buried under the noise level. Moreover the error due to a not sufficient order of the PLL can be even smaller than some residual tracking error due to other effects such as multipath, see 2.1.7. In these cases, a residual SSE due to the PLL can be fairly acceptable.
As described in Appendix B, higher order loops can track input dynamics of higher order. The price of this capability is paid with higher computational effort and stability difficulties.
As stated at the beginning of this Section, the loop order indicates the degree of the characteristic polynomial, i.e. the number of poles of the closed loop transfer function $H(z)$. It has to be pointed out that an important property of a loop is the loop type, which is commonly used within the control system community. The type refers to the number of integrators in the loop. In the PLL considered here, one integrator is in the NCO and a variable number of integrators is present in the loop filter $F(z)$. In this specific case, the order and the type of the loop are the same, since each integrator represents one pole to the transfer function, and no other poles exist. In general, the order can never be less than the type, but additional non-integrator filtering can be present, contributing additional poles and therefore increasing the order, but not affecting the type.

This is the case considered in this thesis, where the PLL order depends only on $F(z)$, and can be defined as $N = M + 1$, where M is the order of $F(z)$. Other architectures exist, where for example some delays are introduced, to take into account hardware delays. This is the case, as instance, of PLLs designed to be implemented on FPGA.
To summarize, it can be said that a PLL of order N PLL can track with zero residual mean error an input function of order $P = N - 1$. This means that a PLL of first order can track a constant phase. If the input order is $P > N$, then the tracking error does not converge. If the input is of order $P = N$, the loop can track the signal with a constant error, which is conversely proportional to the ENB, as discussed in 6.7. This is the case for example when a first order PLL has to track a signal with linear phase, which means constant frequency. Details are provided in Appendix B.
In the following, an overview is given of DPLLs of different order (first, second and third respectively).

First order DPLL

Since the NCO works as a integrator, and it introduces a pole in the closed loop transfer function, a PLL cannot be of type (and order) minus that one. A first order loop filter has the form:

$$F(z) = a = \text{constant} \tag{6.27}$$

It can be demonstrate (see Appendix B) that with this loop filter, the PLL can lock the input phase with SSE = 0 only if it varies with a step change. This case is not very realistic, since constant or step phase implies null frequency.

Second order DPLL

A more realistic case is to consider a linear input phase, which means to have a signal with constant frequency. In order to track this signal with zero SSE, at least a second order PLL is needed (see Appendix B). The DPLL can be in this case naturally related to the analogic one, finding a relation between the loop coefficients and the analogic characteristic parameters (see [67]) . The loop filter

transfer function can be expressed as

$$F(z) = g_0 + g_1 \frac{z}{z-1} \tag{6.28}$$

$$= \frac{g_0 (z-1) + g_1}{z-1} \tag{6.29}$$

If it is set:

$$\begin{cases} a = & g_0 \\ b = & g_1 - g_0 \end{cases} \tag{6.30}$$

the loop trasfer function assumes the form:

$$F(z) = \frac{az - b}{z-1} \tag{6.31}$$

The analogical system can be described by two parameters in particular: the natural frequency ω_n and the damping factor ζ. Since these parameters are suitable to describe a pole pair they can be useful to set the loop coefficients a,b, as detailed in Appendix B.
A second order PLL can lock the input phase even if it changes linearly, i.e. when the input signal has a constant frequency or a frequency step change(see Appendix B).

Third order DPLL

In order to track an input with linear frequency, which means constant Doppler rate and phase that varies quadratically, a third order loop is needed. In the case of a linear Doppler rate, a third order PLL still allows to track the signal, but with a constant, non-zero, SSE. The loop filter in this case is:

$$F(z) = g_0 + g_1 \frac{z}{z-1} + g_2 \frac{z^2}{(z-1)^2} \tag{6.32}$$

$$= \frac{g_0 (z-1)^2 + g_1 z (z-1) + g_2 z^2}{(z-1)^2} \tag{6.33}$$

that, setting

$$\begin{cases} a = & g_0 + g_1 + g_2 \\ b = & -2g_0 - g_1 \\ c = & g_0 - g_1 \end{cases} \tag{6.34}$$

becomes simply

$$F(z) = \frac{az^2 + bz + c}{(z-1)^2} \tag{6.35}$$

Three indipendent parameters are required to characterize a third order DPLL. It is not trivial to set these parameters as a function of the desired ENB, since the stability requirements have to be satisfied at the same time. In Appendix A some details are provided. The method used in this thesis to compute the third order PLL parameters as a function of the ENB is taken from [68].

6.8 Tracking capability

The tracking capability of a PLL can be described in terms of some parameters, indicating a range of input frequencies:

- Hold-in range, the frequency interval within which the loop can track the correct frequency without cycle slips;
- Lock-in range, over which the loop is able to track the phase without cycle slips;
- Pull-in range, the interval within which the PLL can track the phase, even if with some initial cycle slips;
- Pull-out range, the maximum value of frequency that, applied to a PLL perfectly locked, can be tracked without any cycle slip.

These ranges depends on the order of the PLL and in particular on its ENB.

6.9 Loop parameters design

In this thesis the focus is on the design of the loop filter $F(z)$. Keeping constant the other blocks in the loop, the DPLL characteristics can be changed by tuning the order of the loop filter and the loop coefficients.
The goal is to design the loop coefficients, given the loop order. If M coefficients need to be designed, M conditions have to be fixed. A crucial condition that has to be fixed is the loop bandwidth B_L. Given that B_L is related to the loop coefficients through equation 6.26, which contains a circular integration, since

$H(z)$ is a function of the loop filter coefficients, the expression of the coefficients as function of the ENB is not trivial. Anyway, several approximations have been developed ([25], [67], [68]) to find analytical forms to express the coefficients. Furthermore, since one condition is not enough for PLLs of order > 1, other conditions have to be fixed, for example the damping factor for a second order loop. Note that in any case, the stability condition has to be satisfied.
In appendix A the methods applied in this thesis to derive the loop coefficients given a value of B_L, and other fixed conditions, are described.

Chapter 7

Adaptive Bandwidth Algorithm

7.1 Objectives

The aim of the work is to design an adaptive PLL, in order to optimize the performance in terms of tracking jitter, depending on the input dynamics characteristics.
The interest is on the GNSS velocity solution, therefore the scenario in which the PLL needs to be used is a dynamic one.
As described in Section 6.3, in GNSS applications the phase of the input signal depends on the Doppler frequency and on its rate, which in turn depends on the relative motion between the user and the signal source, onboard the GNSS satellites. Furthermore, the user can move too, therefore the Doppler profile can vary randomly.
Given this scenario, the aim of the work described in this thesis is to design an adaptive DPLL able to track signals with variable phase profile. Some cases will be considered with different frequency profiles.
Note that, in a realistic GNSS scenario, second and third order loops are sufficient to track the input signal.
The phase error components in a PLL are the noise jitter related to the input signal-to-noise ratio, and other two components related to the input signal dynamics.
The adaptive DPLL proposed here is designed with the objective to optimize the loop performance in terms of trade-off among all the error contributions. In fact, as stated in Chapter 6, the loop parameters differently impact on the several error components due to different components of the input signal. The narrower

B_L, the smaller the noise effect is, and, on the other side, the wider B_L is, the smaller the errors due to the input dynamics are.
To obtain such a trade-off, the ENB is dynamically changed according to the input signal variation.

7.2 State of the art overview

Different criterion to realize an adaptive loop are possible, as proposed as instance in [69], [70], [71], [72]. The adaptive optimization criterion proposed in this thesis is based on the construction of a cost function, as proposed in [73],[74] and in [75]. In [74] only the Steady State Error (SSE) was taken into account and in [75] only the transient deterministic error was considered, while here both the contributions are taken into account at the same time.
Two main elements of novelty are presented here. The first one consists in the fact that a general case is considered. An adaptive loop is designed for a generic input frequency. Furthermore a cost function is build that takes into account the steady state error, the transient error energy and the tracking jitter due to the noise in the same time, both for a second and a third order loop.
The second element of novelty is that some approximations are done in order to speed up the computations, to satisfy a real time requirement, which is a crucial feature for a GNSS receiver [76]. The goal is to develop an adaptive loop working in real time, to include it in a standard software receiver as N-Gene [24], the fully software GNSS receiver developed by the NavSAS group [77].
A preliminary work was done and is presented in (II). In this thesis it is shown the next algorithm developed, which is more general, as presented in (III).

7.3 Phase error contributions

As stated in Chapter 6 and in Section 7.1, in a PLL the phase error is constituted by two main contributions:

Noise contribution, the tracking jitter

Dynamics contribution:

Transient error

Steady State Error (SSE)

The noise contribution is directly proportional both to the signal-to-noise ratio at the receiver input and to B_L.
The dynamics effects are directly proportional to the input signal dynamics, but inversely proportional to B_L.
These three error contributions stated above are briefly described in the following.

7.3.1 Noise contribution

The noise contribution depends on B_L and on the input noise. Therefore, it is related to the tracking jitter measurement σ_ϕ, which can be expressed, as in [78], by the equation

$$\sigma_\phi = \sqrt{\frac{\sigma_{in}^2}{K^2}\, B_L T} \tag{7.1}$$

where σ_{in}^2 is the variance of the input noise, T is the integration time and the loop gain is $K = 1$.

7.3.2 Dynamics contribution

The transient error and the SSE are deterministic components which depend on the input signal dynamics, related to the Doppler frequency rate. As described in [79], the received carrier is a sinusoidal waveform with instantaneous phase

$$\phi_{RF}(t) = 2\pi f_{RF} t - \frac{2\pi D(t)}{\lambda_{RF}} + \varphi_{RF} \tag{7.2}$$

and frequency

$$f_{RF}(t) = \frac{1}{2\pi} \cdot \frac{d\phi_{RF}(t)}{dt} = f_{RF} + f_d(t) \tag{7.3}$$

where f_{RF} is the carrier frequency, $D(t)$ is the distance between the user and the satellite, and λ_{RF} is the wavelength.
The frequency variation in the time is the Doppler frequency, defined as

$$f_d(t) = -\frac{1}{\lambda_{RF}} \frac{dD(t)}{dt} \tag{7.4}$$

Since both the satellites and the user move, the input frequency and the input phase can assume different shapes and change a lot over the time. An example

is a user moving on a twisty mountain road or a Formula1 race track.
The Taylor expansion of the input phase can be considered:

$$\phi[k] = A_0 + A_1 k + A_2 k^2 + A_3 k^3 + \dots \tag{7.5}$$

from which also the input frequency can be written as a Taylor expansion, thanks to (7.3).

The SSE is due to the insufficient order of the loop and it is dependent on those terms of the expansion which cannot be tracked by the loop. In particular, the SSE is in general highly dependent on the term with the smallest order that the loop cannot track.
A second order PLL can track a linear phase with zero SSE and a parabolic phase with constant SSE. If the phase is a cubic function over the time, a second order PLL does not converge. A second order loop can be successfully used only if the input phase is a function in the time of second order or less, which means that it is parabolic or it can be at least well approximated with a parabolic function. Since a constant non-zero SSE depends in particular on the quadratic term of the phase expansion, here the input phase expansion truncated at the second order is considered for second order PLLs.
Similarly, in a third order loop the input phase expansion is truncated at the third order, since this loop can track a parabolic phase with zero SSE, but with a nonzero constant SSE if the phase is a cubic.
On the other side, the transient error does not depend on the order of the function which describes the input frequency, but it is related in particular to the initial error at the starting up of the DPLL or on step errors in the phase or in the frequency, which can be due to unexpected jumps in the local oscillator. A transient can occur also because of discontinuities in the signal derivative.
In the following the two deterministic error contributions are discussed more in detail, for a second and a third order loop respectively.

Steady State Error

At the steady state, the deterministic error is defined as the difference between the noiseless input phase and the noiseless loop estimate, which is the output of the phase detector in absence of noise. This error can be expressed, in the Zeta-domain, as

$$\rho(z) = \frac{z \cdot (z-1)}{(z-1) + \beta \cdot F(z)} \cdot \Delta_\phi(z) \tag{7.6}$$

where $\Delta_\phi(z)$ is the input phase increment in the Zeta-domain. Applying the final value theorem [66], the SSE can be written as

$$E_\infty[k] = \lim_{z \to +1} \rho(z) \cdot (z-1) \tag{7.7}$$

Second order loop As discussed at the beginning of this section, for a second order PLL the main contribution to the SSE is the quadratic term of the input phase, which depends on the coefficient A_2 of (7.5).
It is:

$$E_{\infty 2} = A_2 \left(\frac{9}{32(B_L T)^2} + 2 \right) \tag{7.8}$$

Third order loop For a third order PLL the SSE mainly depends on the cubic term of the Taylor expansion in (7.5).
Considering only the contribution of the cubic term, the phase increment is

$$\Delta_\phi(z) = A_3 \frac{z^2 + 4z + 1}{(z-1)^3} \tag{7.9}$$

and the SSE is

$$E_{\infty 3} = \frac{6A_3}{\beta(a+b+c)} \tag{7.10}$$

where a, b, c are the filter coefficients in (6.9).
It is possible to express the SSE as a function of B_L, because the filter coefficients depend on the ENB. The resulting equation is a complicated function of B_L, but a simple and very good approximation can be used, which gives:

$$E_\infty 3 \simeq 3 \cdot 2^3 \cdot A_3 \left(\frac{10}{(B_L T)^3} + 1 \right) \tag{7.11}$$

Transient error

The energy of the transient can be expressed as

$$E_t(z) = \frac{1}{2\pi j} \oint \rho_t(z) \cdot \rho_t\left(z^{-1}\right) \frac{dz}{z} \tag{7.12}$$

where $\rho_t(z) = \rho(z) - E_\infty(z)$ is the deterministic error without the SSE contribution.
The phase error in the Zeta-domain is $\rho(z)$ while $E_\infty(z)$ is the SSE computed in (7.8) and (7.11). Considering in (7.12) $\rho_t(z)$ instead of the total global deterministic error $\rho(z)$ gives a finite energy $E_t(z)$, even in the case when a constant

nonzero SSE is present.
Calling A_t the phase error at the starting up, or a step phase variation that may occur whenever, the transient error can be found by computing the circular integral in (7.12). The solution can be found applying to (7.12) the residue theorem ([66]).

Second order loop The energy of the transient computed from (7.12) is a complicated function of B_L, but it can be approximated by neglecting all the terms of the Taylor expansion in 7.5 apart one. For a second order loop, the transient error energy can be written as

$$E_{t2} \simeq \frac{1}{\sqrt{2}} \cdot (\frac{A_t}{B_L T})^2 \tag{7.13}$$

Third order loop Similarly to the case of a second order loop, for a third order loop the energy of the transient can be approximated with the expression:

$$E_{t3} \simeq \frac{1}{\sqrt{2}} \cdot \frac{A_t^2}{B_L T} \tag{7.14}$$

7.4 Estimate of the input noise and dynamics

In the real life, the input signal that comes at the GNSS receiver is unknown and the signal parameters must be estimated.
In particular, in order to compute the error contributions, it is necessary to evaluate:

- σ_{in}, i.e. the signal-to-noise ratio
- A_2, i.e. the Doppler frequency rate
- the error which causes a transient A_t

The evaluation of these parameters is not trivial, in particular in real-time.
The real-time requirement needs to be satisfied, since the problem that has to be addressed is to adapt the loop parameters in order to improve the input signal estimate, while the way to adapt the loop parameter is to estimate the characteristics of the input itself.
Furthermore the error contributions appear in the loop all mixed together.

For example, considering a user driving a car, if a peak is present at the phase detector output it is not possible to understand if it is due to a sudden signal-to-noise ratio decrease, for example due to a tree which represent an obstacle between the user and the satellite, or if it is due to a sudden acceleration of the user. In general in this situation a trade-off is however reached thanks to the techniques used to estimate the input parameters. This means that if a sudden signal variation is present, both the noise and the dynamics estimation increase, but the dynamics weight is much higher. This causes a bandwidth enlargement, which means that in such a situation the dynamics effects have heavier weight, so that the priority is to try and keep the locking condition. If odd behaviours are not present, a better trade-off is reached in terms of signal-to-noise ratio.
The adaptive technique proposed in this thesis needs the input signal parameters estimation in real-time, so for the estimation process only the present and the previous samples of the input signals are available. Therefore this solution gives good results only if the input changes are not very sudden. However, it is realistic to imagine that in the real life this is the case in most cases.

Signal-to-noise estimation

As stated in the previous, the objective is to implement the adaptive DPLL in the software receiver N-Gene [24]. Since N-Gene already computes a signal-to-noise ratio estimate, that estimate is exploited for the adaptive PLL algorithm too. Details about the algorithm used, the squared signal-to-noise variance estimator, are available in [80], where the technique to evaluate the C/N_0 ratio is presented.

Doppler rate estimation

If the PLL is in locking condition, the estimated frequency follows the input frequency.
If this condition is satisfied, the input frequency rate can be estimated by measuring the derivative of the NCO frequency. In order to mitigate the effects of the input noise, the NCO output is filtered with a moving average filter, whose length depends on the chosen update frequency of the adaptive algorithm.
With $T = 1$ ms, a good length has been proved to be $l = 0.5$ s. The frequency slope is estimated by computing the incremental ratio between two consecutive outputs of the filter.

Calling $\eta[k]$ the filter output, it is:

$$\frac{df[k]}{dk} \simeq \frac{\eta[k] - \eta[k-L]}{L} \tag{7.15}$$

Estimation of the parameter A_t

Since the lock is reached and maintained in the PLL by tuning the local frequency, each change in the input phase causes a change in the local frequency. Therefore a phase error A_t modifies the local frequency.
In general, in the real life, the third derivative of the frequency can be approximated with a null value, apart if discontinuities are present. Therefore the third derivative of the local frequency can give an estimate of the contribution to the transient error. According to (7.3), it is

$$\hat{A}_t = \frac{2\pi T^4}{L}\left(\frac{d^2 f[k]}{dk^2} - \frac{d^2 f[k-L]}{dk^2}\right) \tag{7.16}$$

7.5 Cost function

When the input parameters are estimated (as explained in Section 7.4) and can be considered as known parameters, the phase error contributions described in Section 7.3 results to be functions of only one parameter: the loop bandwidth B_L. In this way it is possible to build a cost function of as a weighted sum of the three components of the phase error, which is a function of the parameter $B_L T$:

$$f_C\left(B_L T\right) = w_N \cdot \sigma_\phi + w_t \cdot \sqrt{E_t} + w_\infty \cdot E_\infty \tag{7.17}$$

where w_N, w_t and w_∞ are weight coefficients. In particular it has to be noted that both the SSE and the phase standard deviation represent a phase error, while the term depending on the transient, E_t, is an energy.
The contribution of the transient in the cost function is a term which is proportional to $\sqrt{E_t}$ and it is weighted with a coefficient w_t, which takes into account that the transient energy is the sum of all the phase error samples over the transient time. In other words, the mean value of the transient energy during the transient time is considered.
Ideally, an instantaneous transient is desired, but it is not possible because of the finite bandwidth. Because of this, an interesting parameter of the transient is the transient time.

In this thesis, in order to apply the cost function approach, the transient phase error contribution has been considered.
In order to simplify the computation, a fixed value for w_t is chosen, simply considering the first 2 seconds of transient.
Fixing $w_\infty = 1$, the weight of the noise contribution is chosen to be $w_N = 2.6$ as explained in (II).

7.5.1 Second order loop

Thanks to the approximations applied to simplify the expression of the several error contributions, for a second order loop the cost function is

$$f_{C2}(B_L T) = W_N \cdot \sqrt{B_L T} + \frac{W_{t2}}{B_L T} + \frac{W_{\infty 2}}{(B_L T)^2} + C_{02} \tag{7.18}$$

where $W_{t2} = w_t \sqrt{\zeta} A_t$, $W_{\infty 2} = 9 w_\infty A_2 / 32$, $C_{02} = 2 w_\infty A_2$ and $W_N = w_N \sigma_{in} / K$.
The derivative of (7.18) with respect to the variable constituted by the product $B_L T$, can be easily computed.
In order to simplify the real-time computation, a further approximation is considered which can be valid in the used range of values of EBN. With the hypothesis of $B_L T$ in the order of $2 \cdot 10^{-}2$, the approximated cost function derivative is:

$$f'_{C2}(B_L T t) \simeq \frac{W_N}{2\sqrt{B_L T}} - \frac{2(10^{-2} W_{t2} + W_{\infty 2})}{(B_L T)^3} \tag{7.19}$$

7.5.2 Third order loop

For a third order loop the cost function is

$$f_{C3}\,(B_L T) = W_N \cdot \sqrt{B_L T} + \frac{W_{t3}}{\sqrt{B_L T}} + \frac{W_{\infty 3}}{(B_L T)^3} + C_{03} \tag{7.20}$$

where $W_{t3} = w_t \sqrt{\zeta} A_t$, $W_{\infty 3} = 15 \cdot 2^4 \cdot w_\infty A_3$, while the constant term is $C_{03} = 2 w_\infty A_2$.
Applying an approximation analogous to the one used for the second order loop to obtain (7.19), the derivative of (7.20) can be approximated, if $B_L T$ in the order of $2 \cdot 10^{-}2$, as:

$$f'_{C3}(B_L T) \simeq \frac{W_N}{2\sqrt{B_L T}} - \frac{3(10^{-5} W_{t3} + W_{\infty 3})}{(B_L T)^4} \tag{7.21}$$

Since the cost function is in some way a measurement of the loop phase estimation error, an optimization criterion could be to find the value of B_L which minimizes

the cost function. This minimum exists, as shown in (II), since the noise and the dynamics contribution are directly and inversely proportional to the bandwidth, respectively.

7.6 Optimization of the loop parameters

The criterion applied to optimize the loop is to find the filter coefficients which give the value of B_L that minimizes the cost function. Since both in the second and in the third order loop the function has only one minimum, as explained in (II) and the maximum is at ∞, the condition $f'_C(B_L) = 0$ is necessary and sufficient to find the cost function minimum.

7.6.1 Second order loop

Thanks to the approximations applied, it is possible to compute the derivative of (7.18) and its zero, which is found when the ENB is:

$$B_{OP2} = \frac{1}{T}\left[\frac{4(10^{-2}W_{t2} + W_{\infty 2})}{W_N}\right]^{2/5} \tag{7.22}$$

7.6.2 Third order loop

Similarly, the optimal bandwidth value is computed for a third order loop. It is:

$$B_{OP3} = \frac{1}{T}\left[\frac{6(10^{-5}W_{t3} + W_{\infty 3})}{W_N}\right]^{2/7} \tag{7.23}$$

Thanks to the expressions found for the loop bandwidth (Equation 7.6.1 and Equation 7.6.2), it is possible to compute in real-time the coefficients of the adaptive DPLL, simply by substituting the optimal value of B_L in the coefficient equations in (6.9).

Chapter 8

Adaptive Loop Analysis

8.1 Simulation results

In this chapter some simulation results are shown, to test the adaptive DPLL performance and comparing it with a standard DPLL with fixed bandwidth. Some results are shown for input signals with Doppler frequency that changes linearly or as a parabolic function. The simulated GPS data are generated with N-Fuels, a simulator of GNSS signals developed by the NavSAS group ([81], [77]).
In order to enable fair comparisons between different simulations, the simulation time and the initial frequency error are kept constant. The simulation time is fixed at 120 s and initial frequency error is 10 Hz. Furthermore, also the initialization value of B_L is fixed to a standard value, $B_{L_{init}} = 20$ Hz.
In order to make the system more robust, a minimum and a maximum allowed value for B_L are fixed, $B_{min} = 5$ Hz and $B_{MAX} = 50$ Hz respectively.
Two different levels of SNR are considered. A very common value, $C/N_0 = 45$ dB-Hz, and a high value, unrealistic in standard applications, $C/N_0 = 57$ dB-Hz, in order to show the behavior of the adaptive algorithm if the dynamics component is dominant over the noise effects.
Here the focus is on the PLL behavior. Future works are on the to-do-list, which include more tests with simulated and real data. Further test will be needed to measure the possible improvements on the velocity solution that could be achieved thanks to the enhanced carrier-phase estimated with the real-time adaptive DPLL developed here.

In the first simulation reported here the input frequency is linear, with Doppler

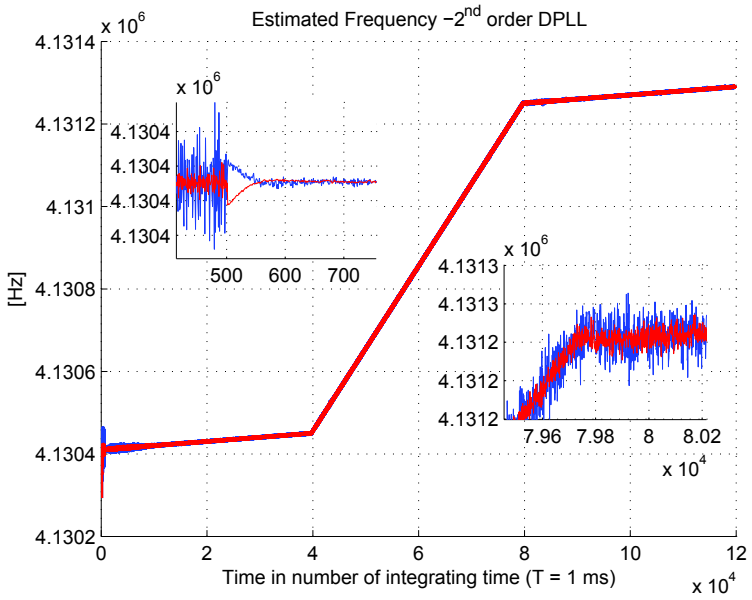

Figure 8.1: 2^{nd} order adaptive loop: NCO frequency with linear input Doppler frequency, when the Doppler rate changes. Frequency error at the starting up of 10 Hz, initial $B_L = 20$ Hz. The simulation result is shown for two signal-to-noise ratio.

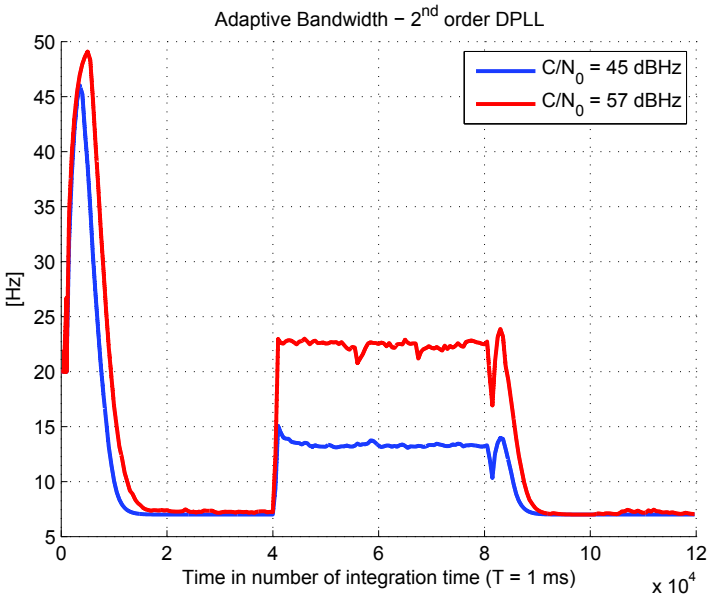

Figure 8.2: 2^{nd} order adaptive loop: PD output, when the NCO frequency is like that in Figure 8.1. The result obtained with the adaptive loop is compared to the result of a standard fixed bandwidth loop, when $C/N_0 = 57$

rate that changes during the simulation. At the beginning it is $D_{rate} = 1$ Hz/s, then after 40 s it becomes $D_{rate} = 10$ Hz/s and in the last 40 seconds of simulation it returns to $D_{rate} = 1$ Hz/s. Some results are shown of a second order adaptive DPLL.

Figure 8.1 shows the NCO frequency of the adaptive loop when $C/N_0 = 57$ dB-Hz and when $C/N_0 = 45$ dB-Hz, respectively. In Figure 8.3 a comparision in terms of discriminator output is done, between the adaptive loop and the standard DPLL with fixed bandwidth B_{Linit}. It is possible to see how in this case the

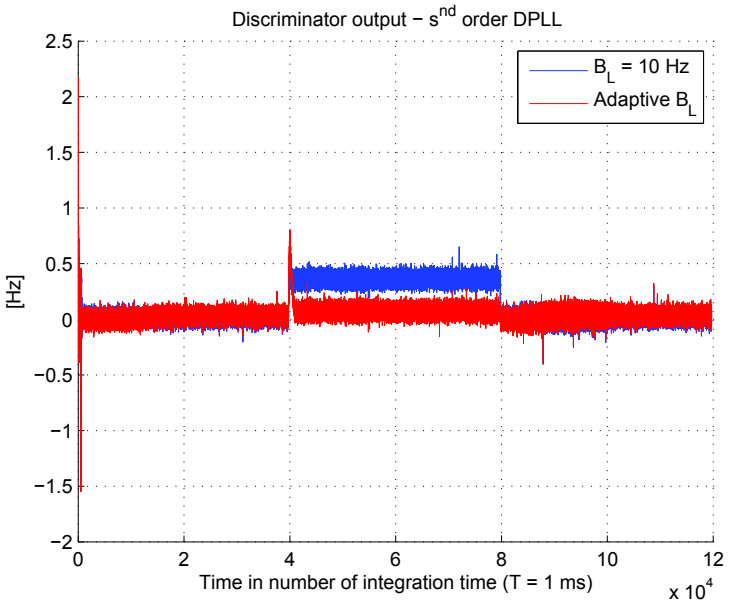

Figure 8.3: 2^{nd} order adaptive loop: adaptive bandwidth trend when the NCO frequency is like in Figure 8.1. The simulation result is shown for two signal-to-noise ratio.

adaptive algorithm allows obtaining a SSE very close to zero, while a fix value of ENB does not guarantee any trade-off between the noise and the dynamics requirements.

For example as shown in Figure 8.3 in this case a fixed bandwidth $B_L = 10$ Hz penalizes the performance towards the dynamics. In this case the signal-to-noise ratio is very high, therefore the algorithm trade-off can assist the dynamics requirement by enlarging the bandwidth by a substantial amount. On the other side, if the signal-to-noise ratio is lower, the trade-off does not allow very high values for B_L. However, at the output of the discriminator it is not visible a very large SSE, since the deterministic error is sunk in the noise which is high. B_L increases as an effect of the transient, then it increases again when the D_{rate}

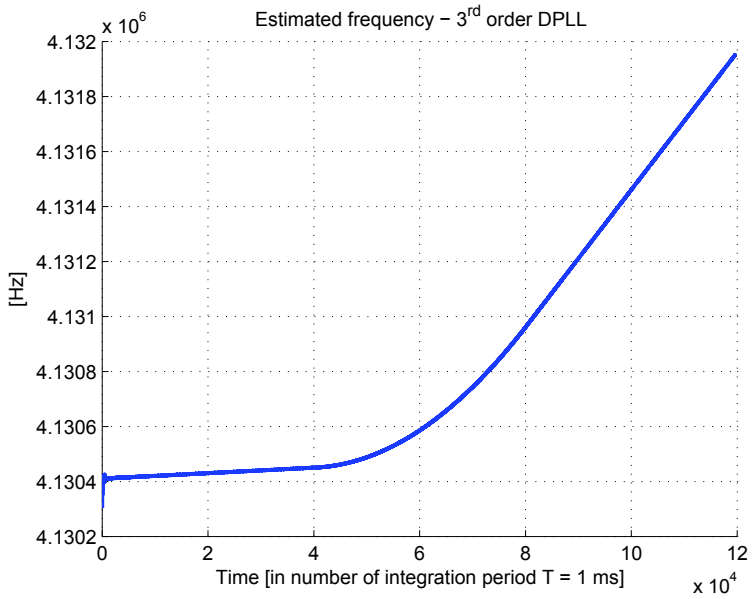

Figure 8.4: 3^{rd} order adaptive loop: NCO frequency with linear and parabolic input Doppler frequency. Frequency error at the starting up of 10 Hz, initial $B_L = 20$ Hz. The simulation result is shown for $C/N_0 = 57$ dB-Hz.

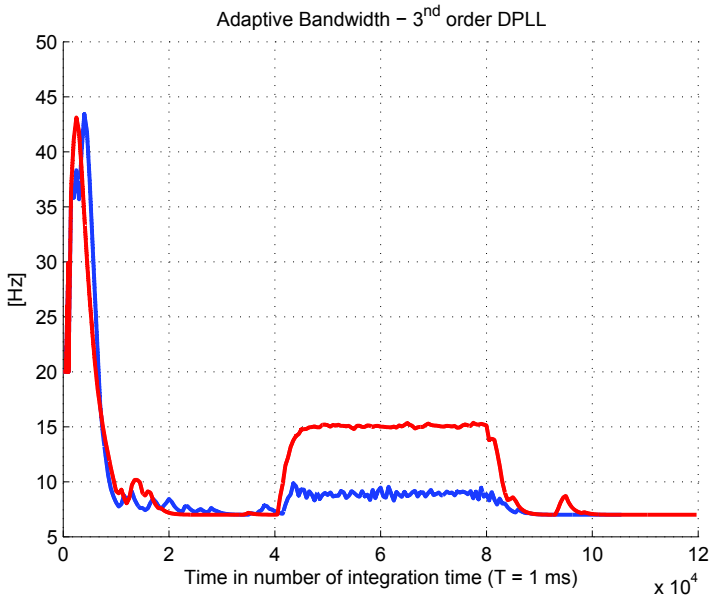

Figure 8.5: 3^{rd} order adaptive loop: adaptive bandwidth trend when the NCO frequency is like in Figure 8.4. The simulation result is shown for two signal-to-noise ratio.

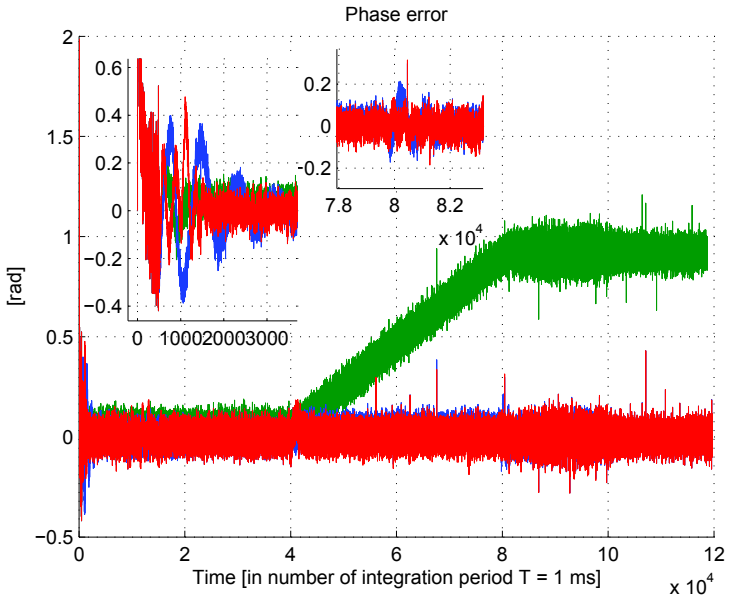

Figure 8.6: Comparison between standard 2^{nd} order DPLL and 3^{rd} order loop, with fixed and adaptive bandwidth. The simulation refers to the same input signal as in Figure 8.4, when $C/N_0 = 57$ dB-Hz.

enlarges, in order to decrease the SSE, then it returns to been smaller in the last part of the simulation. If, with the same dynamics, the signal-to-noise ratio is lower, the adaptive ENB is lower too, according to the trade-off rule between the dynamics and the noise requirements.
In Figure 8.2 the bandwidth trend is shown when the SNR assumes two different values, $C/N_0 = 57$ dB-Hz and $C/N_0 = 45$ dB-Hz respectively. The adaptive algorithm increases the bandwidth in order to supply the order insufficiency, since the SSE is inversely proportional to B_L.

A second simulation is considered with input frequency that in the initial 40 seconds is linear, with $D_{rate} = 1$ Hz/s, then it becomes parabolic for the next 40 s and in the last 40 s the Doppler rate returns to be constant,but higher ($D_{rate} = 10$ Hz/s).
The NCO frequency of a third order adaptive loop is shown in Figure 8.4.
The PLL tracks the input frequency with zero SSE when the frequency is linear and with constant residual SSE when the frequency is of second order. Therefore, when the frequency changes linearly in the linear parts the bandwidth converges, after a certain transient time, towards the inferior limit which has been fixed for the bandwidth. When the frequency follows a parabolic function, B_L converges

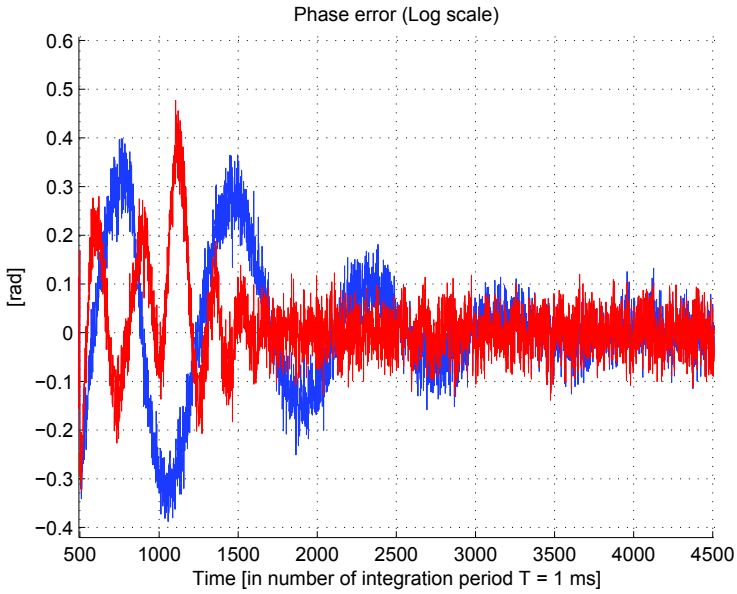

Figure 8.7: Zoom on the transient (discriminator output).

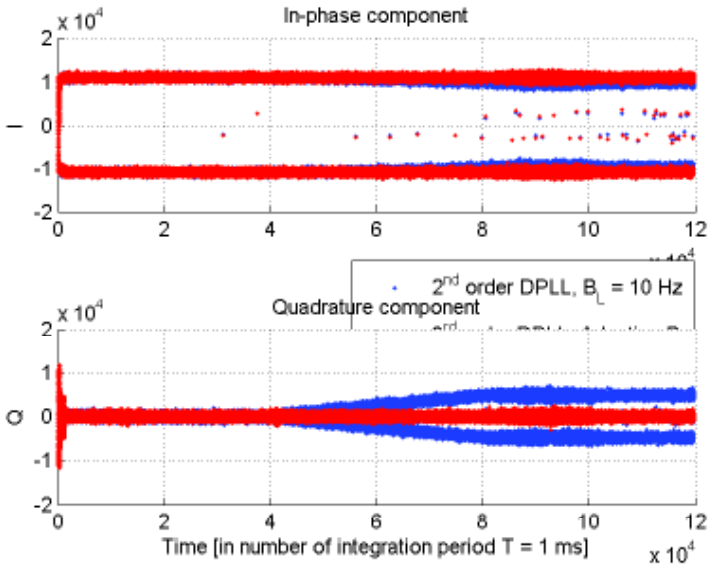

Figure 8.8: Comparison between standard 2^{nd} order DPLL and 3^{rd} order adaptive loop. The simulation refers to the same input signal as in Figure 8.7, when $C/N_0 = 57$ dB-Hz. The *Inphase* and *Quadrature* parts are shown.

towards a higher value, in order to supply the insufficient order, as Figure 8.5 shows.
The adapted values of the ENB are lower when C/N_0 is lower, as for the 2^{nd} order loop, to satisfy the trade-off requirement towards the noise performance.
Figure 8.7 shows a comparison between a standard loop of second order, a standard loop of third order, with fixed bandwidth, and a loop of third order with adaptive algorithm. The output of the phase detector is shown, too. It can be seen how a standard second order loop cannot track the signal, while the residual error in a third order loop converges towards zero.
If the input frequency is a parabola, the phase error in a second order loop diverges. Nevertheless, even in this case the adaptive algorithm can improve the situation by reducing the error divergency speed. However, since it is not possible to enlarge the bandwidth towards infinity, if the input frequency trend remains parabolic the second order loop will lose the lock, sooner or later.
In this case the adaptive third order loop is able to track the input carrier signal with error which converges towards zero, with the adaptive bandwidth changing as depicted in Figure 8.5 and in Figure 8.6 (B_L assumes real finite values).
Figure 8.8 shows the estimation error I and Q, in the case of a standard second order loop compared to the third order adaptive loop. From Figure 8.8 it can be seen how the adaptive third order loop has better performance, since the residual error converges to zero at the steady state.

8.2 Final remarks on the real-time adaptive DPLL

Adaptive DPLLs can have better performance with respect to the fixed bandwidth loops, but obviously the relative improvement amount depends both on the value that is chosen for B_L in the fixed bandwidth PLL and on the input signal characteristics. In general, the more the input signal is variable, the less a fixed bandwidth is suitable.
The adaptive real-time DPLL designed is therefore suitable to be used in dynamics environments, as the environments of interest for this thesis are. The aim in fact is to use the loop to improve the accuracy of the carrier-phase SD measurements, to be used to compute precise velocity solution, together with the precise velocity algorithm [2] and with the local frequency aiding, as described in Part II.

The real-time adaptive algorithm automatically changes the loop coefficients if the Doppler scenario changes, without any operation from the user. Thanks to this method, it is not necessary to know the input frequency shape to adapt the loop characteristics, because the input parameter estimation method is flexible and allows for an improvement in performance whatever the input is.

Part IV

Total Electron Content Measurements

Chapter 9

Earth Ionosphere

9.1 Earth atmosphere overview

The atmosphere of the Earth is a layer composed by gases held around the planet by its gravity field, until a height of about 1000 km.
Because of the gravity, the atmosphere is stratified. Generally, five layers are distinguished in the atmosphere, which include troposphere (6-20 km), stratosphere (20-50 km), mesosphere(50- 85 km), thermosphere (85-700 km) and exosphere (above 700 km), where particles are very rare and above which there is the cosmic void.
At heights lower than 1000 km the gravity is able to maintain an equilibrium, determining the layers thickness. At higher elevations the air becomes more and more ionized due to the cosmic ionizing radiations, in particular from the sun, that arrive less filtered from the upper layers. These radiations cause changing in the layers, i.e. spatial and temporal irregularities of the free ions distribution . Since neutral molecules have refractive index close to 1 while the ionized molecules have different index, such irregularities cause temporal and spatial gradients (the last ones are investigated in depth in [82]).

9.2 Ionosphere vertical profile

As said in Section 9.1, the particles of the atmosphere can be ionized and the layer of the atmosphere where the concentration of ionized molecules is higher is called ionosphere. The ionosphere is constituted by a plasma where the negative

particles (free electrons) and the positive ones (positive ions) exists separately. These particles are attracting each others by means of electrostatic forces, but do not unify to form a neutral molecule because their high energetic content does not allow it.

Figure 9.1 shows the atmosphere layers, including the ionospheric vertical profile.

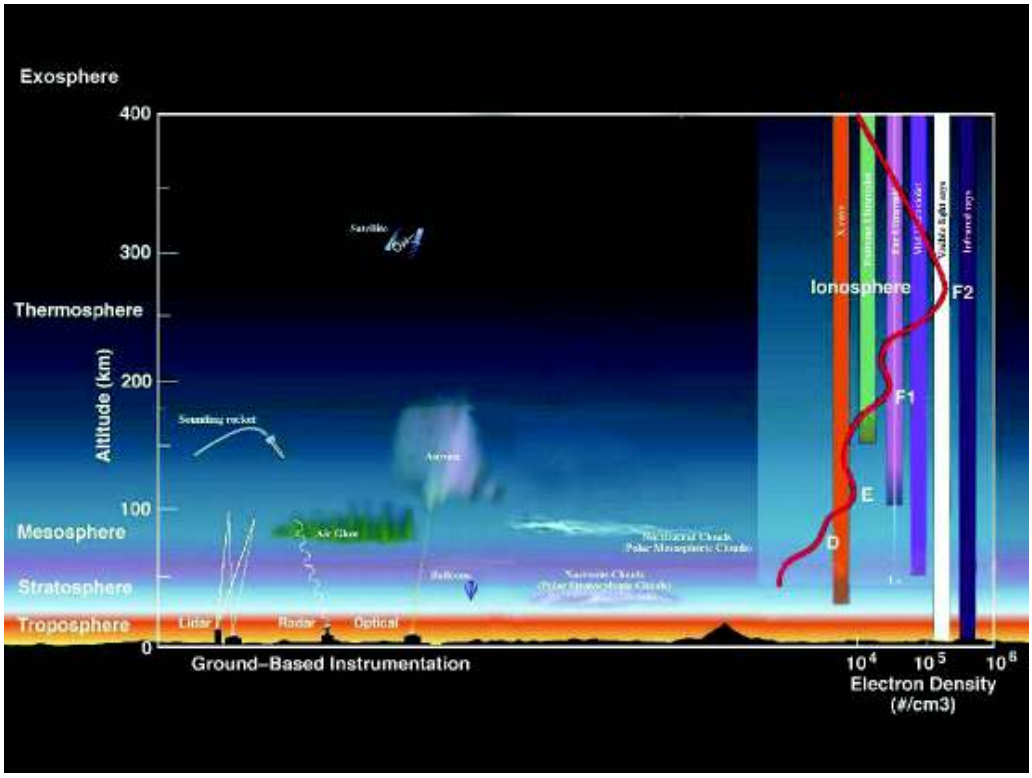

Figure 9.1: Layers of the Earth's atmosphere. The vertical profile of the ionosphere is shown too. Source online: NASA [9].

Ultra-violet rays and X-rays from the sun ionize the atmosphere and at heights higher than 1000 km the ionization is almost complete but, since at that heights the particle concentration is small, the ion and electron number is small as well. At lower elevation, corresponding to the so called ionosphere, the gas density is higher and as a consequence the radiations cause a high number of ionized molecules. At lower altitudes most of the radiation have already been absorbed from the higher layers, so the ionization is almost null. As a consequence, the ionosphere is usually localized in a zone which mainly coincide with the thermosphere, between 85 km and 600 km. The ionosphere thickness is variable and it

depends on the time of the day and of the year, on the location, and on other effects connected to those discontinuities outlined in Section 9.1.

9.2.1 Ionosphere electron density

The ionosphere can be mainly described by four parameters: electron density, electron temperature, ion temperature and ionic composition (different type of ions are present).
Since the main factor affecting the radio propagation is the electron density, the focus of this thesis is on this parameter.
The ionospheric electron density is defined as the number of free electrons per $1m^3$ and it can be measured using different techniques which involves electromagnetic waves, as described in Chapter 10 and in Chapter 11.

9.2.2 Ionospheric layers

Different constituents of the sun ionizing radiation give the maximum ionization at different height. This is because the absorbtion coefficient of the air is different for different constituent of the ionizing radiation, and as a consequence there are various layers of ionization in the ionosphere.
Usually, the ionosphere is divided in two main sub-layers, the E-layer and the F-layer. The E-layer is around 100 km, with ion and electron concentration which is about 10^{-11} of the concentration of neutral particles. The F-layer, at about 300 km of altitude, has a concentration of non-neutral particles which is about 10^{-5} times the concentration of the neutral ones, due to the fact that the neutral particles are more affected by the gravity field effects. In particular, the gravity effects acts on the medium hydrostatic configuration, and this is why the ionized layers are in general horizontally stratified ([83]).

9.2.3 Magnetic effects

Above 1000 km the ionization is total, therefore the medium has a very good electrical conductivity and as a consequence, the earth magnetic field has high influence. This part of the atmosphere is called magnetosphere, which in literature is sometimes indicated to be the highest part of the ionosphere or to be the

upper layer of the ionosphere.
The magnetosphere can be depicted as a blob of conducting fluid, which has inside a magnetic field (the earth magnetic field). The magnetosphere shape anyway is not static since it is influenced by external factors, in particular by the solar wind, that is a ionized plasma coming from the sun. Figure 9.2 depicts these effects. At heights above 1000 km, the atmosphere is not stratified

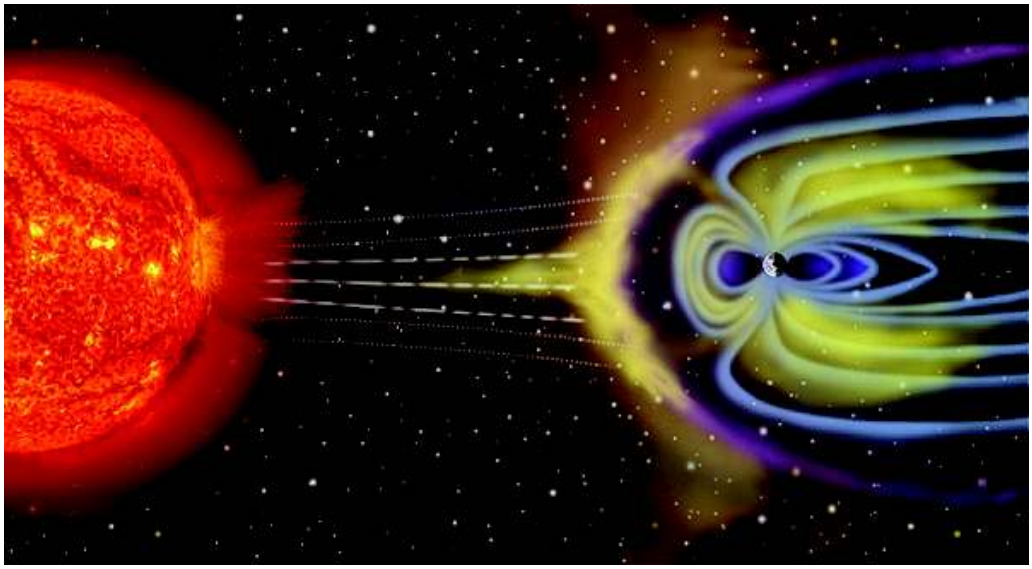

Figure 9.2: Effects of the sun on the Earth's ionosphere. Source online: NASA [9].

anymore since its conformation is not driven by the gravity but by the magnetic field. When the magnetic field in the medium changes, electric fields are induced as well, so currents flow, which generate magnetic field in turn. Also this electron movement causes changes in the ionosphere aspect.

9.3 Ionosphere effects on the GNSS signal propagation

9.3.1 Electromagnetic waves in the ionosphere

The ionosphere affects radio wave propagation in several ways, including attenuation, phase advances, group delays and polarization changes [83]. In the following, some effects are described.

9.3.2 Collision frequency

The attenuation effect that the ionosphere and the magnetosphere have on the electromagnetic waves is due to the collision between the waves and the ionospheric particles, which are not electrically neutral.
The collision frequency is proportional to the pressure since the higher the pressure, the higher the particle density is, which means to have higher probability that the particles are be ionized by an incoming radiation, and also higher probability that a wave collides with a particle.
In general, the collision frequency can be defined as:

$$\nu = \nu_0 exp(-\frac{z}{H}) \tag{9.1}$$

where z is the height and H is the scale height that can be taken equal to 1.5, while ν_0 is a constant (as in [83]).
In isothermal atmosphere with constant composition, the collision frequency is directly proportional to the pressure. In (9.1), ν is defined as a function of an exponential, whose argument is the opposite of the height. This means that the higher the height is, the smaller the collision frequency is. This is perfectly consistent to the pressure idea explained above, since the higher the altitude is, the smaller the pressure is, due to weaker gravity field effects.

9.3.3 Electron density and height

A theory exists to describe the electron concentration as a function of the height. Two factors are involved, with opposite effect:

- gravity effect: the lower the height, the higher the particle density;
- sun radiation effect: the higher the height, the higher the radiation effect.

This issue was studied by Sidney Chapman ([84]) in the early 1930s. The result of the analysis is the so called *Chapman law*, which is not reported here because this level of detail is outside the interest of this work, gives as a result the electron concentration profiles shown in Figure 9.3

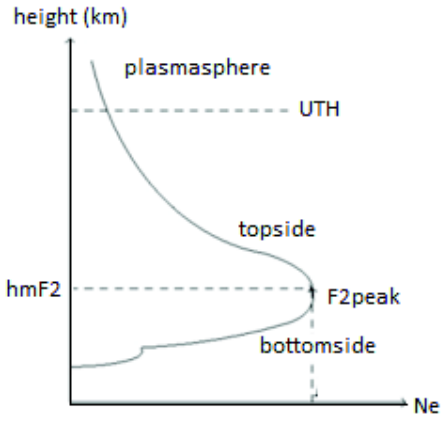

Figure 9.3: Chapman profile. Source online *South African Journal of Science* [10].

9.4 Scintillation

The irregularities in the electron density distribution act like wave scatterers, since irregularities create gradients in the refractive index. This produces ionospheric scintillation on the electromagnetic waves, in particular on the GNSS signals. As schematized in Figure 9.4, the scintillation causes unexpected changes in the signal, such as changes in amplitude and phase. Sometimes the changes are very abrupt, causing very low $\frac{C}{N_0}$ or very large changes in phase, that cannot be tracked from the carrier tracking stage (6). As a consequence, this effect can cause cycle slips or even complete lost of lock.

At present, the analysis of this phenomena in the ionosphere is becoming more and more a challenge (e.g. [85]), with the focus in particular on tracking the GNSS signal in presence of scintillation and modelling or even forecasting the scintillation events.

Ionosphere irregularities are not the focus of this thesis, but they shall be taken into account while doing the data analysis. This means that if the interest is on the measurement of the ionosphere in regular conditions (as in this thesis), before to analyze some real data the eventual presence of irregularities must be considered, in order to study only the data where the presence of scintillation must be excludable. In post processing and static conditions this control is quite easy to be done, checking the carrier-phase measurement and the signal-to-noise ratio continuity.

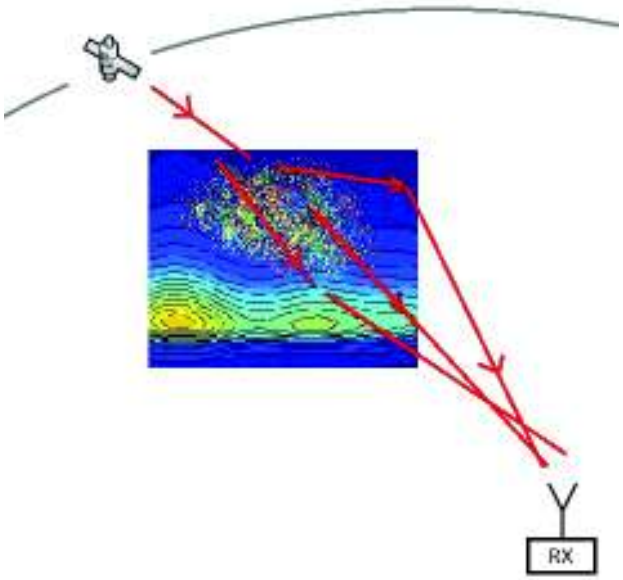

Figure 9.4: Simple representation of ionospheric scintillation.

9.5 Ionosphere models

Even if ionosphere behavior is affected by lots of factors, making any forecast very difficult, models exist that can shape the ionosphere with a quite good grade of reliability. The accuracy of these models is such that they can mitigate up to 70% of ionospheric delay. Some of them will be considered more in detail in Chapter 11.

Chapter 10

GNSS TEC Measurement and Related Error Sources

10.1 Total Electron Content

The effects of the ionosphere on the signal propagation represents one of the largest error sources in GNSS and become a not negligible issue in particular for those applications requiring high-accuracy positioning, which means centimeter or millimeter level accuracy. The ionosphere effect is particularly crucial for applications requiring highly accurate GNSS time transfer, where accuracies better than 10 ns are needed. Examples of applications include experiment synchronization over long baselines, synchronization of pseudolite-based positioning systems, bi-static radar, and multiplatform experiments, as well as remote sensing with satellites above the ionosphere. Furthermore, accurate GNSS-derived TEC measurements could also be used to calibrate (or verify the calibration) of space telescopes.

However, the ionosphere interaction with the GNSS signals can be seen under another perspective, in fact these effects offer a tool to study the ionosphere itself. Although the techniques presented in this thesis can be applied to GNSS in general, for simplicity the focus here is on GPS and all the GNSS measurements in this thesis refer only to GPS signal.

Because of the dispersive nature of the ionosphere, dual frequency GPS receivers enable a direct measurement of the delay introduced by the ionosphere on the signal, which is related to the Total Electron Content (TEC) along the path from the satellite to the receiver antenna.

The TEC in a certain location of the earth is defined as the total number of free electrons which are present in a cylinder of 1 m diameter along the vertical of that point, which are almost entirely contained in the ionosphere layer.
Usually, also in this thesis, with the term TEC it is intended the vertical TEC (vTEC), to be distinguished from the slant TEC (sTEC), which is the number of free electrons in a cylinder of 1 m diameter along the Line-Of-Sight (LOS) between the user and a specified satellite.
TEC measurements are important for several applications, first of all it is related to ionospheric correction needed to provide quite accurate GNSS solution. If high accurate GNSS solution is needed (in particular if highly accurate time-transfer is need), the TEC estimate must be highly accurate as well, and the aim of this thesis is exactly to address this issue.
Anyhow, a number of other applications exists, such as meteorological, climatological, and also of monitoring. For example, thanks to the ionosphere monitoring it could be possible to check the atmosphere polluting situation, to reveal events such as nuclear explosions, or even to forecast earthquakes [86].

10.1.1 Double frequency TEC measurement

The ionosphere is a dispersive medium, affecting on the electromagnetic wave passing through in different ways depending on the signal frequency. This effect enables dual-frequency measurements, which can be performed using the GPS system exploiting the L1 and L2 carrier frequencies (described in Section 1.4).
The delay introduced by the ionosphere is different on the two frequencies, which reflects on a difference in the pseudorange measurement that is proportional to the ionospheric TEC. A simple relationship between the pseudorange difference and the slant TEC exists, that is a good approximation for the sTEC:

$$sTEC(t) = k(PR_L1(t) - PR_L2(t)) \tag{10.1}$$

where

$$k = \frac{1}{40.3} \frac{f_L1^2 f_L2^2}{f_L2^2 - f_L1^2} \tag{10.2}$$

TEC is a function of a lot of variables, including changes (both on long and short term) in solar ionizing flux and magnetic activity. Accidental changes affect the TEC, as described in Section 9.4, anyhow it also largely depends on the time, both of day and of the year (i.e. the season), and on the user location. Note that the slant TEC also depend on the viewing direction.

10.2 Slant to Vertical TEC mapping function

Dual-frequency GNSS receivers measure the slant TEC, which is related to the position of the satellite used. The lower the satellite elevation is, the longer the signal path trough the ionosphere is, as shown in Figure 10.1.

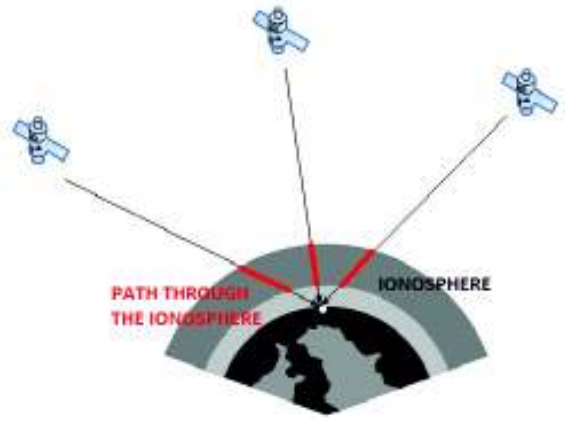

Figure 10.1: Effect of the satellite elevation on the ionospheric delay: the lower the satellite elevation, the longer the signal path through the ionosphere.

A user sees the GPS satellites at the zenith very rarely, especially if the user is located at high latitudes, therefore a mapping function is needed in order to convert slant into vertical TEC estimates.
The Obliquity Factor (OF) enables this conversion and it is derived from models of the ionospheric shell, as explained in [87]. The mapping function suppose that a model is valid for the ionospheric layer, assuming that the ionosphere is a thin shell around the earth, having uniform-density and uniformly located near the mean altitude of maximum TEC, in general considered to be at 350 km. Using this geometric model, as shown in Figure 10.2, a point of intersection can be determined between the user-satellite LoS and the ionospheric maximum layer at 350 km, called Ionospheric Pierce Point (IPP). The mapping function allows to convert the slant TEC into vertical, and viceversa, at the IPP.

In this paper a standard mapping function, largely used in literature as in [88], is considered:

$$OF(t) = \frac{1}{\sqrt{1 - (F \cos \Theta(t))^2}} \tag{10.3}$$

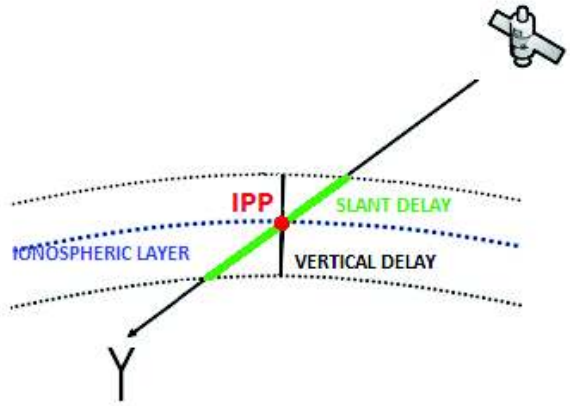

Figure 10.2: Ionospheric Pierce Point.

where $\Theta(t)$ is the satellite elevation angle and F is a fit parameter, which is set to 0.95 as in [88]. As an example, Figure 10.3 and Figure 10.4 show the sTEC measurement obtained with an Ashtec and a CRS receiver respectively, in blue , and the corresponding vTEC computed using (10.3), in red.

10.3 Error sources affecting double-frequency TEC measurement

In this section, the main error sources affecting double-frequency TEC measurement are discussed. In Chapter 2 the main error sources of GNSS are described. In particular the receiver hardware biases, that are negligible in the position solution, constitute a main error source for the TEC estimates and the time measurement, as introduced in Section 2.1.6.

10.3.1 Noise and multipath

Noise and multipath cause inaccurate and unprecise pseudorange measurements, as discussed in Chapter 2. While the noise is a disturbance that can be assumed to be zero-mean over time, the multipath is in general a biased effect. These error sources, that highly affect TEC measurements, will be treated in 11.2, where a method is applied to mitigated them.

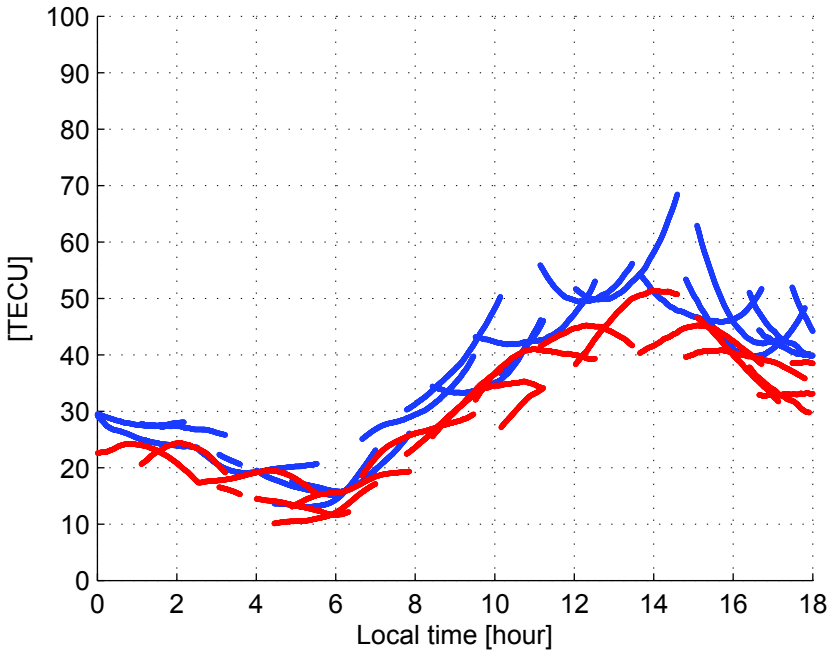

Figure 10.3: Vertical (red) vs Slant (blue) TEC measurement (Ashtec).

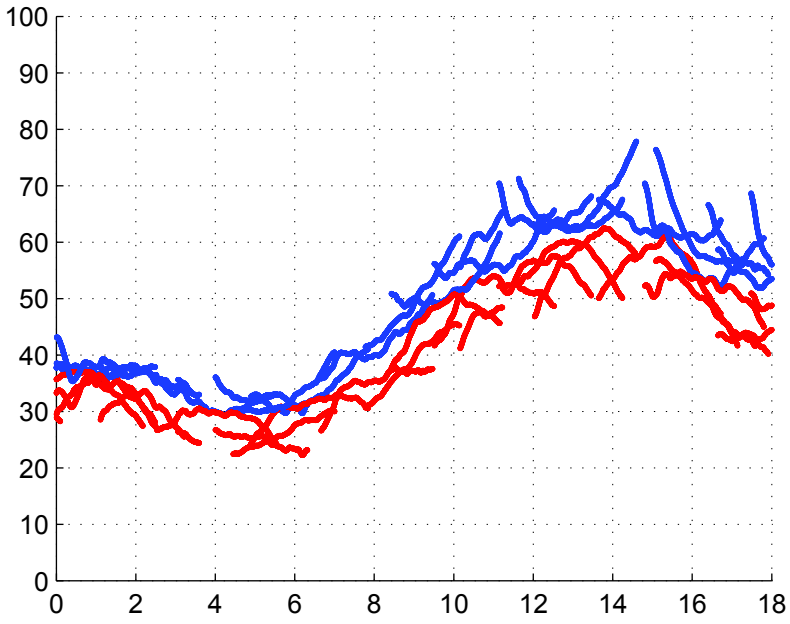

Figure 10.4: Vertical (red) vs Slant (blue) TEC measurement (CRS).

10.3.2 Hardware biases

The hardware biases described in 2.1.6 are mostly negligible when the interest is on the positioning solution, because the biases can be included in the receiver clock bias estimate.
Double-frequency measurements are affected by the different biases introduced by the hardware on different frequencies, and these errors affect the double-frequency TEC measurements.
As a consequence, if the interest is on the ionosphere TEC or the time estimate, these hardware biases must be taken into account, because they can become the limiting factor of the measurement accuracy. In Chapter 11, an analysis of these error sources is be done.

10.3.3 Higher-order ionospheric errors

First-order effects of the ionosphere can be estimated with the linear approach in (10.1). Due to the anisotropic properties of the layer, higher-order effects are also present which cannot be estimated with a linear approximation. However, the higher-order effects were found to be less than 1% of the first-order term, as demonstrated in [89] and in [90]. In [89] it is shown that second-order errors affect TEC measurements with biases lower than 0.2 TECU, which translates to less than 3 cm at the GPS L1 frequency.

10.4 Models of ionospheric delay

Since the ionosphere is affected by several phenomena, such as solar activity and earth magnetic field, it is very difficult to accurately model and forecast the behavior of this layer. Existing models, as Klobuchar [91] and NeQuick [92], of which a brief overview is given in the following, enable up to 70% mitigation of the ionospheric delays.
As detailed in [4], the ionospheric structure is the result of a very complex process. In this thesis, the focus is on the results obtained from measurements and not from the models.

10.4.1 Klobuchar

The so called Klobuchar model [91] provides an estimate of the delay caused on the GPS signal by the ionosphere. The estimate is available for different locations all over the earth at different times of the year and the day. It is interesting to notice that the Klobuchar model has been developed in 1987, which means before the operativity of GPS.
The model is designed to provide information about the iono-delay to those users that do not have a double-frequency receiver and therefore cannot find the double-frequency iono-correction. For those users an algorithm was designed, which uses eight coefficients that are transmitted in the navigation message.
Only eight coefficients are transmitted, even if a higher number of coefficients would allow higher accuracy in the correction. A trade-off between performance and complexity was taken into account when the method was developed, and the correction of only 50% of the ionospheric delays guaranteed by this model is a consequence of a decision in terms of trade-off. At the time when the method was designed, the knowledge of the ionosphere would have allowed corrections of 70-80% of the ionospheric delay, but due to the complexity requirement it was chosen to transmit to the user only a lower number of coefficients.
As described in [91], the method is based on an algorithm which had to take into account several requirements, therefore it is the result of a trade-off between them. The main requirements that it satisfies are:

- user computational complexity and computer storage
- state of the art of the ionospheric models (i.e. knowledge of TEC variation during the day and the year, and with the location)
- number of coefficients broadcasted in the navigation message
- geographic areas of users operating with single-frequency systems.

The Klobuchar model is based on an empirical approach. The assumption is that the electron content is concentrated in a thin layer, at 350 kilometres from the Earth's surface. With this hypothesis, the ionospheric delay along the line of sight (related to the STEC) is computed multiplying by the OF the vertical delay, at the IPP (Ionospheric Pierce Point).

Klobuchar algorithm

The Klobuchar method is based on an algorithm made of the following steps [11]. The equation are not described in details since the models are not the focus of the analysis done in this thesis.

- Calculate earth-centred angle;
- Find IPP characteristics, by computing:
  - latitude;
  - longitude;
  - geomagnetic latitude;
  - local time at the IPP.
- Find ionospheric delay characteristics, by computing:
  - amplitude;
  - period;
  - phase.
- Compute slant factor;
- Compute ionospheric time delay.

10.4.2 Nequick

The NeQuick [92] is the ionospheric model that has been proposed for the Galileo satellite system, to provide the ionospheric correction to single-frequency users, as well as the Klobuchar model for the GPS users. The Nequick model is able to guarantee more accurate corrections with respect to the Klobuchar method, correcting at least 70% of the ionospheric errors.
The NeQuick model is based on the EIL (Effective Ionisation Level) depending on the solar flux, which can be represented by the Az parameter

$$Az = a_0 + a_1\mu + a_2\mu^2 \tag{10.4}$$

where μ is the so called MODIP (MOdified DIP latitude), which is such that:

$$tan(\mu) = \frac{I}{\sqrt{cos\Phi}} \tag{10.5}$$

where I is the true magnetic inclination (0° at the magnetic equator and 90° at the magnetic poles) and Φ represents the receiver geographic latitude. The parameters a_0, a_1,a_2 are the three coefficients that have to be included in the Galileo navigation message to provide the ionospheric correction to single-frequency users. The coefficients updating rate is at least daily.

The reliability of the model depends on several factors, in particular, on unpredictable events such as scintillation and continuous monitoring performed by the monitoring stations allows to be aware of situation that might be problematic. It is very important to be able to alert the users if the corrections provided in the navigation message do not guarantee the specified performance. In these situations, disturbance flags are raised, transmitted continuously and updated at the same updating rate of the Navigation Message, i.e. every 100 minutes.

Since the ionospheric disturbance occur in localized areas, it is useful to consider different zone on the Earth so that a flag can be raised only on specified areas, allowing the users in other regions to use the corrections normally. The Nequick model makes use of five regions, as shown on the map in Figure 10.5.

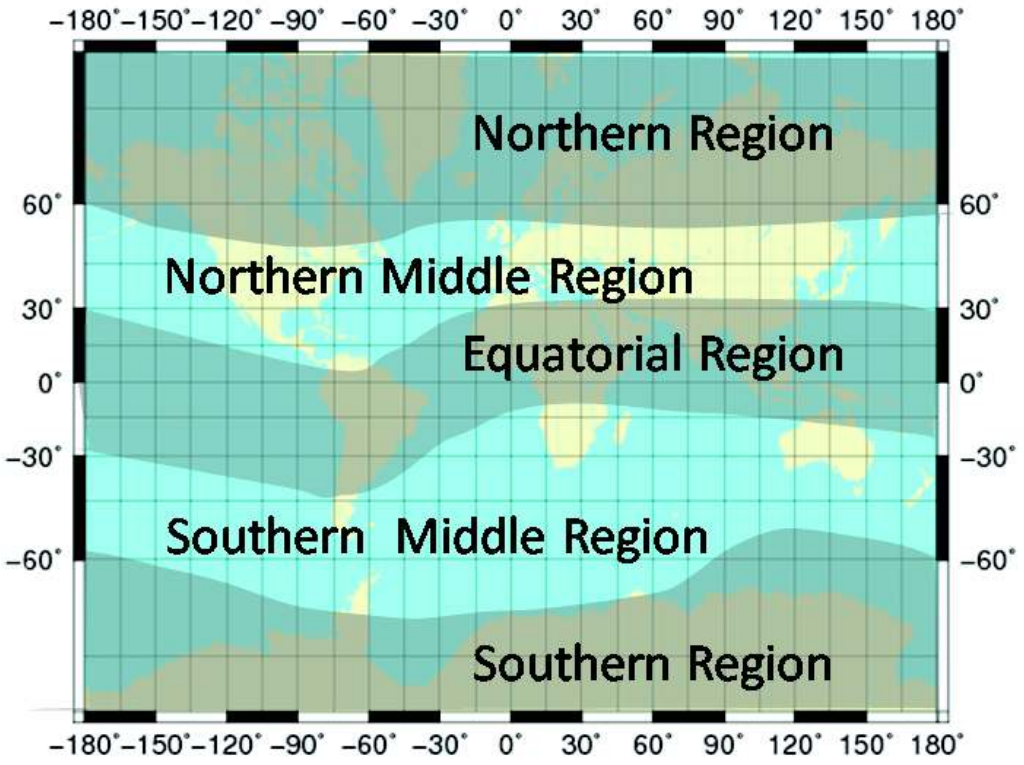

Figure 10.5: Nequick 5 zones. Source online: ESA navipedia, *NeQuick Ionospheric Model* [11].