Jari Nurmi · Elena-Simona Lohan
Henk Wymeersch · Gonzalo Seco-Granados
Ossi Nykänen *Editors*

Multi-Technology Positioning

Springer

Multi-Technology Positioning

Jari Nurmi • Elena-Simona Lohan
Henk Wymeersch • Gonzalo Seco-Granados
Ossi Nykänen
Editors

Multi-Technology Positioning

Springer

Editors
Jari Nurmi
Tampere University of Technology
Tampere, Finland

Henk Wymeersch
Chalmers University of Technology
Gothenburg, Sweden

Ossi Nykänen
Tampere University of Technology
Tampere, Finland

Elena-Simona Lohan
Tampere University of Technology
Tampere, Finland

Gonzalo Seco-Granados
Universitat Autonoma de Barcelona
Barcelona, Barcelona, Spain

ISBN 978-3-319-50426-1 ISBN 978-3-319-50427-8 (eBook)
DOI 10.1007/978-3-319-50427-8

Library of Congress Control Number: 2017932537

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This book has emerged as a joint effort of the fellows and some of the supervisors in the Marie Curie Initial Training Network MULTI-POS, funded by EU FP7 grant number 316528. When we have started to plan the writing of this book in 2015, we knew this would be a challenging endeavour, not only because it was to be written principally by early stage researchers but also because it was aiming at giving the full landscape of what wireless positioning world entails, from physical layer to application layer, and at explaining the inherent links between the navigation and communication worlds, which are typically addressed separately and by different research communities.

Our book aims at addressing all different layers and aspects involved in wireless positioning, from physical layer to application layer, from outdoor to indoor and seamless positioning and from signal and system design to business models and location-based services. Each chapter is dedicated to one or a couple of these different issues, and each chapter aims at presenting the basic overview of the topic, the main challenges and the main existing solutions. The book chapters also aim at raising potential new research questions and pointing out which issues are still currently not solved in the addressed area.

The detailed goals of this book are:

- To make the positioning technology more accessible for people working in other disciplines
- To give an overview of multiple positioning techniques, providing insight on the pros, cons and challenges related to them
- To provide inspiration to the readers on how to utilize the positioning technology and location data available
- To explain some of the available tools existing in the literature to investigate the world of the location-based services
- To open new avenues in ways of thinking about the future joint communication and navigation worlds
- To offer solutions to deal with mobility challenges to fresh graduates interested in cross-sector cross-field international careers.

The target audience of the book includes the following groups:

- Young students and professionals embarking on an international career
- Academic and industrial members involved or interested in Marie Curie EU programmes
- EU policymakers involved in future definitions of research mobility programmes
- People without strong technical background who want to understand the principles of wireless navigation and positioning and the connections between the communication and navigation worlds
- Business developers, city council workers, entrepreneurs and innovators who want to augment their technical background in the location-based services, especially those related to emergency, environment and health.

These being said, we hope that by the current book, we will bridge some of the missing gaps in the existing literature and we will provide to the general audience a comprehensive book on the multi-technology positioning.

Tampere, Finland — Jari Nurmi
Tampere, Finland — Elena-Simona Lohan
Gothenburg, Sweden — Henk Wymeersch
Barcelona, Spain — Gonzalo Seco-Granados
Tampere, Finland — Ossi Nykänen

Acknowledgements

We would like to start by mentioning that most of the work leading to the creation of this book was financially supported by EU FP7 Marie Curie Initial Training Network MULTI-POS (multi-technology positioning professionals) under grant number 316528.

The editors would like to thank each and every one of the MULTI-POS fellows who participated in the book writing, for their tremendous work in providing timely inputs and for their enthusiasm in cooperating in such a book project that will hopefully benefit future generations of fellows and researchers.

The MULTI-POS fellows are listed in an alphabetical order (according to the last name) in here: Anahid Basiri, Luis Bausa Lopez, Ondrej Daniel, Nunzia Giorgia Ferrara, Pedro Figueiredo e Silva, Markus Fröhle, Andreas Gehrmann, Anna Kolomijeca, Maciej Jerzy Paśnikowski, Pekka Peltola, Paolo Pilleggi, Jussi Raasakka, Alejandro Rivero, Susanna Sanchez, Jan Sanroma, Arash Shahmansoori and Enik Shytermeja.

In particular, Pedro Figueiredo e Silva and Ondrej Daniel are warmly thanked also for their help with the LaTeX editing issues in creating the book.

We would also like to express our warm thanks to the following persons for providing valuable criticism and feedback on our book chapters: Simo Ali-Löytty, David Cuesta Frau, Armin Dammann, José A. Del Peral, Luca De Nardis, Fabio Dovis, Ignacio Fernández Hernández, David Gómez Casco, Jan Johansson, Ondrej Kotaba, Helena Leppäkoski, L. Srikar Muppirisetty, Robert Piché, Ronald Raulefs, Georg Rehm, Jesus Selva Vera, Joaquín Torres-Sospedra and Mikko Valkama. We thank also Hana Pelikanova for creating the MULTI-POS logo.

Contents

Chapter 1
Introduction and Book Structure

Elena-Simona Lohan, Henk Wymeersch, Ossi Nykänen, Jari Nurmi, and Gonzalo Seco-Granados

Wireless positioning and navigation area is a prevalent area embroidered in the majority of wireless communication devices and applications. The number of technologies supporting wireless navigation has been continuously increasing in the last decade, from those based on classical satellite navigation systems to technologies employing inertial sensors, Wireless Local Area Networks (WLAN) and cellular systems, and even ultra-sound and visible light systems.

The motivation to write such a book came mainly from the fact that the current literature is still missing compact overviews of the different multi-layered aspects of wireless positioning. For example, while many books have been dedicated to satellite navigation systems, especially to the well-known and widespread Navstar Global Positioning System (GPS) system [8, 11, 13], the books dealing in a compact and unified manner with the most recent positioning technologies, such as WLAN, cellular (from 2nd to 5th generation), Radio Frequency Identification Radio Frequency Identification (RFID), Ultra WideBand Ultra Wide Band (UWB), visible light and so on, are still hard to be found in the literature.

On the other hand, books dealing with non-GNSS positioning technologies are typically focused only on few aspects and few layers of the non-GNSS system, such as physical layer only or Location-Based Services (LBS) layer only. For example, in

E.-S. Lohan (✉) • O. Nykänen • J. Nurmi
Tampere University of Technology, Korkeakoulunkatu 10, 33720 Tampere, Finland
e-mail: elena-simona.lohan@tut.fi; ossi.nykanen@tut.fi; jari.nurmi@tut.fi

H. Wymeersch
Chalmers University of Technology, Department of Signals and Systems, 41296 Gothenburg, Sweden
e-mail: henkw@chalmers.se

G. Seco-Granados
Universitat Autonoma de Barcelona, Barcelona, Barcelona, Spain
e-mail: gonzalo.seco@uab.es

J. Nurmi et al. (eds.), *Multi-Technology Positioning*,
DOI 10.1007/978-3-319-50427-8_1

[4], the focus is on indoor non-satellite positioning systems, such as Bluetooth and other short-range wireless-positioning methods. The authors in [1] also present an overview of terrestrial radiolocation techniques, addressing also the assisted GPS and assisted GNSS concepts from physical layer point of view. The details about the standalone GNSS systems are not included, neither in [4] nor in [1]. The book [12] provides a unified treatment of issues related to all wireless access and wireless localization techniques. The focus is on physical layer principles of localization, medium access methods and terrestrial networks infrastructure design. Security aspects are also discussed. GNSS systems are again not addressed.

Similarly, there are currently only few books dedicated to LBS, no doubt due to the huge amount of existing and possible LBS, which makes the unified and compact views rather difficult to tackle. One example of such books is [9], which covered the latest research in LBS at the time of its writing (in 2011). By using the location and any additional information available on the user and the surroundings, the services can be made context-aware and thus improve the quality and usefulness of the service for the user. Location data can also enable new business models both for commercial and governmental operators, e.g. context-dependent advertisements, guidance, road tolls and location-based insurance fees.

The book addresses three distinct areas of wireless positioning, as follows: the first part after the introductory chapters (Chaps. 3–6) is dedicated to GNSS. The second part (Chaps. 7–10) deals with non-GNSS positioning systems. The third part of the book (Chaps. 11–15) focuses on LBS, applications and business models in wireless positioning and navigation. The book aims at covering the multi-disciplinary fields of wireless localization, ranging from applications and markets down to the data streams in navigation devices.

Chapters 3–6 follow the GNSS typical transmitter–receiver chain, starting from the signal models at the transmitter side and going through the channel impairments and receiver processing.

More specifically, Chap. 3 presents the basics of GNSS signals, focusing on their time-domain and frequency-domain characteristics and on the noise characterization.

Chapter 4 describes the main three error sources in GNSS, namely multipath, interference and ionospheric distortion, and gives an overview of some strategies used to mitigate these error sources. Error mitigation is an active research and development area in the positioning technologies sector. MULTI-POS has participated actively in different error sources mitigation. In 2015, a collaboration was established between Pildo Labs, Tampere University of Technology and GMV Aerospace and Defence to propose a robust GNSS receiver architecture, through a combination of innovative algorithms for mitigation of ionospheric error, multipath and interference [5]. Chapter 4, with the additional participation of École Nationale de l'Aviation Civile and Honeywell International, presents the mitigation techniques for this large error sources.

Chapter 5 focuses the advanced signal processing or GNSS/sensor integration approaches, in order to assure the navigation quality of service, especially in terms of accuracy and availability.

Chapter 6 describes the four GNSS currently operational or under current development and addresses various issues pertaining to multi-GNSS receivers.

Chapter 7 does the transition between GNSS and non-GNSS positioning systems, by describing various signals of opportunity and basic principles about how a seamless navigation can be obtained. In order to bridge the gap between lower technology layer and upper application layer, this chapter deals with the issues related to urban environments and in advancing the availability of the lower layer navigation solution. The sensor and method review offered in Chap. 7 enables the researchers to focus and pick the most suitable sensors and methods for their specific purposes.

Chapter 8 goes into the details on non-GNSS positioning, by describing the fingerprinting and path-loss modeling concepts.

Chapter 9 introduces the main concepts of 5G positioning, overviewing the main techniques from 5G systems which can be used to locate the mobile terminal.

Chapter 10 addresses the issue of uncertainty in controlled multi-agent systems and introduces the main challenges and solutions related to cooperative positioning. This chapter emphasizes also the fact that the so-called cognitive methods for positioning and communication, where cognition refers to the property of the positioning and communication sub-systems, are not be considered as separate, but rather as closely interacting entities. Part of this work was performed in collaboration with the German Aerospace Center, and it is based on the published research papers [6, 7, 14].

Chapter 10 is dedicated to the applications of positioning and remote sensing to solve environmental challenges.

Chapter 11 explains how the user positioning information can be used to provide personalized services and which are some typical context-aware engines and architectures. Indeed, nowadays, there are many applications in the market that use the user position to offer a service, such as weather information providers. Chapter 11 discusses how to model user context is discussed, as well as how to acquire other type of information different than position, in order to enrich mobile services. In some cases, simple information such as user age or user language can be relevant to provide a personalized service and a better user experience. These other types of information which can be relevant to improve the context-aware personalized services are summarized in Chap. 11.

Chapter 12 discusses some business models regarding Galileo open service signals, and it assesses the financial aspects of Galileo, including costs, market share and potential revenues.

Chapter 13 presents another possible market segment of LBS, the e-health market segment, and it introduces the concepts of an Analytical Hierarchical Process Analytical Hierarchical Process (AHP) modeling in the context of e-health LBS. The proposal to use AHP in the context of wireless positioning originated from the team of MULTI-POS fellows at the University of Nottingham in 2013. Since then, an active collaboration between Tampere University of Technology and the University of Nottingham has taken place in this research area and several joint or individual results were published, for example, in [2, 3, 10].

Chapter 14 deals with the public safety applications of wireless positioning and more specifically with autonomous drones for emergency responses.

In addition, this book, created through the collective effort of the fellows and supervisors in a Marie Curie European network, is also talking about experiences within a Marie Curie ITN and it presents the lessons learnt by all network participants in Chap. 16. The discussions in this chapter could offer valuable ideas for those interested in pursuing a cross-sector career and/or to work in a different country than the home country.

The last book chapter, Chap. 17, presents the conclusions and points out the main findings.

This book is written in a tutorial style, presenting the challenges and raising the questions for the reader; no ready-made solutions are given, which, we hope, will motivate the reader to find his/her own path of addressing the unsolved challenges.

References

1. D. Bartlett, *Essentials of Positioning and Location Technology* (Cambridge University Press, Cambridge, 2013)
2. A. Basiri et al., Indoor positioning technology assessment using analytic hierarchy process for pedestrian navigation services, in *International Conference on Localization and GNSS (ICL-GNSS), 2015* (2015), pp. 1–6. doi:10.1109/ICL-GNSS.2015.7217157
3. A. Basiri et al., Overview of positioning technologies from fitness-to-purpose point of view, in *International Conference on Localization and GNSS (ICL-GNSS), 2014* (2015), pp. 1–7
4. A. Bensky, *Wireless Positioning Technologies and Applications (Technology and Applications)* (Artech House Publishers, Boston, 2008)
5. N.G. Ferrara et al., Combined architecture. Enhancing multi-dimensional signal quality in GNSS receivers, in *Inside GNSS Working Papers* (2016), pp. 54–62
6. M. Frohle et al., Multi-step sensor selection with position uncertainty constraints, in *2014 IEEE Globecom Workshops (GC Wkshps)* (2014), pp. 1439–1444. doi:10.1109/GLOCOMW.2014.7063636
7. M. Frohle, L.S. Muppirisetty, H. Wymeersch, Channel gain prediction for multi-agent networks in the presence of location uncertainty, in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2016)
8. E. Kaplan, *Understanding GPS - Principles and Applications*, 2nd edn. (Artech House, Boston, 2005)
9. Y. Liu, Z. Yang, *Location, Localization, and Localizability: Location-Awareness Technology for Wireless Networks* (Springer, New York, 2011)
10. E.S. Lohan et al., Analytic hierarchy process for assessing e-health technologies for elderly indoor mobility analysis, in *5th EAI/ACM International Conference on Wireless Mobile Communication and Healthcare - "Transforming Healthcare Through Innovations in Mobile and Wireless Technologies", Mobihealth 2015* (2015). doi:978-1-63190-088-4
11. J. Nurmi, E.S. Lohan, H. Hurskainen (eds.), *GALILEO Positioning Technology* (Springer, Dordrecht, 2014)
12. K. Pahlavan, P. Krishnamurthy, *Principles of Wireless Access and Localization* (Wiley, Chichester, 2013)
13. F. van Diggelen, *A-GPS: Assisted GPS, GNSS, and SBAS* (Artech House, Boston, 2009)
14. S. Zhang et al., Location-aware formation control in swarm navigation, in *IEEE Globecom Workshops* (2015)

Chapter 2
MULTI-POS: Multi-Technology Positioning Professionals Training Network

Elena-Simona Lohan, Henk Wymeersch, Ossi Nykänen, Jari Nurmi, and Gonzalo Seco-Granados

2.1 How Everything Has Started

During Fall 2011, when a group of senior researchers and professors at Tampere University of Technology started to think about putting together an EU training network in the field of localization, the wireless localization landscape had recently undergone a major paradigm shift, from GNSS global localization solutions to emerging non-GNSS indoor mass-market localization solutions. Indeed, several big players in the indoor localization market had already shown their interest into providing global indoor navigation solutions. For example, HERE started acquiring floor maps and site maps of several indoor locations in 2011, Apple was already using on its mobile devices the Skyhook solutions, and Google was already starting to map the indoor locations. Also, new companies focusing on indoor localization or RTLS were born, such as Meridian Apps or Sonic Notify Inc. It was clear that the wireless positioning was becoming an increasingly important part of future mobile devices.

E.-S. Lohan (✉) • O. Nykänen • J. Nurmi
Tampere University of Technology, Korkeakoulunkatu 10, 33720 Tampere, Finland
e-mail: elena-simona.lohan@tut.fi; ossi.nykanen@tut.fi; jari.nurmi@tut.fi

H. Wymeersch
Chalmers University of Technology, Department of Signals and Systems, 41296 Gothenburg, Sweden
e-mail: henkw@chalmers.se

G. Seco-Granados
Universitat Autonoma de Barcelona, Barcelona, Barcelona, Spain
e-mail: gonzalo.seco@uab.es

J. Nurmi et al. (eds.), *Multi-Technology Positioning*,
DOI 10.1007/978-3-319-50427-8_2

The European Commission (EC) statistics at that time showed that the global navigation market was expected to grow to around €244 billion EUR (incl. sales of GNSS-enabled products) in the next decade, with significant growth being driven by cellular mobiles and LBS.

The wireless positioning research world at that point was split into two large and distinct research communities: the technical community (focused on location technologies and techniques) and the business community (focused on understanding users and markets needs and on finding new LBS in order to enhance the social well-being). Moreover, the existing graduate schools in EU in 2011 related to wireless location focused heavily on GNSS (e.g., ESA International summer school on GNSS or the EU Initial Training Network TRANSMIT). Therefore, there was still a missing link between these two communities, the technical and business community, and the need for creating a graduate school to create expert workforce with good background in all the layers involved in a wireless positioning system was becoming urgent. Thus the idea of creating an international network bringing together experts from various fields of wireless localization was born. Indeed, experts from fields such as mathematics, statistics, signal processing, telecommunications, computer systems, machine learning, aerospace, economics and business administration were contacted and a global network has been created in order to address the gaps of the wireless positioning-related training at EU level. The created network, baptized Multi-technology Positioning Professionals, or MULTI-POS in brief, received the funding from the EU FP7 Marie Curie Initial Training Network (ITN) program (grant nr. 316528) and officially started in October 2012 with a kick-off in Belgium. The network fellows were recruited during the late 2012 and early 2013 and we have had a total of 12 Early Stage Researchers and 3 Experienced Researchers recruited from all over the world in our network. We were lucky to be able to recruit also four female fellows (27%), as vast majority of the competent resources available in the research fields consists of male researchers.

The ten partners of MULTI-POS include six universities and four companies (including two SME). In addition, there are ten associate partners, of which six are companies (five SME and one large enterprise), three research institutes, and a university (see Table 2.1).

2.2 MULTI-POS Aims and Structure

The MULTI-POS network has been built as a fully integrated research training program with a strong contribution of private sector partners, focusing on training Early Stage Researchers through doctoral programs and on offering complementary academic-industrial skills to all the recruited fellows. The network started to exist during Fall 2012, the fellows' recruitment took place mainly around the mid of 2013 and the network officially ended at the end of September 2016. The goals of MULTI-POS network have been:

Table 2.1 MULTI-POS network

Abbreviation	Organization name	Country	Organization type
Full partners			
TUT	Tampere University of Technology, Coordinator	FI	University
CUT	Chalmers University of Technology	SE	University
ENAC	Ecole Nationale de l'Aviation Civile	FR	University
GMV	GMV Aerospace and Defense	ES	Large Enterprise
HON	Honeywell International	CZ	Large Enterprise
PCG	Ptolemus	BE	SME
PLD	Pildo Labs Consulting	ES	SME
UAB	Universitat Autonoma de Barcelona	ES	University
UNOTT	University of Nottingham	UK	University
VU-VUMC	Stichting VU (Vrije Universiteit Amsterdam)	NL	University
Associated partners			
AVEA	AVEA Iletişim Hizmetleri	TR	Large Enterprise
DLR	German Aerospace Center	DE	Research Center
FGI	Finnish Geospatial Research Institute	FI	Research Center
FPPNT	Focal Point Positioning	UK	SME
GEO	Geodan	NL	SME
ITMO	National Research University of Information Technologies Mechanics and Optics	RU	University
SP	SP Technical Research Institute of Sweden	SE	Research Center
SSF	Space Systems Finland	FI	SME
T6ECO	T6 Ecosystems	IT	SME
WPS	Wirepas	FI	SME

- To bridge the gap between the lower technology layer and the upper application layer involved in wireless mobile positioning and navigation.
- To offer high-quality comprehensive training to young researchers in the broad field of the wireless localization and will promote the exchange of fellows in mixed academic-industrial R and D trajectories and in multiple European cultures, thus strengthening also the academic-industrial collaborations and the European collaboration.
- To ensure a superior scientific and engineering design expertise, combined with agile development of new kinds of applications in the area of wireless navigation.
- To create novel methods, technologies and business models for the future location-enabled wireless devices.
- To enhance EU competitiveness in multiple localization-related sectors, as its technology is clearly transferable to such multiple industry sectors.
- To initiate an educational and research framework that unifies the currently fragmented research activities on technological and applications aspects of wireless.

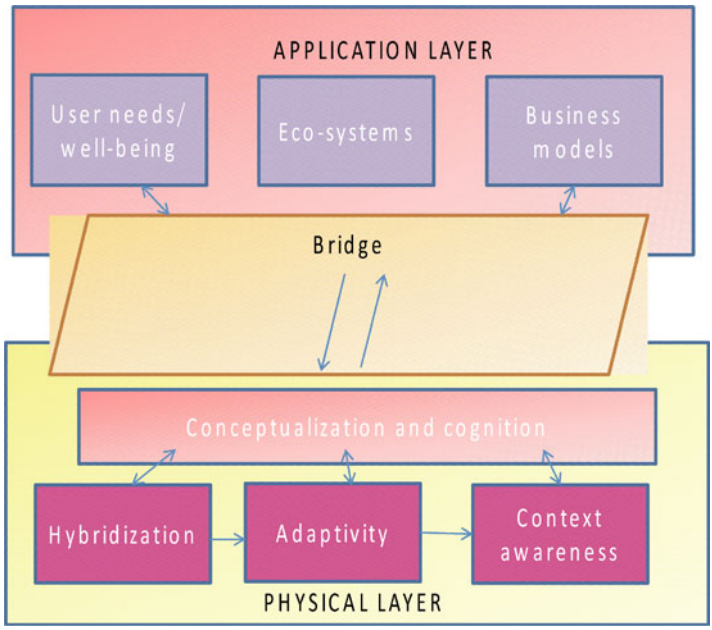

Fig. 2.1 MULTI-POS research overview at various layers involved in wireless positioning

- To create a community of researchers capable of exploiting synergies and complementarities between various fields involved in location and navigation, such as engineering, social sciences, and business world.
- To enhance the career prospects of young researchers by training them to use the tools from various research fields and to tackle the problems in a holistic manner.

The methodology to achieve these goals has been based on exposing the fellows to both technical and economic aspects, to both academic and industrial environments, and offering them first-hand experience and understanding of positioning needs through own mobility between several countries. The logic of the scientific research themes is illustrated in Fig. 2.1.

In MULTI-POS, there have been three main research themes, and several sub-themes per each research theme, as enumerated below:

1. ***Location-Based Applications and Business Models***: focusing on investigating novel business models for companies involved in wireless localization, on creating novel LBS with special focus on enhanced social well-being and crisis management, and on developing the cognitive bridge between the application layer and physical layer. The sub-themes addressed here have been:
   - Location techniques, ecosystem, and business model for 2020 urban mobility
   - Deriving knowledge and social benefits from mass mobile network user data
   - Developing markets for LBS

- Location integration in crisis management life cycle
- Advanced assistance services of high performance in harsh environments

2. ***Cognitive Long-Term Location Approaches***: focusing on developing advanced technologies and algorithms for flexible, context-aware, and dynamically adaptive positioning, and on feeding into the location-based applications and business models theme the cognitive information needed to fill the gap between the application and physical layers. The sub-themes addressed here have been:

   - Building a cognitive prototyping platform
   - Developing cognitive methods for positioning and communications
   - Deriving innovative interference mitigation techniques for dynamically changing environments
   - Context-aware semantic processing
   - Adaptive flexible sensor integration for indoor location

3. ***Hybrid Positioning Technologies***: dealing with novel hybridization solutions that take into account also the trustworthiness of the offered solutions, and with the definition of the technology gaps that can be filled by the cognitive positioning long-term approaches and to feed them into the location-based applications and business models. The sub-themes addressed here have been:

   - Large scale data analytics systems for GNSS data
   - Signals of Opportunity
   - Signals for combined positioning and communication systems
   - GNSS integrity monitoring in urban environment
   - Multi-GNSS positioning

2.3 Main Results and a Look Ahead

MULTI-POS covered a wide and varied range of research topics, as also demonstrated by this book. Regular training and networking events have been organized within the network, with an average of an event at each 4 months, as seen in Table 2.2.

The cross-sector training has been strongly emphasized in MULTI-POS, as all but one among the long-time recruited fellows have got two-sector experience. The MULTI-POS fellows have been exposed to complementary industrial and academic training, with at least 5 months of secondment in a sector different from the sector of their host unit. Also, fellows have been involved in international mobility. A map of the international mobility of the network is shown in Fig. 2.2. The arrows point out to the secondment country. The red circles indicate the locations of the partners of the network.

MULTI-POS greatly helped the fellows getting access to top researchers and sharing ideas and thoughts with them. Some fellows experienced “personal paradigm shifts” when tackling jointly certain research problems, by getting inspiration from the other network members and their collaborators.

Table 2.2 MULTI-POS events

Network event	Location	Date
Kick-off meeting	Brussels, Belgium	Sep 2013
First workshop	Tampere, Finland	Sep 2013
Special session in DASIP 2013 conference	Cagliari, Italy	Oct 2013
Second workshop Prague	Prague, Czech Republic	Dec 2013
Winter school	Ruka, Finland	Feb 2014
Summer school	St. Petersburg, Russia	Jun 2014
Special session in ICL-GNSS 2014 conference	Helsinki, Finland	Jun 2014
Midterm review	Amsterdam, Netherlands	Jul 2014
Special session in ISWCS 2014 conference	Barcelona, Spain	Aug 2014
Third workshop	Barcelona, Spain	Nov 2014
Spring school	Toulouse, France	Apr 2015
Special session in ICL-GNSS 2015	Gothenburg, Sweden	Jun 2015
Fourth workshop	Gothenburg, Sweden	Jun 2015
Spring school and fifth workshop	Nottingham, UK	Apr 2016
Special session in ICL-GNSS 2016	Barcelona, Spain	Jun 2016

Fig. 2.2 MULTI-POS international mobility of fellows

Among the main research results generated within this network, we mention:

- new algorithms and receiver designs for GNSS, such as Vector Delay Frequency Lock Loop (VDFLL)-based dual constellation GNSS receiver design and developments of a computationally efficient algorithm which is capable to perform a binary detection whether GNSS signal is present or not,
- novel jamming detection schemes for GNSS receivers,
- integration of data driven methodologies for ionospheric models generation and development of a methodology for ionospheric threat assessment for low latitudes equatorial data,
- multi-GNSS performance investigation based on full constellation simulated data,
- proposing an NLOS measurement rejection technique prior to the navigation module through the use of a 180° Field-of-View Fisheye camera mounted on the vehicle rooftop,
- extensive comparative analysis in terms of tracking robustness and positioning performance for the automotive usage in case of multipath and signal outages of the conventional GNSS scalar tracking receiver and the vectorized tracking and navigation module,
- evaluation of hybrid cooperative positioning in low signal-to-noise ratio environments, explicitly accounting for scenarios where receivers can receive signals but cannot decode navigation messages,
- development of a new channel prediction framework for mobile multi-agent networks which incorporates the agents' location uncertainty,
- new architectures and frameworks for semantic analysis of location-related parameters,
- proposing a new software component, called the Context Engine, that is responsible for dealing with user context in mobile devices,
- working on information inference, i.e., the prediction of relevant (unknown) user information based on available information and making available two information inference experiments,
- new methods for navigation sensor integration and detection and hybridization of Signals of Opportunity,
- studies about end-user preferences in current GNSS applications and market opportunities for future developments,
- detailed surveys on LBS in environmental protection, safety applications and drones, and commercial Galileo services,
- studies about the possibilities for applying geographic information systems and remote sensing for environmental purposes.

Several of the reports and deliverables within MULTI-POS are available in public access from the project webpage www.multi-pos.eu. The MULTI-POS list of publications is also available at the above-mentioned project webpage.

Chapter 3
Understanding the GNSS Signal Model

Ondrej Daniel, Nunzia Giorgia Ferrara, Pedro Figueiredo e Silva, and Jari Nurmi

3.1 Introduction

The aim of this chapter is to provide a basic understanding of GNSS signals and their widespread mathematical model. Even though it might seem that all these aspects have already been addressed properly in the available literature, the authors' feeling is that the topic still deserves attention, especially for someone who is freshman in the GNSS world.

Very often, the literature on satellite navigation starts directly with the description of a given GNSS signal model in its complex-number representation. The spectrum of this signal is located around zero frequency, which is known as baseband. This signal baseband representation is very useful and therefore it is usually the starting point for any task related to the GNSS signal processing. However, in reality, all the radio signals—including the signals used for navigation—are real and their spectra occupy bands around their carrier frequencies. Due to the widespread usage of the complex representation, the link between the passband reality and the baseband model naturally vanishes. So one of the motivations behind this chapter is to remind the relation between these two representations. The chapter also attempts to explain the way how the noise, which ubiquitously disturbs the useful navigation signal, is transformed from the real world to the baseband representation. This understanding is crucial for proper noise setting in a computer simulation of GNSS signals and this chapter provides a recipe to properly make the settings. Even though the final result is relatively simply, the path to this result is not so straightforward and requires broader understanding. Thus, the chapter outlines several different aspects related

O. Daniel (✉) • N.G. Ferrara • P. Figueiredo e Silva • J. Nurmi
Tampere University of Technology, Korkeakoulunkatu 10, 33720 Tampere, Finland
e-mail: ondrej.daniel@tut.fi; nunzia.ferrara@tut.fi; pedro.silva@tut.fi; jari.nurmi@tut.fi

J. Nurmi et al. (eds.), *Multi-Technology Positioning*,
DOI 10.1007/978-3-319-50427-8_3

to the noise modelling. Finally, the chapter focuses on very important phenomena in GNSS world—the Doppler effect. It provides derivation of typical values of the Doppler frequency shift, and it explains the impact of the effect on the overall navigation signal beyond the well-known narrow-bandwidth approximation. On an intuitive basis it explains how the traditional GNSS signal model can be adjusted to fully incorporate the Doppler effect.

In the entire chapter, the focus is on intuition rather than on mathematically rigorous approach. However, a reference to literature where the reader can obtain more thorough explanation is always provided.

The notation used in this chapter is as follows. In contrast to complex baseband signals, the real passband signals are denoted with tilde, for example $\tilde{x}(t)$. Fourier transform of signal $x(t)$ (or its samples $x[n]$) is denoted as $\mathscr{S}_x(f)$ and signal Power Spectral Density as $\mathscr{G}_x(f)$.

3.2 Signal Model Basics

The journey of the navigation signal starts at the moment when it is transmitted from a satellite towards the Earth. However, there is not only one navigation signal, but as it can be seen from Chap. 6, there are many of them. Each GNSS system allocates a few frequency bands (for example, Galileo uses bands denoted as E5, E6, and E1) and each satellite belonging to a particular system can simultaneously transmit a navigation signal in each of them. Thus, each of the bands is occupied by signals originating from different satellites. The GNSS bands are spread over the so-called L-band, which lies between 1 and 2 GHz and which has been selected due to its favorable signal propagation characteristics. The overall situation is depicted in Fig. 3.1, in which $\tilde{x}_i(t)$ denotes ith navigation signal.

It is typical in GNSS that the navigation signals, which belong to the same system and share the same band, completely overlap in the frequency domain since their carrier frequencies are the same. Also, the frequency bands used by different navigation systems can fully or partially overlap as, for example, in case of Galileo E5, GPS L5, and GLONASS G3 bands. Moreover, the signals overlap in time domain, since all the GNSS signals are transmitted continuously. If the signals overlap in both, frequency and time, domains, it raises a question how they can be distinguished by the GNSS receiver. The brief answer is that the receiver is able to separate them by means of some special signal properties, more particularly based on their correlation characteristics. Almost all GNSS systems operate in this frequency/time overlapping mode, which is in the wireless communication community known as Code Division Multiple Access (CDMA).

The situation is slightly different only for GLONASS. The frequency bands used by this system are split into several sub-bands which are assigned to individual signals. Thus, the signals are not fully overlapping in frequency domain. This operational mode is known as Frequency Division Multiple Access (FDMA) and

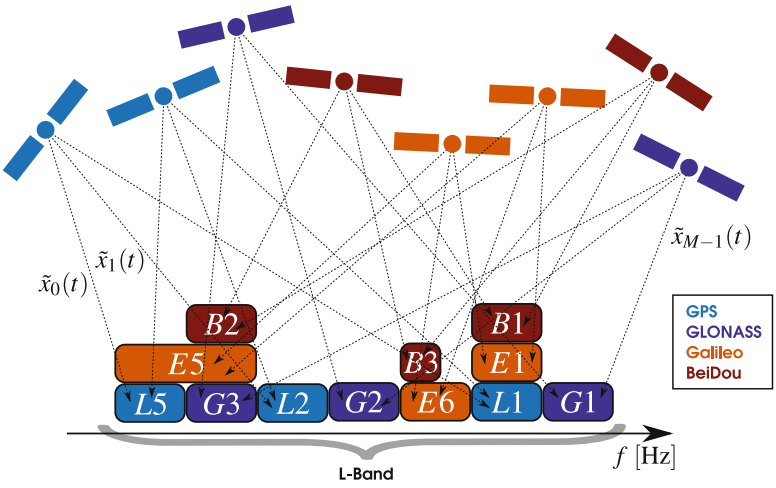

Fig. 3.1 Schematic illustration of GNSS frequency bands

it allows the receiver to distinguish the individual signals simply by means of frequency filtering. It should be noted, however, that a new CDMA-based signal has been incorporated into the GLONASS system (particularly G3 band in Fig. 3.1).

In general, CDMA is prioritized over FDMA in the GNSS world since it facilitates interoperability among the individual navigation systems and it also simplifies the design of the RF part of the GNSS receiver.

3.2.1 Passband Signal Model

The RF passband signal available at the receiver input can be modeled as

$$\tilde{r}(t) = \sum_{i=0}^{S-1} \tilde{x}_i(t) + \tilde{n}(t), \tag{3.1}$$

where S is the number of all individual navigation signals $\tilde{x}_i(t)$ originating from various GNSS systems as well as occupying various frequency bands, and $\tilde{n}(t)$ is a noise process generally representing various fundamental natural noise phenomena.

The dominant type of noise is the thermal noise caused by a random movement of electrons inside any electrical conductor and it is naturally created in the receiver hardware. Another important type of noise is the antenna noise and it comprises an influence of external natural noise sources on the receiver. Basically every object with a temperature higher than the absolute zero radiates a noise in form of electromagnetic waves which is then collected by the antenna. More information

on noise details can be found, for example, in [11] and [2]. From the mathematical point of view it is important and very convenient that the total noise effect can be modeled as Additive White Gaussian Noise (AWGN). The word *additive* suggests that the noise can be added to the useful signal simply using the plus sign as in (3.1), and the word *white* means that the noise has Power Spectral Density (Power Spectral Density (PSD)) which is constant over all frequencies in the range from $-\infty$ to ∞. The PSD is equal to $\mathscr{G}_{\tilde{n}}(f) = N_0/2\,\mathrm{W/Hz}$. It holds that $N_0 = kT$, where k is the Boltzmann's constant and T is the receiver system temperature in Kelvins, which models the joint effect of thermal and antenna noise. It is worth noting that a navigation signal travels from a satellite along its long journey (roughly 20,000 km) towards the receiver without any additive noise. The signal is degraded mainly by the receiver noise during the signal reception.

It should be clear that all the signals considered so far are *real*-valued, since only real signals can exist in the real world and since it is not possible to directly transmit a *complex*-valued signal from an antenna.

The rest of this section focuses on the signal processing of only one particular navigation signal $\tilde{x}_i(t)$ and therefore the subscript i is hereafter omitted. It is not defined explicitly into which GNSS system the signal $\tilde{x}(t)$ belongs to and hence it can be understood as a generic or template GNSS signal. Only the assumption is that the signal is transmitted at the carrier frequency f_c and it is band-limited to a bandwidth B, as shown in Fig. 3.2. The figure also depicts the constant PSD of noise $\tilde{n}(t)$. Moreover, the figure shows the signals on the frequency axis including also its negative part. It should be clear, however, that there is nothing like negative frequency axis in the real world. For example, it is not possible to transmit a harmonic signal with frequency -1 MHz. However, the concept of negative frequency axis is beneficial especially from a mathematical point of view since it eases some mathematical derivations.

Since the signal $\tilde{x}(t)$ is real its frequency spectrum $\tilde{X}(f)$ possesses special symmetry in form of $\tilde{X}(-f) = \tilde{X}(f)^*$. This result, coming directly from Fourier analysis, implies that all information carried by the signal is stored only in one half of the spectrum and that the second half is—from the information point of

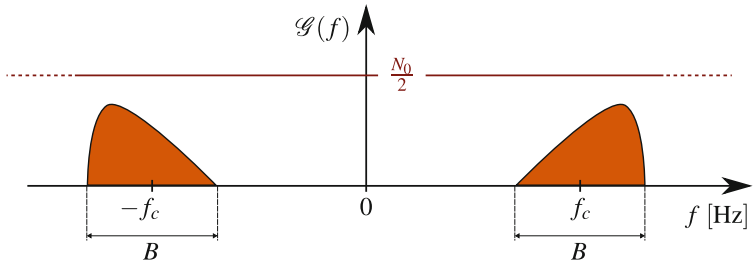

Fig. 3.2 Schematic illustration of PSD of navigation signal $\tilde{x}(t)$ and noise $\tilde{n}(t)$. The signal is transmitted from a satellite at the carrier frequency f_c and it is band-limited to bandwidth B. PSD of noise is constant with value $N_0/2$

view—redundant. This is intuitively clear, since the concept of negative frequency axis has been introduced only to facilitate some mathematical operations performed on signals, so it would not make sense if this expedient extension would add new information. Also, the signal $\tilde{x}(t)$ is transmitted at relatively high RF carrier frequency f_c. In principle, the receiver can completely process this signal directly at this high frequency, but this approach would require an unnecessary effort. For example, all hardware parts of the receiver would need to be designed and operated at a high frequency. Instead, the receiver performs only truly necessary operations at the RF frequency (for example, RF filtering and low noise amplification) and immediately after that it translates the RF signal from the carrier frequency to zero frequency. Since then the signal occupies considerably lower frequencies, the requirements on the signal processing hardware are also reduced accordingly.

The initial signal processing performed by the receiver can be summarized as follows. First, the receiver filters the signal $\tilde{x}(t)$ using an RF filter and then it amplifies it using a low noise amplifier. Afterwards the receiver multiplies the signal by the complex exponential $\exp\{-j2\pi f_c t\}$ which in frequency domain causes a shift of the overall spectrum to the left side, as depicted in Fig. 3.3. Since the information is hidden only in one half of the spectra, the second half is filtered out using a low pass filter. The resulting signal, called $x(t)$, is now complex and is known as the complex envelope of the original RF passband signal $\tilde{x}(t)$. The complex envelope lies in the band around zero frequency known as baseband. This procedure can be implemented according to the diagram depicted in Fig. 3.4.

The signal $x(t)$ is generally complex, because its spectrum is not necessarily symmetric around the zero frequency. Due to the signal complexity, the overall frequency axis now carries useful information in contrast to passband (real) signals.

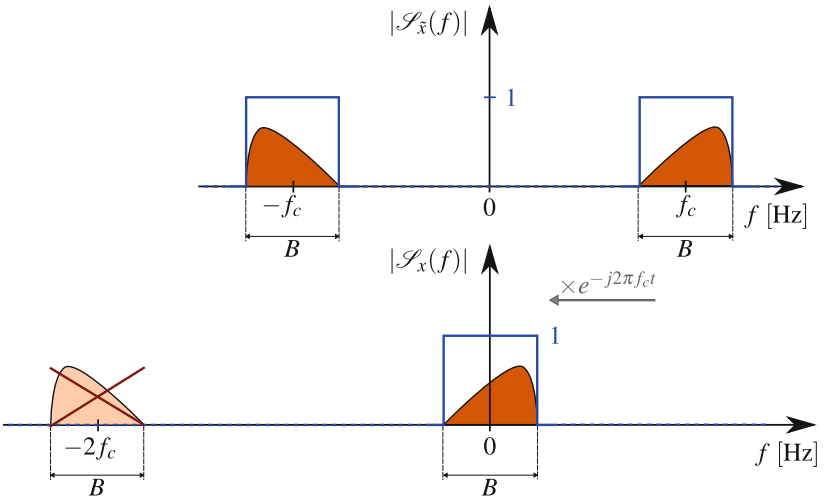

Fig. 3.3 An illustration of the frequency transformation from passband to baseband. After the frequency translation using $\exp\{-j2\pi f_c t\}$, the left part of spectra located at $-2f_c$ is filtered out

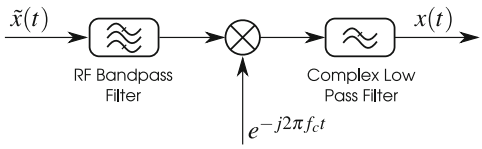

Fig. 3.4 Fundamental implementation of the signal transformation from passband to baseband

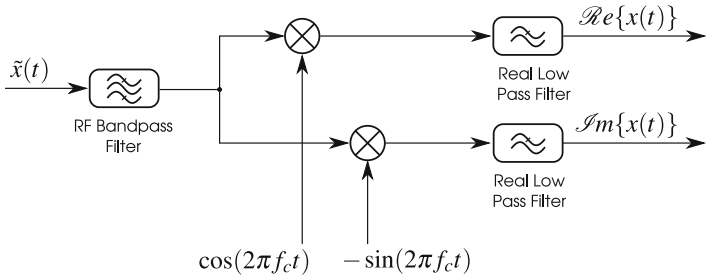

Fig. 3.5 Practical implementation of the signal transformation from passband to baseband

Therefore, from the perspective of signal processing algorithms operating with the baseband signals, the negative axis becomes actually reality. However, it is reality of numbers used by some algorithms, not reality of a real signal which can be sent in form of electromagnetic waves through space. Moreover, the complexity of the signal is not an issue. In many cases it is mathematically easier and more elegant to work with complex representation. And this is especially true for passband signals. Also, the low pass filter in Fig. 3.4 is complex in general. However, it can be shown that if its frequency response is symmetric around zero frequency, it can be implemented as two real filters, one filtrating real and the second the imaginary part of the signal. It leads to practically preferred implementation shown in Fig. 3.5 which provides the same output as the implementation according to Fig. 3.4.

3.2.2 Baseband Signal Model

The baseband navigation signal $x(t)$ is traditionally modeled as

$$x(t) = A\, s(t-\tau)\, e^{j2\pi f_D t + j\phi(t)}, \tag{3.2}$$

where $A, \tau, f_D, \phi(t)$ represent signal amplitude, signal delay due to finite propagation speed, frequency shift due to the Doppler effect, and residual carrier phase, respectively.

The multiplicative factor A captures signal amplitude variations while the signal is travelling from satellite to the receiver. The signal attenuation is caused not only by power dissipation into the space (so-called free-space loss) but also by other various phenomena, such as atmospheric attenuation due to the oxygen and attenuation caused by various obstacles. The parameter also captures the effect of amplifiers at the transmitter and receiver stage.

The main parameter of interest is the signal delay τ, since it reveals the distance to the satellite and therefore it forms the basis for the position calculation. However, the signal delay is not caused only by the distance itself, but also by some other physical phenomena. The main source of additional delay is the ionosphere which contains free electrons causing significant dispersion of GNSS signals [16]. The effect of the dispersion on the signal delay causes its artificial extension since the dispersion reduces the group velocity of the signal.

Parameter f_D represents frequency shift (or an error) between expected and actual frequency of received signal caused by the Doppler effect. The effect is discussed in more detail in Sect. 3.5. However, it should be highlighted that f_D can also be used to model the frequency error due to inaccurate local oscillators at receiver as well as transmitter. This is especially the case at the receiver side, since on the satellites very accurate atomic clocks are employed.

The residual phase $\phi(t)$ is the difference between phase of received signal and signal of the local oscillator, beyond the phase difference already modeled by the term $2\pi f_D t$. In hypothetical case of static receiver and static transmitter, the residual phase would be a constant value. However, due to the movement the phase varies. The change in the phase over time interval corresponds to the change of the distance. Thus, this variation can be used for positioning purpose. Similarly to the time delay, the phase is also significantly affected by the ionosphere. However, in contrast, the phase is advanced (or sped up) since the phase velocity of the navigation signal is increased. Interestingly, it is advanced by the same amount as the delay τ is increased [16].

All the constant signal parameters described above generally vary in time and therefore they should be thought of as functions of time. However, during a short time duration they can be considered constant. This is a valid assumption for most practical applications of interest.

The signal $s(t)$ is generally formed based on two components. The first one is specially designed to allow the receiver to estimate the signal delay τ and the generic name for this component is *code*. The second component carries data message containing several important parameters needed for actual determination of position in the receiver and it is known as *navigation message*. The data rate of the navigation message is significantly lower than the code rate. The exact structure of each component as well as the way how the components are combined together differs across the GNSS systems. The modernized GNSS signals exploit more complex techniques than their predecessors. The historically oldest but still in use publicly available GNSS signal is the Global Positioning System (GPS) L1 C/A. It uses relatively simple solution in which both code and navigation message are first multiplied together. Then the resulting discrete sequence is modulated using Binary Phase-Shift Keying (BPSK) modulation technique with rectangular modulation pulse. It finally leads to the baseband signal $s(t)$. The modernized navigation signals across all GNSS systems still rely on BPSK modulation but they also use some of the variants of the so-called Binary Offset Carrier (BOC) modulation. Thus, BPSK and BOC modulations can be understood as two main signal building blocks in the GNSS world. For the sake of clarity, the navigation message is omitted for the

following discussion and the focus now is only on the code component $c(t)$. This simplification can be justified by the fact that the data rate of the navigation message is slower than the rate of the code. Roughly saying, it implies that the important properties of signal $s(t)$ are mainly determined by the properties of the code $c(t)$.

The code component $c(t)$ is periodic with period T and is based on pseudorandom discrete sequence $\boldsymbol{c} = [c_0, c_1, \ldots, c_{M-1}]$ with M elements. Each element of the sequence is called a *chip* (as opposed to bit known from digital communication, chip does not carry any information), it has duration of $T_C = T/M$ s. The pseudo-randomness means that the signal *looks like* it would be a random process. However, the signal cannot be random completely, because the processing in the GNSS receiver completely relies on knowledge of the code. The way how the code is constructed is beyond the content of this chapter. But in the simplest case, it can be thought of just as a random periodic sequence of values -1 and 1.

It should be noted that each navigation satellite uses unique codes and therefore the receiver distinguishes the individual satellite signals based on them. This is true for the CDMA approach. However, in case of FDMA, GLONASS shares the same pseudorandom sequence across all satellites.

For the sake of clarity, the signal $c(t)$ is now expressed mathematically for the case of BPSK as well as BOC modulations. Both of these modulations use a modulation pulse with rectangular shape $h(t)$, which is depicted in Fig. 3.6.

Assuming BPSK modulation, the signal $c_{\mathrm{BPSK}}(t)$ can be expressed as

$$c_{\mathrm{BPSK}}(t) = \sum_{n=-\infty}^{\infty} \boldsymbol{c}[n \ (\mathrm{mod}\ M)]h(t - nT_C), \tag{3.3}$$

whereas for BOC modulation the signal $c_{\mathrm{BOC}}(t)$ is

$$c_{\mathrm{BOC}}(t) = \underbrace{\sum_{n=-\infty}^{\infty} \boldsymbol{c}[n \ (\mathrm{mod}\ M)]h(t - nT_C)}_{c_{\mathrm{BPSK}}(t)} \ \mathrm{sign}\{\sin(2\pi f_{\mathrm{SC}} t)\}), \tag{3.4}$$

where the sign() is a function returning a sign of its argument and f_{SC} is called as sub-carrier frequency. Time domain examples of both signals are depicted in Fig. 3.7. In the figure, it is assumed that $f_{\mathrm{SC}} = 1/T_C$. This leads to the basic form of BOC modulation, labeled in literature as BOC(1,1). More details on the BOC modulation can be found in [13].

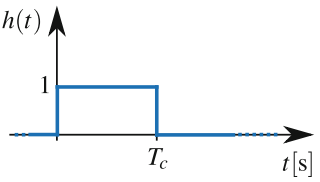

Fig. 3.6 The rectangular modulation pulse used in GNSS

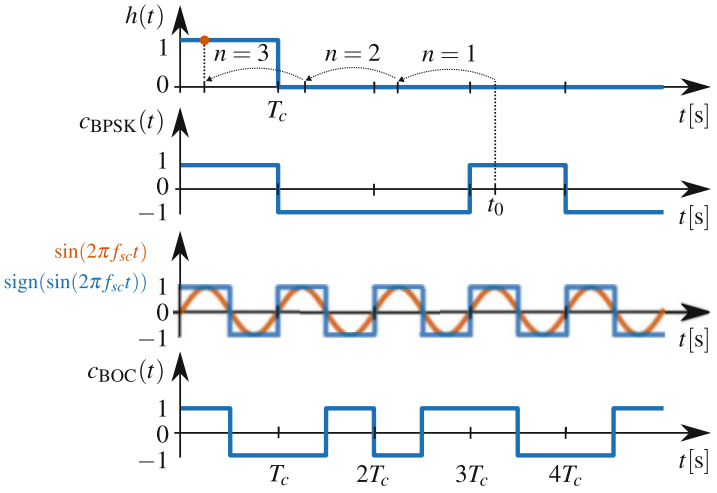

Fig. 3.7 Time domain examples of generating BPSK and BOC modulated signals

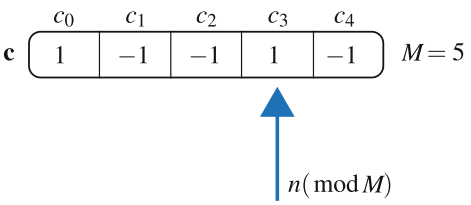

Fig. 3.8 Illustration of the code vector containing the pseudorandom sequence

It is instructive to properly understand the math expressed by the modulation Eqs. (3.3) and (3.4). For that reason, Fig. 3.7 focuses on generation of the signal $c_{\mathrm{BPSK}}(t)$ for a particular time moment $t = t_0$. The time is plugged into Eq. (3.3) and it turns out that the value of modulation pulse $h(t_0 - nT_C)$ shall be evaluated next. For that fixed time $t = t_0$, the summation index $n \in \mathcal{N}$ increments its value by one starting from $-\infty$ and it continues towards ∞. Only when n takes such value that the input argument of the modulation pulse $(t_0 - nT_C)$ falls into the interval $[0, T_C)$, the function $h(t_0 - nT_C)$ returns non-zero value. In this particular case, $h(t_0 - nT_C)$ is non-zero only for $n = 3$. This value of index n is then used for addressing the code vector $\boldsymbol{c}$. Since this sequence is periodic, the modulo operation ensures that when the index n reaches the end of the sequence, the value $n + 1$ again points to the first element of the vector. The code vector is illustratively depicted in Fig. 3.8 for a sequence with short length equal to $M = 5$. The signal modulated using BOC is obtained similarly.

What is the advantage of BOC modulation with respect to BPSK? Positioning accuracy is tightly coupled with the ability to accurately estimate the time delay τ. The variance of time delay estimate σ_τ^2 is theoretically bounded, for example, by Cramér-Rao Lower bound (CRLB). It determines the lowest possible value of the variance that can be achieved by any unbiased estimator. As shown in [9], the bound specifies that

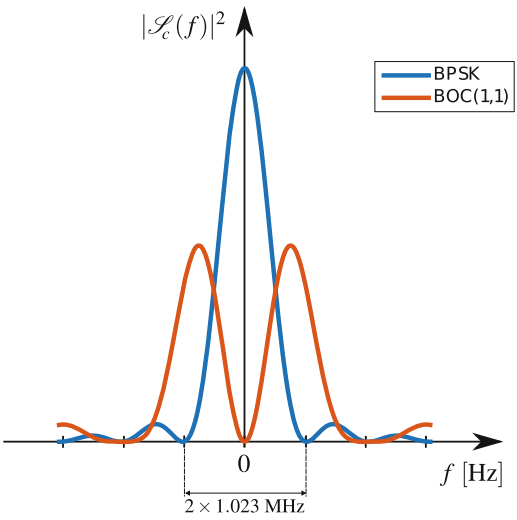

Fig. 3.9 Squared absolute value of BPSK and BOC(1,1) spectrum

$$\sigma_\tau^2 \geq \frac{1}{\text{SNR} \times F^2}, \tag{3.5}$$

where SNR stands for the Signal-to-Noise Ratio and F^2 represents the mean square bandwidth expressed as

$$F^2 = \frac{\int_{-\infty}^{\infty} (2\pi f)^2 |\mathscr{S}_c(f)|^2 df}{\int_{-\infty}^{\infty} |\mathscr{S}_c(f)|^2 df}. \tag{3.6}$$

Now, the spectra of $c_{\text{BPSK}}(t)$ and $c_{\text{BOC}}(t)$ are depicted in Fig. 3.9. Considering infinity frequency axis, the area under both curves is the same. Thus, it is clear that the mean square bandwidth F^2 is higher for BOC modulation than for BPSK. Therefore, for the same value of SNR, BOC modulation can achieve higher accuracy of time delay estimate and therefore better positioning accuracy.

The overall baseband signal model (3.2) is expressed in a continuous time domain. However, the receiver usually processes this signal using a digital signal processing techniques that operate on discrete samples. These samples can be obtained by means of an Analog-to-Digital converter. Moreover, while performing a computer simulation in order to experiment with these navigation signals, the discrete signal model is also necessary. If the signal (3.2) is sampled with sampling frequency f_S, the discrete signal representation reads as

$$x[n] = x(nT_s) = A\, s(nT_s - \tau) e^{j2\pi f_D n T_s + j\phi(nT_s)}, \tag{3.7}$$

where $T_S = 1/f_S$ is the sampling period. Since the signal bandwidth is B, the sampling frequency f_S must satisfy the condition that $f_S \geq B$. Only in this case the discrete signal is not degraded by the effect known as aliasing (see Fig. 3.10).

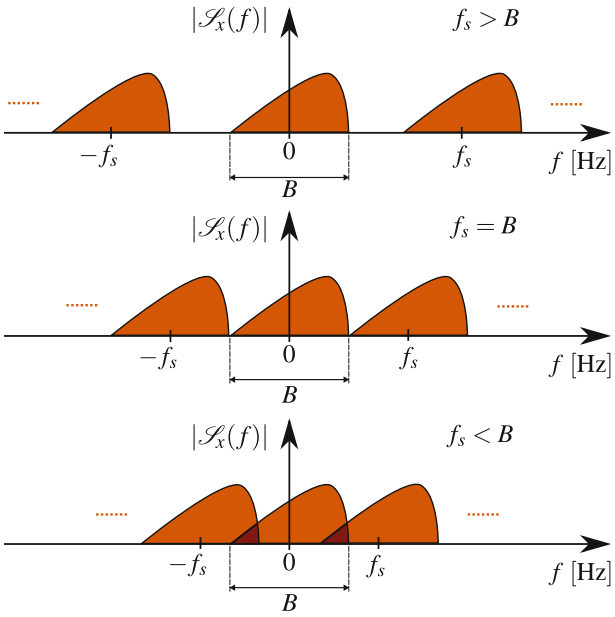

Fig. 3.10 Aliasing effect. For sampling frequencies smaller than the signal bandwidth, the periodic parts of spectra are overlapping causing degradation of the discrete signal

3.3 Signal-to-Noise and Carrier-to-Noise Density Ratios

The ultimate goal of the GNSS systems is to allow the receiver to determine its position. Due to the ubiquitous noise, the calculated position is always only an estimate of the true position. Intuitively, the quality of the estimate is affected by the strength of the navigation signals with respect to the amount of noise. The vague words amount and strength should be replaced by the term power since it turns out that this is the crucial quantity. Therefore, the power of useful signal relatively to the power of noise has significant impact on the error in the position estimate. The term Signal-to-Noise Ratio, introduced previously in (3.5), is nothing but the ratio between these two powers and it is the subject of this section.

First of all, a conceptual issue with the power of the noise should be tackled now. In the previous section, it was defined that the PSD of the noise $\tilde{n}(t)$ is constant with value equal to $N_0/2$. Since the noise power is obtained by integrating PSD over bandwidth, which is however infinite in this case, the overall noise power seems to be also infinite. But an arbitrary signal (in this case it is a signal representing the noise) with infinite power cannot exist in practice. The key to figure this paradox out is to realize that the noise model with constant PSD is only an idealization of the real world. The model makes sense in practice only if the noise is understood as an input to a system with limited bandwidth. Then the power of the noise propagating into this system (in this case into the GNSS receiver) is finite and it is given directly

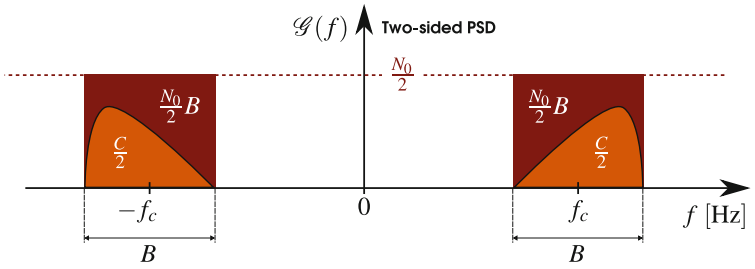

Fig. 3.11 Schematic two-sided PSD illustration of navigation signal $\tilde{x}(t)$ and noise process $\tilde{n}(t)$

by PSD and the system bandwidth. In other words, from the GNSS receiver point of view, it is only important that the noise PSD takes constant value within the input bandwidth. It really does not matter how it behaves outside of the system bandwidth. More thorough explanation can be found in [21]. Typically, the receiver contains an input RF filter which is the first filter determining this bandwidth. It should be clear now that the SNR concept implicitly assumes a limited system bandwidth, because only then the noise power takes finite value.

This is depicted in Fig. 3.11. The figure shows the PSD of the noise before and after filtration at the input of the GNSS receiver as well as the PSD of the useful signal. The bandwidth of the useful signal is limited already at the transmitter side, but the bandwidth of the noise is limited only at the receiver. In Fig. 3.11, the power of useful signal is denoted as C and it can be calculated as the area under the PSD. Since Fig. 3.11 shows two-sided (meaning that negative frequencies are also considered) PSD spectrum, half of the signal power is concentrated around frequency $-f_c$ and the second half around f_c. The power of noise is calculated in the same way as power of useful signal and hence the total noise power after the RF filter (together over both positive and negative frequency axes), denoted as N, is equal to $N = 2B(N_0/2) = BN_0$. Therefore, for SNR it holds

$$\text{SNR} = \frac{C}{N} = \frac{1}{B}\frac{C}{N_0}, \tag{3.8}$$

where the quantity C/N_0 is known as the Carrier-to-Noise Power Spectral Density Ratio. Obviously, the SNR as well as the C/N_0 can be equivalently used to describe the relation between signal and noise powers.

As it has been already discussed above, the concept of negative frequency axis is only theoretical and is used for mathematical convenience. To adapt the situation shown in Fig. 3.11 closer to the reality, the negative axis might be simply deleted and overall signal and noise power transformed to the positive axis. It leads to Fig. 3.12 where it can be seen that power of the signal around carrier f_C is C and PSD of the noise is now N_0. It is said that this representation is single-sided, since only the positive side of frequency axis is taken into account. It is obvious that SNR is still

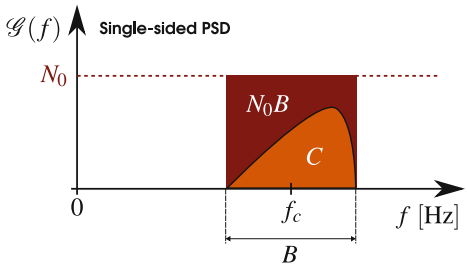

Fig. 3.12 Schematic single-sided PSD illustration of navigation signal $\tilde{x}(t)$ and noise process $\tilde{n}(t)$

the same as before since signal and noise power remains the same. It was mentioned that the situation depicted in Fig. 3.12 is closer to reality than the situation depicted in Fig. 3.11. Why? Simply because the situation according to Fig. 3.12 is something what an radio engineer can directly observe on a spectral analyzer. The level of the noise PSD is N_0 simply because it is a level directly seen on the analyzer. If one wants to exploit the concept of two-sided representation, she or he needs to adapt the noise PSD to $N_0/2$ and assume that the power of the useful signal is also equally distributed between both axes. This is exactly the reason why the model (3.1) defines that the noise PSD as $N_0/2$.

One should be aware of the fact that the term SNR is preferred in wireless communication community besides C/N_0. The reason is that the definition of signals for digital communication is very tightly connected with its bandwidth. To properly demodulate the data from a communication signal, the receiver is always supposed to use the bandwidth defined according to the standard. Hence the bandwidth might be thought of as a constant for given digital communication signal.

However, in the GNSS world the situation is a bit different and the C/N_0 metric is the most preferred one, as explained next. The bandwidth of the navigation signal $s(t)$, as transmitted from satellite is typically relatively large. For example, as it can be seen from Chap. 6, Fig. 6.1, the bandwidth of GPS L1 C/A signal is around 24 MHz. Of course, the receiver can use this entire bandwidth to achieve best positioning performance, however, it does not have to since it is possible to determine the position even with smaller bandwidth. Regarding the GPS L1 C/A signal, the majority (around 90%) of its power is concentrated in main lobe located around carrier frequency f_C with bandwidth 2×1.023 MHz (See Fig. 3.9). It is considered as a de facto standard that minimum operational bandwidth should always cover the main signal lobe. However, a GNSS receiver can also exploit higher signal bandwidth to achieve higher positioning accuracy. To conclude, the bandwidth B is receiver-dependent parameter.

There are two important remarks. The first comment deals with the intuition behind the reason why a larger bandwidth leads to higher positioning accuracy. Simply, higher bandwidth means sharper edges of the signal in time domain, which are important from delay accuracy point of view. The effect of the signal bandwidth on time delay accuracy can also be seen from the mean squared bandwidth given

by (3.6). In the equation it is assumed that the overall signal bandwidth is available and therefore the integration of the signal spectrum is over infinity frequency range. If the bandwidth is limited at the receiver, however, then the integration frequency interval is limited accordingly. This leads to smaller value of the mean squared bandwidth and thus to higher theoretical variance of delay estimator according to (3.5).

The second comment deals with the fact that the input bandwidth B determines also how much power of useful signal propagates through the input RF filter into the GNSS receiver. Equation (3.8) implicitly assumes that the overall signal power available on the receiver antenna passes completely through the input RF filter. However, as explained above, it does not need to be the case, since GNSS receiver input bandwidth might be smaller than the overall signal bandwidth of the transmitted navigation signal. Therefore, to be precise, the power of the useful signal before and after the RF filter should be distinguished. If the power before the filter is denoted as C and the power after the filter as S, Eq. (3.8) can be extended into the form of

$$\mathrm{SNR} = \frac{S}{N} = \frac{\mu(B)\, C}{N} = \frac{\mu(B)}{B} \frac{C}{N_0}, \tag{3.9}$$

where $\mu(B)$ is a correction factor specifying the fraction of useful signal power that goes through the input RF filter (see Fig. 3.13). The factor depends on filter bandwidth as well as on the shape of the actual signal spectrum. For example, if the GPS L1 C/A signal would be captured with the bandwidth 2×1.023 MHz, then $\mu(B) = 0.9$ as mentioned above. Equation (3.9) makes clear distinction between SNR and C/N_0. It implies that the SNR captures the relationship between signal and noise after the input filter, whereas the C/N_0 does that before it. From a practical point of view the factor $\mu(B)$ is usually close to one and therefore can be neglected. Hereafter, this assumption is applied for the rest of this text, and therefore it is assumed that $S = C$.

The previous comments also outline the justification why among the GNSS community the C/N_0 is the preferred metric over the SNR. Imagine a situation in which an engineer is given a task to compare the positioning accuracy of two different GNSS receivers, each of them with different bandwidth B. She or he has at his disposal a GNSS RF generator and noise generator. To make a fair and practically useful comparison, the simplest approach is to provide both receivers with a useful signal with power C and to generate noise with PSD equal to N_0 over the range of frequencies not smaller than input bandwidths of the receivers. Since the C/N_0 is bandwidth independent metric, it might serve as common reference for comparison of both receivers. On the other side, working with SNR might be more intuitive, since it directly compares power of signal and power of noise. Therefore it directly expresses how noisy a signal is without a need to consider the bandwidth. General conclusion might be that when the receiver is treated as a black box then the C/N_0 might be the preferred metric. When the internal receiver signal processing algorithms are concerned then the SNR metric might be more useful.

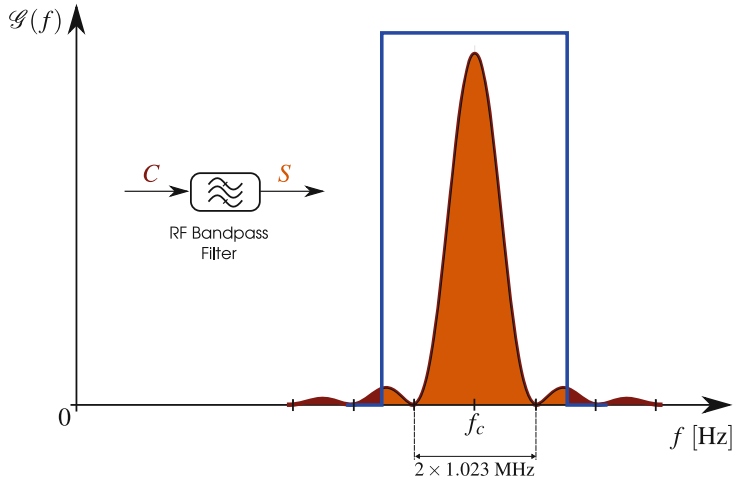

Fig. 3.13 Power and PSD of useful signal before and after the input RF filter

As summarized in [7], both metrics are important and their usage depends on the context.

3.4 Noise in Computer Simulation

A computer simulation involving radio signals is almost always performed in baseband, rather than directly in passband. The reason is that the simulation in baseband does not pose so high computational burden since the signal occupies lower frequencies. Moreover, mathematical operations on a signal are simplified and more intuitive when the complex signal representation is utilized. Therefore, working with radio signals in baseband is a very convenient approach.

A typical task while working with GNSS signals is to simulate a baseband navigation signal $x(t)$ according to (3.2) and then to add a baseband noise $n(t)$ with such scaling that the resulting noisy signal $y(t) = x(t) + n(t)$ attains a given value of C/N_0. However, C/N_0 is a metric to describe relation between signal and noise power exclusively in passband. Therefore, great care must be taken to ensure that power of the baseband signals $x(t)$ and $n(t)$ is such that signal $y(t)$ properly models the corresponding passband signal with the given value of C/N_0. This task is often subject of 3 dB error, meaning that the generated baseband noisy signal $y(t)$ has C/N_0 lower or higher by 3 dB than it should have in order to exactly model the corresponding passband signal. Zhang and Miller [24] is completely devoted to discussion and explanation of this kind of error in general communication context and is recommended as a relevant source of information. The ultimate aim of this section is to provide a recipe how to properly generate the baseband signal $y(t)$ to correctly capture the effect of C/N_0.

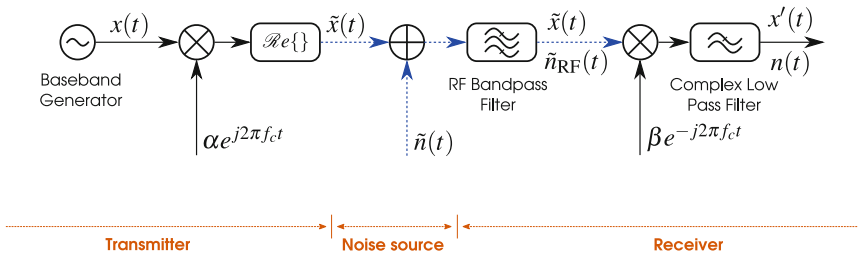

Fig. 3.14 The generic radio chain formed by transmitter, channel and receiver. The signals expressed using *dashed blue lines* are real-valued, whereas *solid black lines* are used to denote complex-valued signals

One of the reasons causing the 3 dB error might be the fact that the useful signal is generated in baseband, but the baseband noise, which should be added to the signal, is specified with the C/N_0 parameter applicable to passband. Therefore, in this section the journey of the useful signal is observed from its origin in baseband to passband where the noise is added. Then the signal and noise are treated in isolation while passing from passband to baseband in the receiver. Even though in reality the signal and noise are summed together already in passband, the separate analysis of their transformations into baseband is possible, due to the fact that overall system is considered to be linear and therefore the superposition principle holds.

Another reason for the 3 dB error might be caused by different scaling constants (typically factors 2 or $\sqrt{2}$) introduced to the signal path by various authors. Therefore, in the subsequent analysis, all possible scaling constants are kept in a symbolic notation to make their impact visible.

Before proceeding further, the power of a generally complex signal $u(t)$ calculated in time domain is defined as [1]

$$\mathscr{P}\{u(t)\} = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} |u(t)|^2 dt = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} u(t)u^*(t)dt, \tag{3.10}$$

where it is assumed that the limit exists. The definition holds for deterministic functions as well as for random processes. However, for general random process, the quantity $\mathscr{P}\{u(t)\}$ is also random. Fortunately, the band-limited Gaussian noise process with constant PSD, which is considered here, belongs to the class of the so-called ergodic signals with the convenient property that the power according to (3.10) is a non-random quantity. For more details on ergodic processes, see [15] or [4]. For the following discussion it is useful to note that if the signal $u(t)$ is multiplied by a real constant γ, the power of the resulting signal is $\mathscr{P}\{\gamma\, u(t)\} = \gamma^2 \mathscr{P}\{u(t)\}$. Moreover, multiplying the signal $u(t)$ by complex exponential $\exp\{\pm jU\}$ does not affect the power at all.

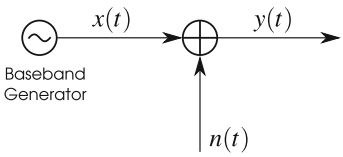

Fig. 3.15 The generic radio chain simplification when the signal modelling is performed solely in baseband

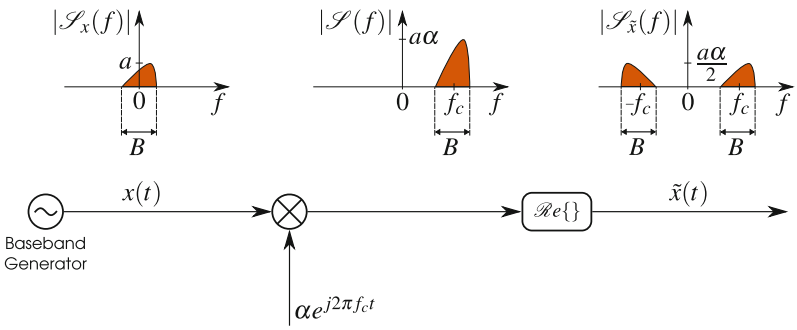

Fig. 3.16 Schematic diagram of the transmitter part

3.4.1 From Baseband to Passband and Back

An overall system diagram which mimics a generic radio chain formed by transmitter, channel and receiver is depicted in Fig. 3.14. The signal $x(t)$ is generated in baseband. Then it is transformed to passband, where the noise is added. Finally, the signal together with the noise is then transformed back to the baseband. Our aim now is to investigate how the noise added to the signal in reality in passband can be added into the signal directly in baseband. This finding will allow us to perform all operations with the signal solely in baseband according to Fig. 3.15, which significantly simplifies the situation shown in Fig. 3.14. This simplification is the beauty of a radio channel simulation in baseband.

First, the focus is on the transmitter part, as depicted in Fig. 3.16. The baseband generator produces the signal $x(t)$. Absolute value of its spectrum $|\mathscr{S}_x(f)|$ is shown in the figure only schematically. Then this signal is multiplied by signal $\alpha \exp\{j2\pi f_c t\}$, $\alpha > 0$, causing the frequency translation of the signal spectrum from baseband, where it is located around zero frequency, to the carrier frequency f_c. This signal is still complex, since its spectrum is non-symmetric around the zero frequency. This signal is known in literature as analytical signal. The following operation $\mathscr{R}e\{\}$ ensures that the output signal is real, as only real signal can be directly transmitted from the antenna. Note the effect of the operation $\mathscr{R}e\{\}$ on the spectrum. The operation creates symmetrical image of the spectrum at frequency $-f_c$. However, it also decreases spectrum value from $(a\,\alpha)$ to $(a\,\alpha/2)$, where a is the height of the absolute spectrum $|\mathscr{S}_x(f)|$ (see Fig. 3.16). This is due to the fact that the operation $\mathscr{R}e\{\}$ applied to a signal $u(t)$ is equal to $\mathscr{R}e\{u(t)\} = \frac{1}{2}\{u(t) + u^*(t)\}$.

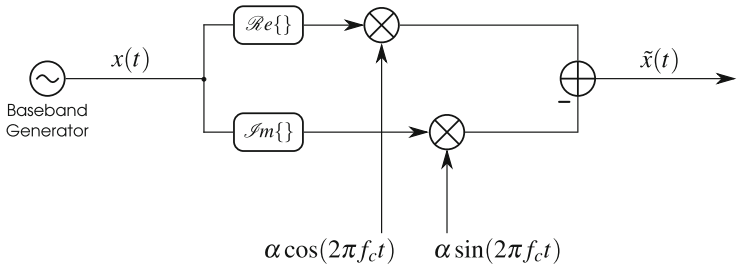

Fig. 3.17 Practical implementation of the transmitter part

The diagram in Fig. 3.16 shows only the principle of the creation of the passband signal $\tilde{x}(t)$. In practice, a radio transmitter is traditionally implemented according to Fig. 3.17, which is derived from Fig. 3.16 using the Euler's formula $\exp\{jx\} = \cos x + j \sin x$. Of course, both implementations provide exactly the same passband signals.

Now, consider that the power of baseband signal $x(t)$ is $\mathscr{P}\{x(t)\}$. What is the power of the signal $\tilde{x}(t)$, denoted as $\mathscr{P}\{\tilde{x}(t)\}$? It is clear from Fig. 3.16 that $\tilde{x}(t) = \mathscr{R}e\{\alpha x(t) \exp\{j2\pi f_c t\}\}$. Plugging this into (3.10) results in

$$\begin{aligned}\mathscr{P}\{\tilde{x}(t)\} &= \mathscr{P}\left\{\mathscr{R}e\left\{\alpha x(t) e^{j2\pi f_c t}\right\}\right\} = \mathscr{P}\left\{\frac{1}{2}\alpha x(t) e^{j2\pi f_c t} + \frac{1}{2}\alpha x^*(t) e^{-j2\pi f_c t}\right\} \\ &= \frac{\alpha^2}{4}\mathscr{P}\{x(t)\} + \frac{\alpha^2}{4}\mathscr{P}\{x(t)\} + \underbrace{\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}\frac{\alpha^2}{2}\mathscr{R}e\left\{x^2(t)e^{j4\pi f_c t}\right\}dt}_{\to 0},\end{aligned} \tag{3.11}$$

where the last term vanishes. Roughly saying, it is because the integrand oscillates due to the exponential function and hence the integral does not grow faster than T which goes to ∞. Or even more intuitively, since the signal $\tilde{x}(t)$ consists of two separated pieces of spectra, the total power should be the sum of the powers associated with the individual pieces. Thus

$$\mathscr{P}\{\tilde{x}(t)\} = \underbrace{\frac{\alpha^2}{4}\mathscr{P}\{x(t)\}}_{\text{Power around } -f_c} + \underbrace{\frac{\alpha^2}{4}\mathscr{P}\{x(t)\}}_{\text{Power around } f_c} = \frac{\alpha^2}{2}\mathscr{P}\{x(t)\} = C, \tag{3.12}$$

where the symbol C is used again to denote the total power of useful passband signal as it was done previously.

Now the focus is on the input of the receiver, as seen in Fig. 3.18. It is assumed that the passband signal $\tilde{x}(t)$ passes the input RF filter unaffected. This is definitely not the case for noise $\tilde{n}(t)$, whose theoretically infinite bandwidth is limited by the

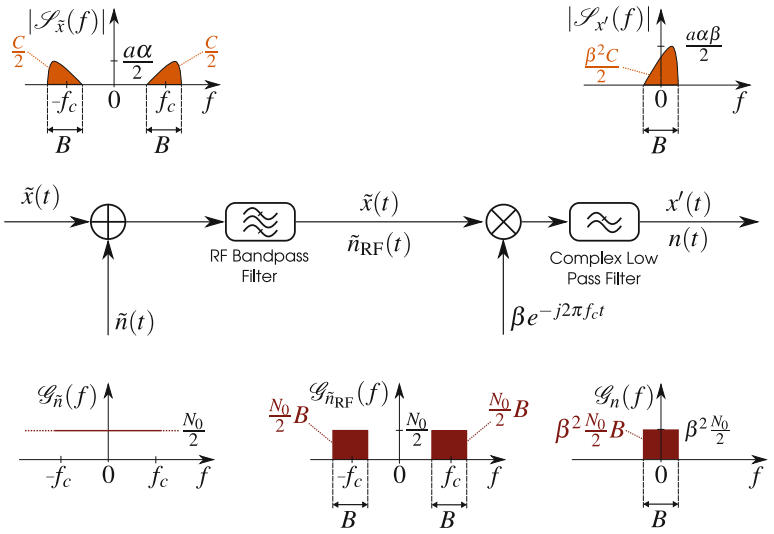

Fig. 3.18 Schematic diagram of the receiver part together with the noise source

filter as depicted in the figure. The noise after the RF filter is denoted as $\tilde{n}_{\mathrm{RF}}(t)$ and its power is $\mathscr{P}\{\tilde{n}_{\mathrm{RF}}(t)\} = 2(\frac{N_0}{2}B)$. Therefore, SNR after the RF filter is expressed as follows:

$$\mathrm{SNR} = \frac{\mathscr{P}\{\tilde{x}(t)\}}{\mathscr{P}\{\tilde{n}_{\mathrm{RF}}(t)\}} = \frac{\frac{\alpha^2}{2}\mathscr{P}\{x(t)\}}{2(\frac{N_0}{2}B)} = \frac{1}{B}\frac{C}{N_0}, \tag{3.13}$$

which is nothing else than (3.8), but now this equation also incorporates the power of baseband signal $x(t)$.

The useful signal $\tilde{x}(t)$ and noise $\tilde{n}_{\mathrm{RF}}(t)$ are now supposed to be transformed into baseband. The transformation is now done separately, even though at this point in the receiver the signals are summed inextricably together. This can be done due to the principle of superposition. This transformation has already been discussed (see Fig. 3.3) and therefore the explanation here is only brief. First, the focus is on signal $\tilde{x}(t)$. The signal is multiplied by $\beta \exp\{-j2\pi f_c t\}$, $\beta > 0$, and then filtered by a low pass filter resulting in the baseband signal $x'(t)$, which is the same as the original signal $x(t)$. They differ only by a scaling constant and therefore they have different powers. Power of the signal $x'(t)$ is equal to

$$\begin{aligned}\mathscr{P}\{x'(t)\} &= \mathscr{P}\left\{\mathscr{O}\left\{\beta\tilde{x}(t)e^{-j2\pi f_c t}\right\}\right\} = \frac{1}{2}\mathscr{P}\left\{\beta\tilde{x}(t)e^{-j2\pi f_c t}\right\} \\ &= \frac{\beta^2}{2}\mathscr{P}\left\{\tilde{x}(t)\right\} = \frac{\beta^2}{2}C = \frac{\alpha^2\beta^2}{4}\mathscr{P}\{x(t)\},\end{aligned} \tag{3.14}$$

where $\mathcal{O}\{\}$ represents low-pass filtration as shown in Fig. 3.18. This filtration cuts half of the power away. Please note that the constants α and β are usually set in theoretical analyses to keep powers of these baseband signals at the same value, typically $\alpha = \beta = \sqrt{2}$ in [21] or $\alpha = 1$ and $\beta = 2$ in [1].

The same transformation is also applied to the noise resulting in baseband noise $n(t)$ with power equal to

$$\begin{aligned}\mathscr{P}\{n(t)\} &= \mathscr{P}\left\{\mathcal{O}\left\{\beta\tilde{n}_{\mathrm{RF}}(t)e^{-j2\pi f_c t}\right\}\right\} = \frac{1}{2}\mathscr{P}\left\{\beta\tilde{n}_{\mathrm{RF}}(t)e^{-j2\pi f_c t}\right\} \\ &= \frac{\beta^2}{2}\mathscr{P}\left\{\tilde{n}_{\mathrm{RF}}(t)\right\} = \beta^2\left(\frac{N_0}{2}B\right).\end{aligned} \tag{3.15}$$

Since after the RF filter both signal and noise have been translated to baseband by the same procedure, the baseband SNR should stay the same as immediately after the RF filter. Clearly, SNR in the baseband at the receiver is

$$\mathrm{SNR} = \frac{\mathscr{P}\{x'(t)\}}{\mathscr{P}\{n(t)\}} = \frac{\frac{\alpha^2\beta^2}{4}\mathscr{P}\{x(t)\}}{\beta^2(\frac{N_0}{2}B)} = \frac{\frac{\beta^2}{2}C}{\beta^2(\frac{N_0}{2}B)} = \frac{1}{B}\frac{C}{N_0}. \tag{3.16}$$

Indeed, Eqs. (3.13) and (3.16) provide the same results. Therefore, SNR takes the same values in passband as in baseband, even though the powers of individual components changed. From Eq. (3.16), it can be directly observed that

$$\mathrm{SNR} = \frac{\alpha^2\mathscr{P}\{x(t)\}}{2N_0B} = \frac{1}{B}\frac{C}{N_0}. \tag{3.17}$$

This equation can be interpreted as follows. To preserve a given C/N_0 corresponding to the situation in passband, the baseband signal $x(t)$ can be generated with power $C' = \alpha^2\mathscr{P}\{x(t)\}$ and noise $n(t)$ with power $N' = 2N_0B$. It is worth noting that under this interpretation, as the equation suggests, PSD of noise in baseband is equal to $2N_0$, however, PSD of the noise after RF filter is equal to $N_0/2$. The equation reveals that the noise power, which should be provided by the noise generator, is equal to

$$N' = 2N_0B = \frac{\alpha^2\mathscr{P}\{x(t)\}B}{C/N_0}. \tag{3.18}$$

Now the powers of signals $x(t)$ and $n(t)$ are determined to properly correspond to the given value of C/N_0. It is also clear that the PSD of the noise is constant in the bandwidth B. What is still missing is to discuss the correlation characteristics of the noise. Moreover, now the noise $n(t)$ is a complex signal and it is not directly evident, how the real and imaginary components should be generated and what is the, if any, relation between them. This is the content of the following discussion.

3.4.2 Noise Correlation

The Gaussian noise with constant PSD is an ergodic and thus also stationary process. For the stationary process $u(t)$ it holds that its auto-correlation function $\mathscr{R}_{uu}(t_1, t_2) = \mathrm{E}[u(t_1)u^*(t_2)]$ is only a function of time difference $\tau = t_1 - t_2$ and hence $\mathscr{R}_{uu}(t_1, t_2) = \mathscr{R}_{uu}(\tau) = \mathrm{E}[u(t)u^*(t-\tau)]$. Moreover, for stationary and ergodic processes, the auto-correlation function is equal to inverse Fourier transform of its PSD $\mathscr{G}(f)$, hence $\mathscr{R}_{uu}(\tau) = \int_{-\infty}^{\infty} \mathscr{G}(f) \exp\{j2\pi f\tau\} df$.

For the sake of clarity, the auto-correlation functions of all the noise signals considered so far, namely $\tilde{n}(t)$, $\tilde{n}_{\mathrm{RF}}(t)$, and $n(t)$, are now evaluated. The noise before RF filter, $\tilde{n}(t)$, has auto-correlation function

$$\mathscr{R}_{\tilde{n}\tilde{n}}(\tau) = \frac{N_0}{2} \int_{-\infty}^{\infty} e^{j2\pi f\tau} df = \frac{N_0}{2} \delta(\tau), \tag{3.19}$$

where $\delta(\tau)$ is the Dirac delta function. The equation means that $\tilde{n}(t)$ and $\tilde{n}(t+\tau)$ are completely uncorrelated for $\tau \neq 0$. Since it is assumed that the noise is Gaussian, they are also independent. Now, if the noise passes the RF filter, the resulting signal $\tilde{n}_{\mathrm{RF}}(t)$ becomes correlated as can be seen from the following equation

$$\begin{aligned} \mathscr{R}_{\tilde{n}_{\mathrm{RF}}\tilde{n}_{\mathrm{RF}}}(\tau) &= \frac{N_0}{2} \int_{-f_c-\frac{B}{2}}^{-f_c+\frac{B}{2}} e^{j2\pi f\tau} df + \frac{N_0}{2} \int_{+f_c-\frac{B}{2}}^{+f_c+\frac{B}{2}} e^{j2\pi f\tau} df \\ &= N_0 B \,\mathrm{sinc}(B\tau) \cos(2\pi f_c \tau), \end{aligned} \tag{3.20}$$

where it is used that

$$\int_{\frac{-B}{2}}^{\frac{B}{2}} e^{j2\pi f\tau} df = \frac{e^{j\pi B\tau} - e^{-j\pi B\tau}}{j2\pi\tau} = B \,\mathrm{sinc}(B\tau). \tag{3.21}$$

Finally, the noise in baseband $n(t)$ is correlated as well since its spectrum is also band-limited. Since the PSD of this noise signal is $2N_0$, its correlation function is

$$\mathscr{R}_{nn}(\tau) = 2N_0 \int_{-\frac{B}{2}}^{+\frac{B}{2}} e^{j2\pi f\tau} df = 2N_0 B \,\mathrm{sinc}(B\tau). \tag{3.22}$$

All the auto-correlation functions are depicted in Fig 3.19.

Equation (3.22) is the important one since it describes the auto-correlation of the complex noise $n(t)$ that is supposed to be generated in the simulation. To get further insight, it is valuable to express the noise as $n(t) = n_R(t) + jn_I(t)$ and to examine auto-correlations $\mathscr{R}_{n_R n_R}(\tau)$ and $\mathscr{R}_{n_I n_I}(\tau)$ as well as cross-correlations $\mathscr{R}_{n_I n_R}(\tau)$ and $\mathscr{R}_{n_R n_I}(\tau)$. It follows directly that

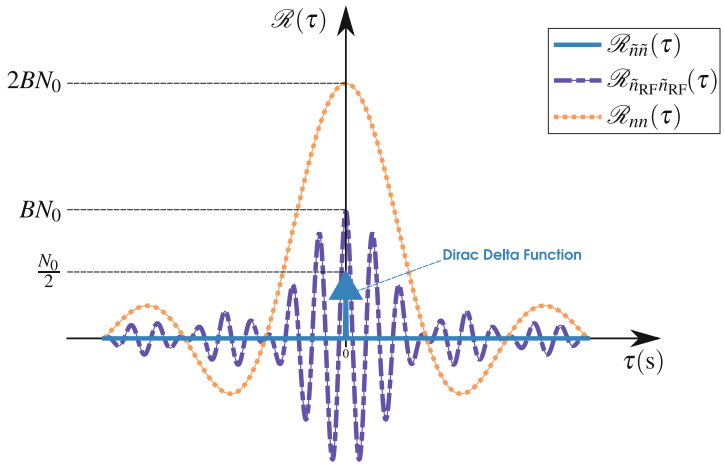

Fig. 3.19 Auto-correlation functions of noise signals $\tilde{n}(t)$, $\tilde{n}_{\mathrm{RF}}(t)$, and $n(t)$

$$\begin{aligned}\mathrm{E}[n(t)n^*(t-\tau)] &= \mathrm{E}[\{n_R(t)+jn_I(t)\}\{n_R(t-\tau)-jn_I(t-\tau)\}] \\ &= \mathscr{R}_{n_Rn_R}(\tau)+\mathscr{R}_{n_In_I}(\tau)+j\{\mathscr{R}_{n_In_R}(\tau)-\mathscr{R}_{n_Rn_I}(\tau)\} \\ &= \mathscr{R}_{nn}(\tau).\end{aligned} \tag{3.23}$$

Next, as it is shown in [21], for $f_c > B$ it holds that $\mathrm{E}[n(t)n(t-\tau)] = 0$. Thus,

$$\begin{aligned}\mathrm{E}[n(t)n(t-\tau)] &= \mathrm{E}[\{n_R(t)+jn_I(t)\}\{n_R(t-\tau)+jn_I(t-\tau)\}] \\ &= \mathscr{R}_{n_Rn_R}(\tau)-\mathscr{R}_{n_In_I}(\tau)+j\{\mathscr{R}_{n_In_R}(\tau)+\mathscr{R}_{n_Rn_I}(\tau)\} \\ &= 0\end{aligned} \tag{3.24}$$

Expressions (3.23) and (3.24) form a set of equations. Noting that auto-correlation function of the complex noise (3.22) is real, the solution of the set is as follows:

$$\mathscr{R}_{n_Rn_R}(\tau) = \mathscr{R}_{n_In_I}(\tau) = \frac{1}{2}\mathscr{R}_{nn}(\tau) = N_0B\,\mathrm{sinc}(B\tau), \tag{3.25}$$

and

$$\mathscr{R}_{n_Rn_I}(\tau) = \mathscr{R}_{n_In_R}(\tau) = 0. \tag{3.26}$$

Equations (3.25) and (3.26) already provide a receipt how the noise $n(t)$ should be generated in a simulation. First, from (3.26) it can be seen that noise components $n_R(t)$ and $n_I(t)$ are not mutually correlated at all. Again, since they are supposed to be Gaussian processes they are also independent. This simplifies the noise generation since real and imaginary parts can be generated completely independently.

It is worth of noting that the components are not correlated only because the original spectrum is symmetric around the carrier frequency. In other words, if the integration interval in (3.22) would not be symmetric, then $\mathscr{R}_{nn}(\tau)$ would not be real. This would cause a non-zero cross-correlation between real and imaginary noise components.

Moreover, (3.25) says that real and imaginary noise components are correlated in exactly the same way. Since the power of each of the noise components can be determined by evaluation of its auto-correlation function for $\tau = 0$, the equation also reveals that $\mathscr{P}\{n_R(t)\} = \mathscr{R}_{n_R n_R}(0) = N_0 B$ and also $\mathscr{P}\{n_I(t)\} = \mathscr{R}_{n_I n_I}(0) = N_0 B$. Remember that the total power of the noise in baseband is $N' = 2N_0 B$. Therefore, it can be seen that the power of complex noise is equally distributed between real and imaginary components.

Only the remaining issue is how to deal with the fact that both components, $n_R(t)$ and $n_I(t)$, are correlated. Is it really necessary to generate correlated noise components in our simulation? The brief answer is *no*. Why? Note that (3.25) represents correlation functions of real and imaginary components of continuous noise $n(t)$. To get correlation functions of corresponding discrete noise samples, it should be realized that in discrete world, τ is expressed as $\tau = kT_s = k/f_s$, where $k \in \mathscr{N}$. In other words, τ is discrete with step T_s. Therefore, the discrete correlation function is obtained from the continuous correlation function which is evaluated at delays $\tau = kT_s$ as follows:

$$\mathscr{R}_{n_R n_R}[k] = \mathscr{R}_{n_R n_R}(\tau)\big|_{\tau = kT_s} = N_0 B \,\mathrm{sinc}(BkT_s). \tag{3.27}$$

The same holds also for $\mathscr{R}_{n_I n_I}[k]$. Now the important question is how to properly select the sampling frequency in the simulation. Since the aim is to capture the continuous signal band-limited in the bandwidth B, the minimum sampling frequency is equal to $f_s = B$ to avoid aliasing. One can select sampling frequency higher than bandwidth, but the choice $f_s = B$ has a valuable consequence. Indeed, for $f_s = B$ then

$$\mathscr{R}_{n_R n_R}[k] = N_0 B \,\mathrm{sinc}(f_s k T_s) = N_0 B \delta[k], \tag{3.28}$$

where $\delta[k]$ is the Kronecker delta. The auto-correlation function is depicted in Fig. 3.20. It says that for $f_s = B$ the noise samples are uncorrelated for $k \neq 0$. Again, due to the fact that they are Gaussian, they are also independent. Therefore, it can be concluded that to properly generate the complex discrete noise samples corresponding to underlying continues noise $n(t)$, it is sufficient to just generate two independent Gaussian random discrete processes, each with power $N'/2 = N_0 B$. It is worth highlighting again that this is possible only because the spectrum with bandwidth B is symmetric around carrier frequency f_c and because the sampling frequency f_s is simply equal to B.

Moreover, it is worth of noting that the generated noise components are with zero mean value. The constant PSD directly determines that its mean value is equal to

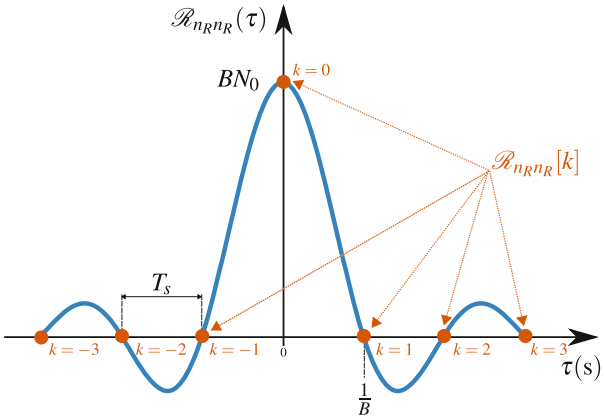

Fig. 3.20 Auto-correlation function of real part of noise $n(t)$ together with auto-correlation function of noise samples for sampling frequency $f_s = B$. The auto-correlation function of imaginary noise component is the same

zero, since a non-zero mean value would cause a discrete spectral peak (Dirac delta) at zero frequency and thus violating the constant PSD property.

3.4.3 Simulation Setup

This section introduces a baseband simulation setup which properly handles generation of a noisy navigation signal with required C/N_0.

For the actual simulation, it is assumed that a random generator providing normally distributed independent zero-mean real samples is available. The variance of these samples is supposed to be equal to one. Since the simulation is performed in discrete domain, we are going to compute the power of a discrete sequence $u[n]$, which is obtained by sampling of continuous signal $u(t)$ with sampling period T_s. The signal $u(t)$ is supposed to have total duration T, which corresponds to $M = T/T_s$ samples. The power in (3.10) can be approximated for this finite-duration signal by changing the integration interval and then replacing the integral by summation as follows:

$$\begin{aligned}\mathscr{P}\{u(t)\} &\approx \frac{1}{T}\int_0^T |u(t)|^2 dt = \frac{1}{T}\int_0^T u(t)u^*(t)dt \\ &\approx \frac{1}{MT_s}\sum_{n=0}^{M-1}|u[n]|^2 T_s = \frac{1}{M}\sum_{n=0}^{M-1}|u[n]|^2 = \frac{1}{M}\sum_{n=0}^{M-1}u[n]u^*[n] \\ &= \mathscr{P}\{u[n]\}. \end{aligned} \tag{3.29}$$

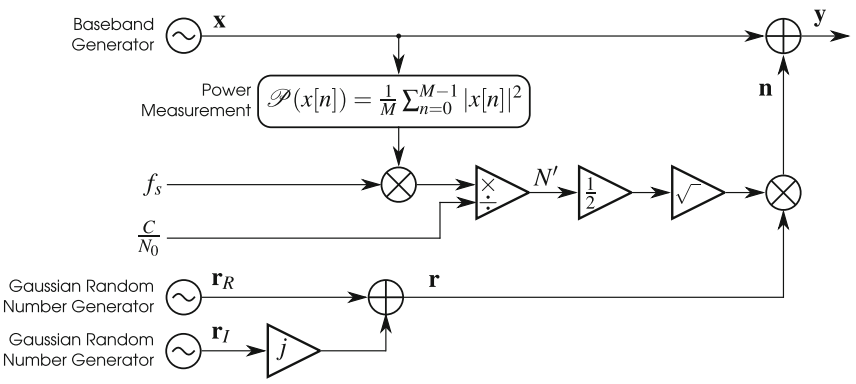

Fig. 3.21 Simulation setup for setting of the noise power to correctly capture specified C/N_0. It is assumed that $f_s = B$ and $\alpha = 1$

The overall procedure leading to the proper C/N_0 setting of the signal $y[n] = x[n] + n[n]$ in a computer simulation is summarized in Fig. 3.21. The sampling frequency is $f_s = B$. The baseband generator generates M samples of useful navigation signal $x(t)$ according to (3.2) and the samples are grouped into vector $\boldsymbol{x}$. The noise generator generates M samples from Gaussian distribution with variance equal to one. These samples are stored in vector $\boldsymbol{r}_R$. Similarly, the vector $\boldsymbol{r}_I$ is formed. From these two real noise vectors, the complex vector $\boldsymbol{r}$ is created. Then the power of the useful signal $\boldsymbol{x}$ is measured based on (3.29). Based on this power estimate the noise vector $\boldsymbol{r}$ is properly scaled to provide the noise vector $\boldsymbol{n}$ which has the required power N' according to (3.18). Finally, the useful signal and the noise vector are summed together leading to vector $\boldsymbol{y}$. Note that $C/N0$ is in linear scale. Without loss of generality, it is assumed that $\alpha = 1$. To remind, α is the scaling factor used in the transmitter to amplify the generated signal (see Fig. 3.14).

3.5 The Doppler Effect

The aim of this section is to provide an intuitive explanation of the Doppler effect, to numerically evaluate the value of frequency Doppler shift experienced by a GNSS receiver, and finally to discuss the overall impact of the effect on the radio navigation signal. The results of this section lead to an extension of the signal model (3.2) to fully capture the influence of the effect.

3.5.1 The Doppler Effect Fundamentals

It is well known that the Doppler effect causes changes of received signal frequency due to non-zero radial speed between the transmitter and receiver. This section explains the effect using a simple plane wave in unbounded, homogeneous, and nonconducting medium. The wave formula considered here represents the simplest solution of the so-called wave equation. The wave equation itself is a differential equation derived from the Maxwell equations. It poses some constraints which must be satisfied by the vector components of the electric and magnetic fields. A simple form of the wave equation is one-dimensional equation. Its solution leads to the plane wave–a wave with constant values of electromagnetic field in an infinite plane perpendicular to the direction of travelling. Even though this wave is overall simplistic, its analysis still provides valuable results. It is due to the fact that in many practical situations the real electromagnetic waves can be approximated by the plane wave. The simplest form of the plane wave is a harmonic wave given by the formula

$$\psi(t,z) = \psi_0 \cos\left\{2\pi f_{\mathrm{TX}} t - \frac{2\pi f}{c} z\right\} = \psi_0 \cos\left\{\phi(t,z)\right\}, \tag{3.30}$$

where t, z, ψ_0, f_{TX}, and c represent the time, position on z-axis, wave amplitude, frequency of transmitted harmonic signal (the subscript TX stands for *transmitter*), and wave propagation speed, respectively. The overall argument of the cosine function represents the instantaneous phase, denoted as $\phi(t,z)$. It is assumed that the wave travels in vacuum (therefore c actually represents the speed of the light) in positive direction along z-axis. It is worth noting that Eq. (3.30) shall be understood as a template; meaning that the form of the equation holds for all vector components of the electromagnetic field. To obtain more information on the theory related to the explanation above, the reader might see the classical books [5, 6], and [22].

Equation (3.30) simply describes a value of a component of the electromagnetic field at given time t and position z, provided that the wave has already propagated to that position from location of its source. Figure 3.22 depicts a scenario in which both transmitter and receiver are static and they are located at positions $z = 0$ and $z = z_0$, respectively. Then Fig. 3.23 illustrates the wave according to (3.30) for this scenario. The figure on left depicts the situation for particular time $t = t_0$, whereas the figure on the right shows the wave observed by the receiver during a time period,

Fig. 3.22 An illustration of the static scenario. The transmitter is placed at position $z = 0$ and the receiver is located at position $z = z_0$

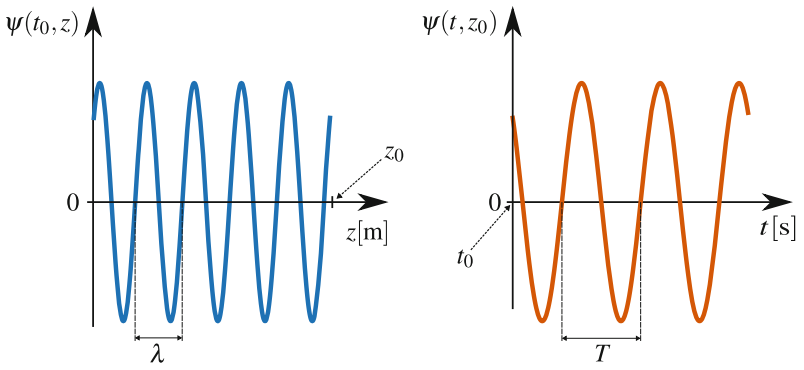

Fig. 3.23 An example of a plan wave according to (3.30) and corresponding to scenario on Fig. 3.22. On the *left side*, the wave is depicted as it propagates along the z-axis. The figure on the *right side* shows the wave as it is observed by the receiver at its location $z = z_0$. Here the wave is a function of time, starting from the moment $t = t_0$

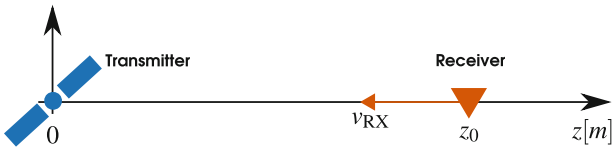

Fig. 3.24 An illustration of the scenario with the static transmitter and receiver that is moving towards the former by the constant speed v_{RX}

starting from t_0. The figure also illustrates two important wave parameters which are signal period $T = 1/f_{TX}$ and wavelength $\lambda = c/f_{TX}$.

The explanation of the Doppler effect slightly differs depending on which of the two objects (transmitter or receiver) is at rest and which is moving. The derivation below first deals with static transmitter and moving receiver. The opposite scenario is investigated afterwards.

3.5.1.1 Static Transmitter and Moving Receiver

It is assumed that the receiver is moving at a constant speed v_{RX} (the subscript RX stands for *receiver*) from position $z = z_0$ in the direction towards transmitter, as depicted in Fig. 3.24. The movement can be easily modeled in (3.30) as a replacement of z-coordinate z by the term $(z_0 - v_{RX}t)$. This leads to the instantaneous phase of the signal observed by the moving receiver given by

$$\phi'(t) = \phi(t, z_0 - v_{RX}t) = 2\pi f_{TX}t - \frac{2\pi f_{TX}}{c}(z_0 - v_{RX}t). \tag{3.31}$$

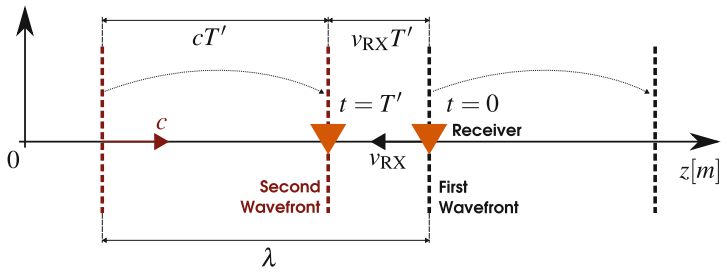

Fig. 3.25 An illustration supporting an explanation of the Doppler effect for moving receiver

In general, it holds that the first derivative of phase with respect to time is equal to the angular frequency $\omega = 2\pi f$. Thus the signal frequency observed by the moving receiver, f'_{RX}, can be expressed as follows:

$$f'_{RX} = \frac{1}{2\pi}\frac{d\phi'(t)}{dt} = f_{TX}\left(\frac{c + v_{RX}}{c}\right). \tag{3.32}$$

It can be seen from the equation that if the speed of the receiver increases in the direction towards the transmitter, the received frequency also increases. On the contrary, if the receiver is moving in the opposite direction, the sign of the speed becomes negative and thus the higher constant speed results in lower frequency of the received signal.

Another explanation, inspired by Young et al. [23], is based on Fig. 3.25. It depicts two moving wavefronts of the plane wave. The wavefront basically defines a set of points at which the wave oscillates with the same phase. The distance between wavefronts is equal to λ and they travel at speed c. It is assumed that the user meets the first wavefront at time $t = 0$ and the second one at time $t = T'$. During this time period the receiver moves by the distance $v_{RX}T'$. Since the second wavefront is supposed to reach the receiver at time $t = T'$, it must have already travelled the remaining distance cT'. From the diagram it is clear that $\lambda = cT' + v_{RX}T'$. It directly implies that

$$T' = \frac{\lambda}{v_{RX} + c} \implies f'_{RX} = 1/T' = f_{TX}\left(\frac{c + v_{RX}}{c}\right), \tag{3.33}$$

which is the same result as provided by (3.32).

Intuitively, the receiver moving in the direction towards the transmitter interacts with the wave faster than before and thus it observes a particular value of the signal phase more often than in the static case. The received frequency is increased because the faster phase change leads to higher frequency.

3.5.1.2 Moving Transmitter and Static Receiver

Now the situation is different since the receiver is supposed to be at rest while it is being approached by the transmitter with the constant speed v_{TX}. Similarly as in the previous case, the key idea behind derivation of the Doppler effect in this situation is to discover the change of phase at the receiver.

At the beginning, it is assumed that the transmitter is located at the position $z = 0$ and receiver at $z = z_0$, as depicted in Fig. 3.26. The transmitter starts moving at time $t = t_0$ and at the same time it also starts generating the radio wave with the initial phase $\phi_1 = \phi(t_0, 0) = 2\pi f_{TX} t_0$. Now the question to answer is at which time this particular phase of the wave reaches the receiver located at position z_0. Since the wave propagates with speed of the light, it arrives to the receiver at time $t_1 = t_0 + z_0/c$. This observation can be easily verified by evaluation of the phase $\phi(t, z)$ at time $t = t_1$ and position $z = z_0$ leading indeed to $\phi(t_1, z_0) = 2\pi f_{TX} t_0 = \phi_1$. So, the summary up to this point is that at the time t_1 the receiver experiences the signal phase ϕ_1.

Now, during the following time period, denoted as Δt, the transmitter moves over the distance $v_{TX}\Delta t$. Since the phase of the wave being generated by the transmitter is solely dictated by its local oscillator, the current phase does not depend on the transmitter position at all. Therefore, at time $t = t_0 + \Delta t$ the instantaneous phase of generated wave is $\phi_2 = \phi(t_0 + \Delta t, 0) = 2\pi f(t_0 + \Delta t)$. When does this phase of the wave reach the receiver at the position $z = z_0$? It is clear that the wave needs to travel distance $(z_0 - v_{TX}\Delta t)$. Thus, it reaches the receiver at time $t_2 = t_0 + \Delta t + (z_0 - v_{TX}\Delta t)/c$. Hence it can be seen that at the time t_2 the receiver observes the signal with phase ϕ_2.

Therefore, the receive experiences the temporal phase change equal to

$$\frac{\phi_2 - \phi_1}{t_2 - t_1} = \frac{2\pi f_{TX}(t_0 + \Delta t)\left(\frac{c}{c - v_{TX}}\right) - 2\pi f_{TX} t_0 \left(\frac{c}{c - v_{TX}}\right)}{\Delta t}. \tag{3.34}$$

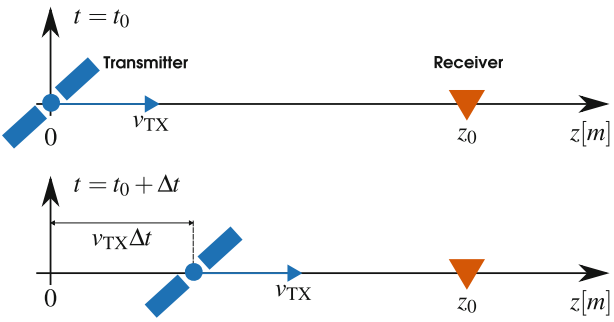

Fig. 3.26 An illustration of the scenario with a static receiver and moving transmitter which is moving towards the receiver by the speed v_{TX}

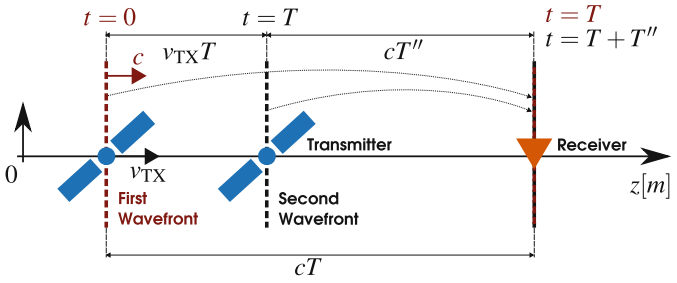

Fig. 3.27 An illustration supporting an explanation of the Doppler effect for moving transmitter

Now, an auxiliary phase function is defined as $\phi''(t) = 2\pi f_{\mathrm{TX}} t \left(\frac{c}{c-v_{\mathrm{TX}}}\right)$. Plugging it into (3.34) yields

$$\frac{\phi_2 - \phi_1}{t_2 - t_1} = \frac{\phi''(t_0 + \Delta t) - \phi''(t_0)}{\Delta t}. \tag{3.35}$$

When $\Delta t \to 0$, the right-hand side of (3.35) is nothing but the definition of the first derivative of $\phi''(t)$ at the point $t = t_0$. Since the first derivative of the phase divided by the factor 2π gives a frequency, it turns out that the frequency of the signal observed by the static receiver is

$$f''_{\mathrm{RX}} = \frac{1}{2\pi} \frac{\mathrm{d}\phi''(t)}{\mathrm{d}t}\bigg|_{t=t_0} = \frac{1}{2\pi} \lim_{\Delta t \to 0} \frac{\phi''(t_0 + \Delta t) - \phi''(t_0)}{\Delta t} = f_{\mathrm{TX}} \left(\frac{c}{c - v_{\mathrm{TX}}}\right). \tag{3.36}$$

Therefore, it can be seen that the transmitter moving towards the static receiver increases the frequency of the observed signal, and vice versa.

A simplified explanation is based on the following reasoning together with the help of Fig. 3.27. First, the transmitter emits a wavefront with signal period T. Afterwards, during the period duration, the wavefront moves to point in distance cT from the origin. It is assumed that the receiver is located exactly at this distance. While the signal has been travelling, the transmitter moved over the distance $v_{\mathrm{TX}}T$. When the transmitter reaches its new position, it starts generating the wavefront with exactly the same phase as generated at the origin. This is due to the fact that the signal period is just equal to T. The new wavefront reaches the receiver exactly T'' s after the previous wavefront. Therefore, from the receiver point of view, it seems that the signal period is equal to T''. From the diagram it is clear that $cT'' = (cT - v_{\mathrm{TX}}T)$. It directly implies that

$$T'' = T\left(\frac{c - v_{\mathrm{TX}}}{c}\right) \implies f''_{\mathrm{RX}} = \frac{1}{T''} = f_{\mathrm{TX}} \left(\frac{c}{c - v_{\mathrm{TX}}}\right), \tag{3.37}$$

which provides the same result as (3.36).

3.5.1.3 Considering the Special Relativity

Einstein's second postulate of special theory of relativity says that the speed of light in vacuum is the same in all inertial frames of [23]. In the previous derivation of the Doppler effect, this principle was used only partially when it was considered that the electromagnetic wave emitted by the moving transmitter propagates in vacuum with the speed of light (and not as the sum of this speed and actual speed of the transmitter). There is, however, another effect directly related to this postulate. The name of the effect is the *time dilation*. Consider a clock running in an inertial reference frame (for example at a spacecraft) and an observer located in another reference frame (for example on the Earth). Assume that the frames are moving with constant speed v with respect to each other. Then the effect basically states that, according to the observer, the clock runs slower (with respect to a nominal clock rate measured in the clock reference frame) by the factor

$$\gamma(v) = \frac{1}{\sqrt{1 - v^2/c^2}}. \tag{3.38}$$

Now we use the factor to correct the previously derived Doppler shift equations. The case of *moving transmitter and static receiver* is considered first. It is assumed that the transmitter is equipped with a clock with the period of one tick equal to T_0. The transmitter exploits this clock to generate the transmitted harmonic wave with the frequency $f_{\mathrm{TX}} = 1/T_0$. If the transmitter and receiver would be static with respect to each other, then the receiver would be able to measure the period T_0 exactly. However, when the transmitter moves, then it appears in the reference frame of the receiver that all the physical processes happening in the transmitter, including the clock period, are slower. Particularly, it appears that the frequency emitted by the moving transmitter is not f_{TX} but rather $f''_{\mathrm{TX}} = 1/\left(T_0\gamma(v_{\mathrm{TX}})\right) = f_{\mathrm{TX}}/\gamma(v_{\mathrm{TX}})$. Replacing f_{TX} by f''_{TX} in (3.36), the expression for the frequency of the received signal reads as

$$f''_{\mathrm{RX,rel}} = f''_{\mathrm{TX}}\left(\frac{c}{c - v_{\mathrm{TX}}}\right) = f_{\mathrm{TX}}\left(\frac{c}{c - v_{\mathrm{TX}}}\right)\sqrt{1 - \frac{v_{\mathrm{TX}}^2}{c^2}} = f_{\mathrm{TX}}\sqrt{\frac{c + v_{\mathrm{TX}}}{c - v_{\mathrm{TX}}}}. \tag{3.39}$$

Now the focus is on case of the *static transmitter and moving receiver*. To facilitate the following explanation, an additional static observer is placed in the reference frame of the transmitter. The overall situation is depicted in Fig. 3.28, in which the observer is represented by a green dot. Consider that the receiver is moving during a limited time period which is given by one tick of its internal clock. The observer watches this movement and it measures that the receiver needs T_0 s to travel between points z_0 and z_1. Therefore, it appears to the observer that one tick of the receiver's clock takes time T_0.

Due to the movement, all receiver's physical processes appear to the observer to run slower than as they are experienced by the receiver itself. Therefore, one

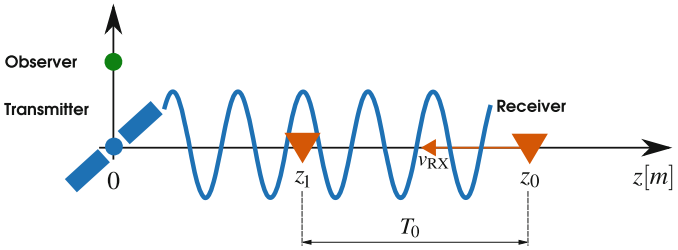

Fig. 3.28 An illustration of the scenario with the static transmitter and receiver that is moving towards the former by the constant speed v_{RX}

tick of the clock period measured by the observer is longer than the clock period as measured in the reference frame of the moving receiver. Thus, one tick of the receiver's clock takes less than T_0, particularly $T_0/\gamma(v_{RX})$, seconds.

When the receiver moves between the points z_0 and z_1, it passes a certain number of signal periods, as illustratively shown in Fig. 3.28. This number is the same irrespectively if it is counted by the observer or by the receiver itself. Indeed, if we say, for example, that the receiver's antenna is just capturing the wave maximum, then both, the observer and receiver, agree on that [3]. Thus, according to the observer, the receiver is passing a given number of periods during the time T_0. The receiver is experiencing the same number of periods but now in shorter time interval equal to $T_0/\gamma(v_{RX})$. Since the frequency of received signal can be determined as the number of passed signal periods over the corresponding time interval, the receiver measures higher frequency of the received signal than the observer by the factor $\gamma(v_{RX})$.

If the time dilation would be ignored, the duration of the clock tick on the receiver would be T_0 instead of $T_0/\gamma(v_{RX})$, and the formula for the received frequency would be (3.32). To incorporate the relativistic effect, the received frequency is just increased by the factor $\gamma(v_{RX})$. Therefore, the frequency of the received signal is

$$f'_{RX,rel} = \gamma(v_{RX}) f_{TX} \left(\frac{c + v_{RX}}{c} \right) = f_{TX} \frac{\left(\frac{c+v_{RX}}{c} \right)}{\sqrt{1 - \frac{v_{RX}^2}{c^2}}} = f_{TX} \sqrt{\frac{c + v_{RX}}{c - v_{RX}}}. \tag{3.40}$$

It can be clearly seen that Eqs. (3.39) and (3.40) are the same. Thus, they perfectly obey the Einstein's first postulate of relativity that states that the laws of physics are the same in every inertial frame of reference. Therefore, it is not necessary to distinguish explicitly who is moving and who is at rest, the only thing that matters is the relative speed of the transmitter and receiver. Finally, the general formula of the Doppler effect including the special relativity is given by

$$f_{RX} = f_{TX} \sqrt{\frac{c + v}{c - v}} = f_{TX} \eta, \tag{3.41}$$

where the Doppler factor η, including all the speed parameters, is defined. The speed sign convention followed in the equation says that if the speed v takes a positive value then the objects (receiver and transmitter) are approaching each other.

Please note that Eqs. (3.39) and (3.40) can be derived directly from the model of the plane wave (3.30) using the Lorentz transformation, as shown in [3].

3.5.2 *Practical Values of the Doppler Shift in GNSS*

To be able to exploit a navigation signal from a navigation satellite, the GNSS receiver needs to determine the value of the corresponding frequency shift f_D which affects the signal. The frequency shift due to the Doppler effect is defined simply as difference between the received and transmitted frequency and therefore

$$f_D = f_{\mathrm{RX}} - f_{\mathrm{TX}}. \tag{3.42}$$

The aim of this section is to provide an interval of values into which the Doppler frequency shift falls. In other words, it attempts to answer a question what is the maximum and minimum Doppler shift experienced by the GNSS receiver. This information is valuable when the GNSS receiver initially searches for the navigation signals since it limits the frequency search space.

The analysis provided here is overall simplistic, however, it still provides reasonable results. It assumes a static receiver located on static Earth. The flattening of the Earth at the poles is neglected. To further ease the mathematical derivation, the satellite trajectory around the Earth is simplified by considering a circular instead of elliptical orbit. This approach can be justified by small eccentricity of all the orbits used by satellite navigation systems. For example, eccentricity for GPS, GLONASS, Galileo, and BeiDou-M navigation orbital trajectories are less than 0.023, 0.004, 0.001, and 0.003 respectively [17].

The model is depicted in Fig. 3.29, where $R = 6370\,\mathrm{km}$ is the Earth radius, A is the altitude of the satellite circular orbit above the Earth and $r = R + A$ is the orbit radius, d is the distance between receiver and transmitter, and α is an angle between z-axis and a line segment connecting the mass center of Earth with the satellite on the orbit.

The Doppler effect is solely determined by the radial speeds of objects. The term *radial* is of high importance since it means only the relative speed between these objects. In other words, the radial speed is the change of the distance between transmitter and receiver per time unit. The distance d can be expressed as function of α as follows:

$$d(\alpha) = \sqrt{(R + r\cos\alpha)^2 + (r\sin\alpha)^2}. \tag{3.43}$$

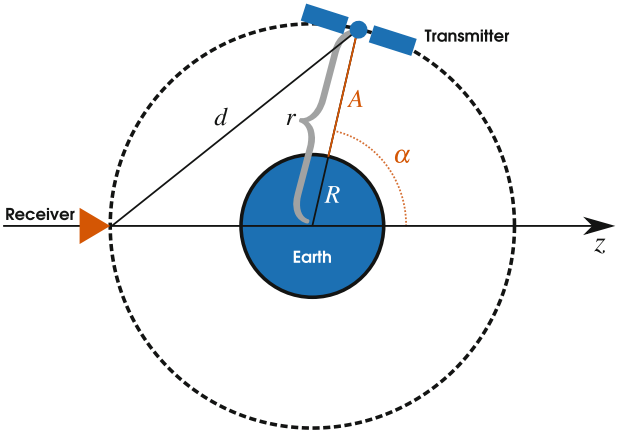

Fig. 3.29 A simplified model of the GNSS satellite orbit and Earth

To include the notion of time, the angle α is simply expressed using the orbital angular frequency ω_P as

$$\alpha = \omega_\mathrm{P} t = \frac{2\pi}{T_\mathrm{P}} t, \tag{3.44}$$

where T_P is the orbital period of the satellite. Plugging (3.44) into (3.43), the distance d as function of time reads

$$d(t) = \sqrt{(R + r\cos\omega_\mathrm{P} t)^2 + (r\sin\omega_\mathrm{P} t)^2}. \tag{3.45}$$

Now, the radial speed v_TX is given by the first derivative of (3.45) with respect to time. However, the negative temporal change of the distance means that the satellite is moving towards the receiver. Since it is desired to keep the speed positive during this direction of movement, the radial speed is actually defined here with the negative sign of the first derivative. It reads

$$v_\mathrm{TX}(t) = -\frac{\mathrm{d}d(t)}{\mathrm{d}t} = \frac{Rr\omega_\mathrm{P}\sin\omega_\mathrm{P} t}{\sqrt{(R + r\cos\omega_\mathrm{P} t)^2 + (r\sin\omega_\mathrm{P} t)^2}}. \tag{3.46}$$

Figure 3.30 shows the radial speed according to (3.46) as a function of α. The underlying orbital parameters are defined in Table 3.1. It can be seen that radial speed is symmetric around $\alpha = 180°$ with extremes around 100° and 260° where the distance varies fastest. Numerical values of the maximum radial speed v_TX taken from the graph are summarized in Table 3.2. This table also shows the maximum carrier frequencies used in a given navigation system. Plugging the maximum radial speed together with the maximum carrier frequency into (3.41) yields the received frequency and also the Doppler frequency shift $f_D = f_\mathrm{RX} - f_\mathrm{TX}$.

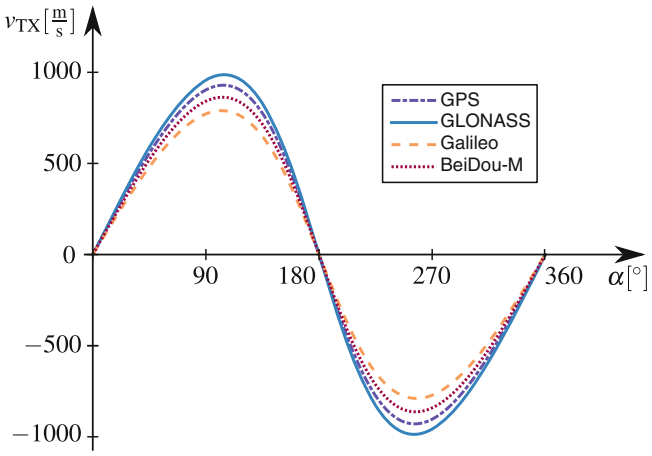

Fig. 3.30 Radial speed

Table 3.1 Orbital parameters based on [17]

Parameter/GNSS	GPS	GLONASS	Galileo	BeiDou-M
Altitude A [km]	20,182	19,130	23,222	21,528
Period T_P [min]	718	676	845	773

Table 3.2 Numerical results of Doppler shift calculation

Parameter/GNSS	GPS	GLONASS	Galileo	BeiDou-M
Maximum radial speed max$\{v_{TX}\}$ [m/s]	929	987	789	863
Maximum carrier frequency in the system f_{TX} [MHz]	1575.420	1605.375	1575.420	1589.742
Maximum carrier frequency of received signal f_{RX} [MHz]	1575.425	1605.380	1575.424	1589.747
Maximum Doppler shift f_D [kHz]	4.879	5.281	4.146	4.573

It can be seen from Table 3.2 that the Doppler shift estimated based on the simple model falls roughly into the range ±5 kHz. However, it should be highlighted that the receiver is supposed to be static. Therefore, to obtain some reserve and hence higher confidence that the actual or real value of the Doppler shift actually falls into the interval, the speed of receiver is now considered. The maximum radial speed of the satellite towards the receiver is roughly 1000 m/s. If the same value would be at the same time also assumed for the radial speed of the receiver towards the satellite, it would result in Doppler shift in a range ±10 kHz. Since 1000 m/s is relatively high speed already (roughly three times more than the speed of sound), this extended interval is in most cases sufficiently confident.

3.5.3 *Impact of the Doppler Effect on the Received Signal*

The previous elaboration on Doppler effect implicitly assumed the signal in form of pure harmonic waveform. It was shown that the Doppler effect effectively changes its carrier frequency as observed by the moving receiver. The carrier frequency change is modeled in the baseband model (3.2) simply as multiplication of the signal $s(t-\tau)$ by complex exponential $\exp\{j2\pi f_D t\}$. However, is this really the entire influence of the Doppler effect on the signal? The simple answer is *no* and this statement is investigated in this section.

The explanation starts by considering the transmitted signal in form of two harmonic signals, each with different frequencies f_1 and f_2. Assume that the difference between these two frequencies is equal to $\Delta f = f_2 - f_1$. Due to the principle of superposition both harmonic components of this signal are affected by the Doppler effect individually. So each component is shifted in frequency domain separately. The received frequencies would be $f_1\eta$ and $f_2\eta$, where the factor η depending on radial speeds of transmitter and receiver is defined in (3.41). It is evident that the frequency difference between the received harmonic components is equal to $\Delta f\eta$. This simple observation is illustrated in Fig. 3.31.

It is, however, more interesting to think about a signal with the continuous frequency spectrum with bandwidth B. How is this overall signal spectrum affected by the Doppler effect? The theory of Fourier transform says that a signal can be represented as an integral of weighted harmonic functions. Therefore, if each harmonic function is treated separately, the explanation mentioned above for the case of the two harmonic signals can also be applied in this case. Thus, from the illustration in Fig. 3.31 it is clear that the received signal bandwidth is extended or shrinked by the Doppler factor η.

This effect in the frequency domain has its own counterpart in the time domain. It is well known from the Fourier analysis that a signal with wide spectrum is sharp

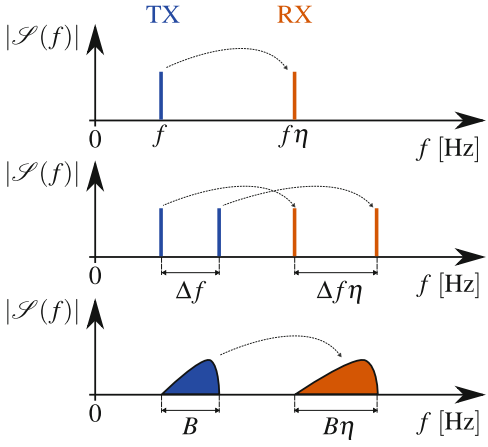

Fig. 3.31 The demonstration of the impact of the Doppler effect on the received signal

in time domain (for example, Dirac delta function) and vice versa (for example, a harmonic signal). Thus the stretching of the signal spectrum speeds up the signal in time domain, and shrinking it slows it down.

3.5.4 GNSS Signal Model Revision

Now it is time to revise the signal model introduced in (3.2). The modelling of the Doppler effect by the complex exponential only considers the frequency shift of the carrier frequency. It completely ignores the influence of Doppler effect on the overall signal bandwidth. Whether the effect is significant or not depends on whether the signal bandwidth can be considered narrow or not. The narrow-band definition depends not only on the actual bandwidth or the ratio between bandwidth and carrier frequency, but mainly on the purpose of the analysis or application for which the model is supposed to be used. If the effect on the bandwidth can be ignored, then the model (3.2) is still applicable and it is known as the model employing a narrow-band approximation.

If this approximation is not acceptable, then the model (3.2) must be improved. Before the model extension is discussed, the Doppler effect on the signal bandwidth is analyzed more carefully in the following. For the mathematical convenience, the analysis is performed in the baseband representation, but this has no impact on its result.

The analysis focuses on a signal $c(t)$ representing a pseudorandom sequence $\boldsymbol{c}$ modulated using a BPSK modulation with rectangular modulation pulse as discussed in Sect. 3.2. For the sake of clarity, only one chip from the overall sequence is analyzed. The considered chip, denoted as $p(t)$, lies on the interval $[-T_c/2, T_c/2)$ as depicted on the left side in Fig. 3.32.

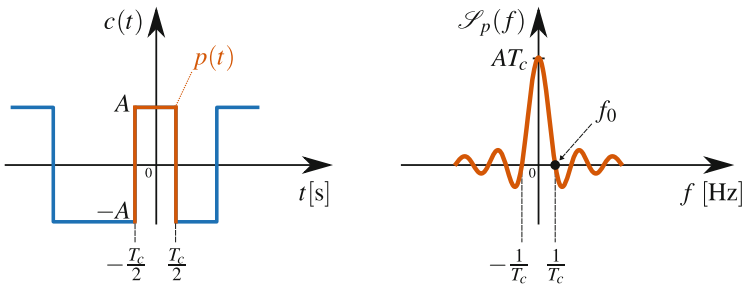

Fig. 3.32 Spectral analysis of the pulse $p(t)$. On the *left side* there is the pulse in time domain, on the *right side* there is its frequency spectrum

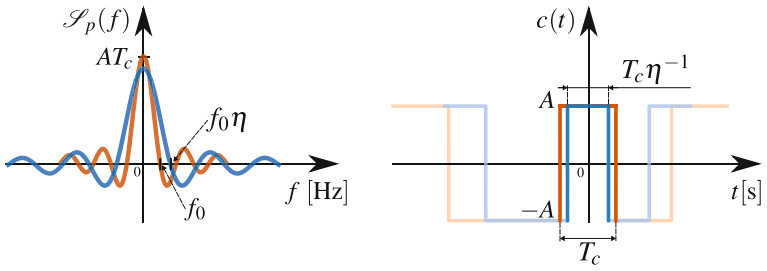

Fig. 3.33 An impact of Doppler shift in frequency as well as time domains

The spectrum of that chip is given by its Fourier transform

$$\mathscr{S}_p(f) = \int_{-\infty}^{\infty} p(t)e^{-j2\pi ft}dt = A\int_{-\frac{T_c}{2}}^{\frac{T_c}{2}} e^{-j2\pi ft}dt = \left[\frac{e^{-j2\pi f\frac{T_C}{2}} - e^{-j2\pi f\frac{-T_C}{2}}}{-j2\pi f}\right]_{-\frac{T_c}{2}}^{\frac{T_c}{2}}$$
$$= AT_C\frac{\sin(\pi fT_C)}{\pi fT_C} = AT_C\mathrm{sinc}(fT_C) \tag{3.47}$$

and it is depicted on the right side in Fig. 3.32. The spectrum takes zero values for the frequencies satisfying the condition $fT_c = n$, where $n \in \mathscr{N}$. Let the frequency f_0 denote the position of the first spectral null. It is evident that $f_0 = 1/T_c$. This also means that the chip duration can be directly determined by the position of the first spectral null and, therefore, there is a clear and simple link between the chip period and signal spectrum, as shown in Fig. 3.33. Under the Doppler effect, the overall spectrum is extended or shrunk based on the factor η. The position of the first null is moved to frequency $f_0\eta$ and, hence, the chip duration changes to its new value equal to

$$T_c' = \frac{1}{f_0\eta} = \frac{T_C}{\eta}. \tag{3.48}$$

There are two conceptual possibilities to incorporate this effect into model (3.2). First, the chip duration T_c can be simply treated as an additional explicit model parameter of the modulation pulse $h(t)$. For example, like $h(t, T_c)$, in which the chip duration can be controlled directly according to (3.48).

The second option does not require the additional input argument to the modulation pulse. In this case, the pulse is still based on the nominal value of T_c, meaning that this parameter is constant and is not adapted according to the Doppler effect at all. Instead, the time variable t is "sped up" or "slowed down" in order to model the change of the chip duration. The explanation is based on Fig. 3.34. In the absence of the Doppler effect, it is clear that the function $h(t)$ ends exactly at time $t = T_c$. However, due to the Doppler effect it appears that it can end before or after

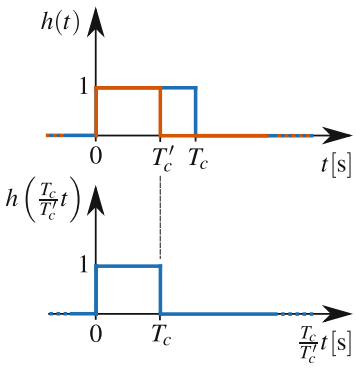

Fig. 3.34 Incorporating the impact of the Doppler effect into the modulation pulse $h(t)$. The period T_c is nominal chip duration, whereas T_c' denotes the changed duration due to the Doppler effect

the nominal time T_c. In this example it is assumed that the pulse is compressed and therefore the falling edge of the pulse should be reached earlier than in the nominal case, particularly at time T_c'. This can be achieved simply by multiplying the time argument t of the nominal modulation pulse $h(t)$ by a factor $T_c/T_c' = \eta$. Indeed, then the modulation pulse $h(\eta t)$ ends at time $t = T_c'$ as expected. This is depicted in the lower part of Fig. 3.34.

Therefore, the extended baseband model (3.2) transforms into the following form

$$x(t) = A\, s(\eta t - \tau) e^{j2\pi f_D t + j\phi(t)} = A\, s\left(\left\{1 + \frac{f_D}{f_{TX}}\right\} t - \tau\right) e^{j2\pi f_D t + j\phi(t)}. \tag{3.49}$$

Whether this extension of the signal model is necessary depends on an application. For example, in [18] it is pointed out that this extension is essential for GNSS-based indoor positioning in which the navigation signals with relatively long duration has to be processed. If the model does not fully consider the Doppler effect as explained above, then it might be interesting to determine, after which time period the received signal differs from the model by time misalignment equal to T_c. For example for the GPS L1 C/A code with 5 kHz Doppler shift this happens approximately after 0.3 s.

3.6 Conclusions

This chapter has provided some basic explanations related to the model of GNSS signals and discussed several aspects that are usually silently skipped in an advanced literature on GNSS. The chapter dealt with the link between passband and baseband signal representations with the main focus on the noise modelling. It also provided thorough explanation of Doppler effect and its full impact on the navigation signal. The chapter leaves the navigation signal at the point when it is ready for the actual signal processing algorithms such as acquisition, tracking, and position estimation. For further reference, these algorithms are explained in classical

books such as [8, 12, 16, 19]. The underlying principles behind the algorithms such as signal, detection, and estimation theories can be found, for example, in [1, 4, 9, 10, 14, 15, 20].

Acknowledgements This work was financially supported by EU FP7 Marie Curie Initial Training Network MULTI-POS (Multi-technology Positioning Professionals) under grant nr. 316528.

References

1. S. Benedetto, E. Biglieri, *Principles of Digital Transmission: With Wireless Applications*. Information Technology Series (Springer, New York, 1999). ISBN:9780306457531
2. L.V. Blake, Antenna and receiving-system noise-temperature calculation. Naval Research Lab, Washington DC (1961)
3. R.P. Feynman, R.B. Leighton, M.L. Sands, *The Feynman Lectures on Physics*. The Feynman Lectures on Physics, vol. 3 (Pearson/Addison-Wesley, Reading, MA, 1963). ISBN:9780805390490
4. W.A. Gardner, *Introduction to Random Processes: With Applications to Signals and Systems* (MacMillan, New York, 1986). ISBN:9780029487907
5. D.J. Griffiths, *Introduction to Electrodynamics* (Pearson Education, Harlow, 2014). ISBN:9780321972101
6. J.D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1998). ISBN:9780471309321
7. A. Joseph, Measuring GNSS signal strength, in *Inside GNSS*, November/December 2010, pp. 20–25
8. E. Kaplan, *Understanding GPS - Principles and Applications*, 2nd edn. (Artech House, Boston, 2005)
9. S.M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory* (Prentice-Hall PTR, Englewood Cliffs, 1993), p. 595. ISBN:0133457117
10. S.M. Kay, *Fundamentals of Statistical Signal Processing Volume II: Detection Theory* (Prentice-Hall PTR, Englewood Cliffs, 1998), p. 560
11. R.B. Langley, GPS receiver system noise, in *GPS World* (1997), pp. 40–45
12. P. Misra, P. Enge, *Global Positioning System: Signals, Measurements, and Performance* (Ganga-Jamuna Press, Lincoln, MA, 2001)
13. J. Nurmi et al., *GALILEO Positioning Technology*. Signals and Communication Technology (Springer, Netherlands, 2014). ISBN:9789400718296
14. A.V. Oppenheim, A.S. Willsky, I.T. Young, *Signals and Systems*. Prentice-Hall Signal Processing Series (Prentice-Hall, Englewood Cliffs, 1983)
15. A. Papoulis, S.U. Pillai, *Probability Random Variables, and Stochastic Processes*. McGraw-Hill Series in Electrical Engineering: Communications and Signal Processing (Tata McGraw-Hill, New York, 2002). ISBN:9780070486584
16. B.W. Parkinson, J.J. Spilker, *Global Positioning System: Theory and Applications, Volume 1*, vol. 1 (American Institute of Aeronautics and Astronautics, Washington, DC, 1996). ISBN:1600864198
17. A. Rossi et al., *Disposal Strategies Analysis for MEO Orbits*. Tech. rep. ESA, General Studies Programme, Internal Number: 12-604-01, 2015, p. 43
18. G. Seco-Granados et al., Challenges in indoor global navigation satellite systems: unveiling its core features in signal processing. IEEE Signal Process. Mag. **29**(2), 108–131 (2012). ISSN:1053–5888. doi:10.1109/MSP.2011.943410
19. J.B.-Y. Tsui, *Fundamentals of Global Positioning System Receivers: A Software Approach* (Wiley, New Jersey, 2005), p. 352. ISBN:0471712574
20. H.L. Van Trees, *Detection, Estimation, and Modulation Theory, Part 1* (Wiley, Hoboken, 2004). ISBN:0471463825

21. H.L. Van Trees, *Detection, Estimation, and Modulation Theory, Radar-Sonar Signal Processing and Gaussian Signals in Noise*. Detection, Estimation, and Modulation Theory (Wiley, New Jersey, 2004). ISBN:9780471463818
22. R.K. Wangsness, *Electromagnetic Fields* (Wiley, New Jersey, 1986). ISBN:9780471811862
23. H.D. Young et al., *University Physics, Nide 1* (Wiley, New York, 1996). ISBN:0201571552
24. W. Zhang, M.J. Miller, Baseband equivalents in digital communication system simulation. IEEE Trans. Educ. **35**(4), 376–382 (1992). ISSN:0018–9359. doi:10.1109/13.168713

Chapter 4
GNSS Vulnerabilities

Susana María Sánchez-Naranjo, Nunzia Giorgia Ferrara, Maciej Jerzy Paśnikowski, Jussi Raasakka, Enik Shytermeja, Raúl Ramos-Pollán, Fabio Augusto González Osorio, Daniel Martínez, Elena-Simona Lohan, Jari Nurmi, Manuel Toledo López, Ondrej Kotaba, and Olivier Julien

4.1 Introduction and Motivation

GNSS positioning principle is based on the solution of an elemental geometric problem, in which the receiver is located at the intersection point of four spheres, each one centered in one satellite with known coordinates, and with radius equal to the satellite receiver distance, or pseudorange, as shown in Fig. 4.1 [32].

S.M. Sánchez-Naranjo (✉) • D. Martínez
Pildo Labs, Carrer Marie Curie 8-14, 08042 Barcelona, Spain
e-mail: smsanchezn@unal.edu.co

N.G. Ferrara • E.-S. Lohan • J. Nurmi
Tampere University of Technology, Korkeakoulunkatu 10, 33720 Tampere, Finland
e-mail: elena-simona.lohan@tut.fi; jari.nurmi@tut.fi

M.J. Paśnikowski • M.T. López
GMV Aerospace and Defence S.A.U., Calle Isaac Newton 11 P.T.M., 28760 Tres Cantos, Madrid, Spain

J. Raasakka • O. Kotaba
Honeywell International s.r.o., V Parku 2325/16, 148 00 Prague, Czech Republic

E. Shytermeja • O. Julien
Telecom Lab, SIGNAV Research Group, École Nationale de l'Aviation Civile, 7 Avenue Edouard Belin, 31055 Toulouse Cedex 4, France

R. Ramos-Pollán
Unidad de Supercomputación y Cálculo Científico, Universidad Industrial de Santander, Carrera 27 Calle 9 Ciudad Universitaria, Bucaramanga, Colombia

F.A.G. Osorio
Facultad de Ingeniería, Universidad Nacional de Colombia, Cra 30 No. 45 03, Building 453 Office 114, Ciudad Universitaria, Bogotá, Colombia

J. Nurmi et al. (eds.), *Multi-Technology Positioning*,
DOI 10.1007/978-3-319-50427-8_4

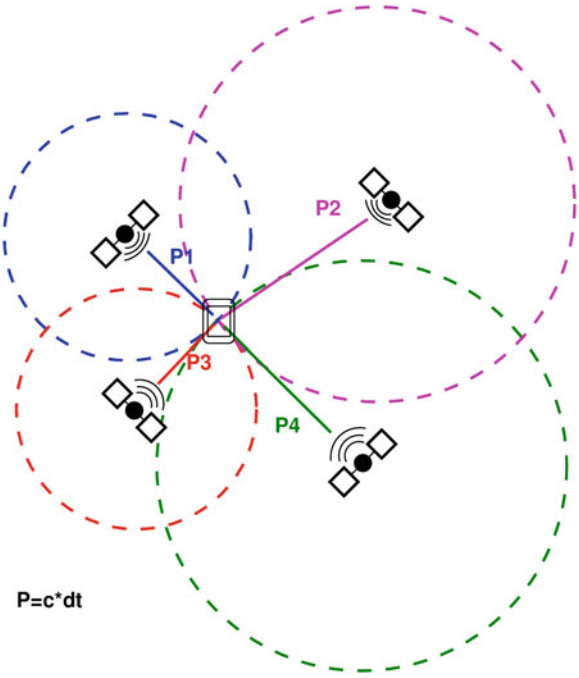

Fig. 4.1 GNSS positioning is based on solving the geometric problem in which the receiver position is unknown, while the pseudoranges to four or more satellites and the satellites positions are known

Distances between satellites and receiver are calculated by using the propagation time of an electromagnetic wave along the path joining them. This travel time, multiplied by the speed of light, provides a measure of their separation. In an ideal situation the clocks are perfectly synchronized, satellites orbits are perfectly known, signal propagates at vacuum light speed, there are no obstacles in the signal path, and there are no interferences. Any deviation from this ideal scenario can contribute to errors in distance measurements. Error sources in GNSS are [32]:

1. Satellite clock: Synchronization is a critical factor as one nanosecond of satellite clock inaccuracy results in a distance measurement error of about 30 cm to that satellite. Even though they are provided with very accurate (Cesium) atomic clocks, there are drift and noise which are modeled and included as part of the message provided by the satellite.
2. Ephemeris error: Satellites orbits are continuously monitored from several stations located around the earth, and their expected location is transmitted to the satellites, which they forward to the receivers. The accuracy of the orbital prediction is in the order of a few meters, which will create about a few meters of error in computing position.

3. Dilution of Precision: The geometric arrangement of satellites, as they are presented to the receiver, affects the accuracy of the position and time solution. Receivers will ideally be designed to use signals from available satellites in a manner that minimizes this effect.
4. Atmospheric effects: GNSS signals pass through the near-vacuum of space, and then they go through the different layers of the atmosphere to the earth. Electromagnetic waves travel in a straight path only in vacuum or in perfectly homogeneous media; any substance or imperfection in the propagation media causes a deviation from the shortest path, increasing the travel time and affecting the distance calculations. Travel through atmosphere introduces this kind of errors in propagation time. For GNSS applications, the atmospheric regions of interest are mainly two: troposphere and ionosphere. Troposphere ranges from the Earth surface to 17–20 km in altitude, whereas ionosphere ranges from 75 to 500 km.
5. Multipath: This effect relates to arrival of signals by two or more paths, due to multiple reflections. The multiple signals cause constructive and destructive interference in the receiving antenna, affecting, hence, the received signal quality.
6. Interference: Satellite signals arrive to earth surface with a low power, hence external signals can interfere in a great extent.

GNSSs have applications in a broad spectrum of fields, ranging from military usage to commercial services such as surveying, mapping, agriculture, transportation, machine control, marine navigation, vehicle navigation, mobile communication, and many others [30]. Such level of involvement demands a higher performance of GNSS-based systems. The assessment and treatment of error sources are vital for delivering a service satisfying the growing needs of each sector.

Table 4.1 presents the typical error ranges for different sources. It can be seen from the table that ionospheric error is a major error source. On the other hand, multipath is highly enhanced for urban environments, and interference sources are also a concern for GNSS users due to the increase of potential interference generators. This chapter presents details on ionosphere, multipath, and RF Interference error sources and the mitigation strategies that can be used to mitigate them.

Table 4.1 Typical impact in GNSS position calculation due to different error sources

Contributing source	Error range (m)
Satellite clock	±2
Ephemeris error	±2.5
Dilution of precision	±2
Ionospheric delays	∼10
Tropospheric delays	±0.5
Multipath	±1

4.2 Ionospheric Error

4.2.1 What is the Ionospheric Error?

Ionosphere is an atmospheric layer ranging from 75 to 500 km, and it is characterized by a large concentration of free electrons that interact with the travelling signal.

Ionosphere can affect GNSS signal in two different ways [51]: First, the ionosphere electron content causes a variation in the medium refraction index, and consequently, in the wave propagation velocity and the signal travel time. Since the calculation of distance from satellite to receiver is based on such travel time, there is an error in the range measurements that can vary from less than 1 m to more than 100 m. Second way is due to irregularities in ionosphere, and it is known as scintillation effect. Such irregularities can produce diffraction as well as refraction, causing short-term signal fading that impacts the receiver tracking capabilities [35]. The scintillations amplitude may be so intense that the receiver can lose lock to the signal, and if a number of satellites are affected simultaneously, positioning may even become impossible. This phenomenon affects GNSS signal specially in low latitude regions (around the equator) [33].

The delay caused on signal by ionosphere is proportional to the Total Electron Content (TEC) met by the signal on its path from the satellite to the receiver. TEC depends on the receiver's geographical location, the hour of day, and the solar activity levels. Figure 4.2 shows the geographical distribution of Vertical TEC on May 2016 [44]. The distribution shows differences in intensity along the latitude lines. Ionospheric activity usually peaks along the sub-equatorial anomaly regions, located on average ∼15° above and below of the geomagnetic equator.

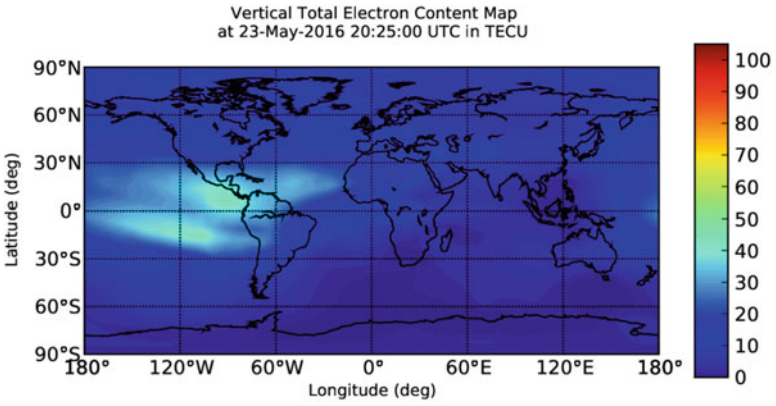

Fig. 4.2 Vertical Total Electron Content in units of 0.1 TECUs ≃ 1.6 cm of delay in the GPS L1 signal at 20:25:00 UT on May 23rd, 2016. Courtesy NASA/JPL-Caltech

Besides the geographic variability, the activity levels change significantly with the epoch within the solar cycle and with the time of day [11]. In particular, characterizations of TEC in low latitude regions such as South America and Asia have found a considerable day-to-day variability [39, 49, 64].

Due to such complex behavior, ionospheric error mitigation has been a challenge for GNSS positioning. Several approaches have been made, including modifications to the receiver architecture, different models used when processing the signals, and the usage of corrections from augmentation systems. Some applications use combinations of several mitigation methods.

4.2.2 Ionospheric Error Mitigation

Different techniques are used to deal with ionospheric error, depending on the applications and/or expected performance levels. Next sections describe the main methods and models, and summarize the existing approaches to mitigation for single frequency users.

4.2.2.1 Dual Frequency Combination

Since ionosphere is a dispersive medium, GNSS signals refraction depends on its frequencies (as the squared inverse). This fact allows to reduce its effect to the first order by using a dual frequency receiver. However, dual frequency devices have a high cost and it is estimated that about 75% of receivers only acquire single frequency [2]. Given this predominance of single frequency receivers, it is necessary to develop mitigation techniques suitable for its use.

4.2.2.2 Physics-Based Data Driven Ionospheric Models

Physical models aim at a simple yet general characterization suitable of being broadcast to single frequency users. The main models are Klobuchar model [34], which was developed for GPS, and NeQuick model [13], developed for Galileo users.

Physical models offer a simple correction, but lack precision at very low or very high latitudes.

4.2.2.3 Ionospheric Maps

Another approach is the creation of maps based on collected data from networks of receivers. The main ionospheric maps used are International Reference Ionosphere (IRI) and GIM. IRI provides monthly averages of the electron density, electron

temperature, ion temperature, and ion composition in the ionospheric altitude range for a specific location, time, and date [7]. GIM provides TEC grids every 2 h, with a resolution of 5° in longitude and 2.5° in latitude[26].

Maps provide a better approach for ionospheric modeling. Disadvantages of maps include a poor performance during years of high solar activity [37], and the low network coverage for some regions, causing that the generated map must rely on interpolation of distant measurements.

4.2.2.4 Ionospheric Data Driven Models

There is a recent increment in data sources and GNSS systems, providing new perspectives to approach the existing problems. Worldwide networks such as the IGS, and regional networks such as GEORED in Colombia [20], REGME in Ecuador [55], EPN in Europe [16], among many others, provide access to large amounts of GNSS data. This allows the generation of purely data-based models generated with machine learning methods that have shown high performance even in high variability scenarios.

Thanks to the availability of data, different machine learning methods have been explored. For example, ANN has been used to model ionosphere in Brazil [38], Puerto Rico [40], and India [65] and predict electron content with high accuracy. MARS have been compared to IGS map GIM [31] and provide higher accuracy. Random Forest Regression [18] has shown better results than physical models and maps for low latitudes.

Machine learning and big data technologies, in conjunction with the large amounts of data available from different sources, allow the development of fast, reliable, and locally adapted models.

4.3 Multipath

Multipath effect in GNSS appears when the signal emitted by satellite arrives at the receiver along multiple propagation paths. This situation can occur when the signal is reflected from surfaces surrounding the receiver. All the GNSS systems were designed to function in ideal line-of-sight (LOS) conditions. The situation becomes more complicated in environments where the GNSS signal arrives distorted, such as urban canyons and indoor locations.

In order to mitigate this effect, it is necessary to formulate a model of multiple path propagation. The most straightforward model of multipath can be defined as a set of discrete reflected signals arriving with larger delays, different amplitudes, and carrier phases compared to the signal of the direct path, as it is described in [32]. Usually, there is a relative motion between receiver, satellites, and the objects reflecting the signal and generating multipath can be taken into account in more advanced models, where the parameters are time-varying. Some of the

models assume that the received carrier frequencies are equal for non-direct and direct propagation paths. This can be helpful, e.g., in urban environments, where a receiver on a vehicle moving with considerable velocity can filter out the reflected signals and the multipath mitigation process can be augmented. Moreover, many reflections and refractions causing the multipath effect make the secondary paths to have longer propagation time and also cause the shift in phase and attenuation of the received signal.

According to [59], one can distinguish two types of multipath:

- *Coherent* multipath—typically specular reflections with the time-delay difference on the order or smaller than inverse of signal bandwidth and a Doppler difference smaller than the difference of coherent correlation level;
- *Noncoherent* multipath—typically diffuse reflections with the time-delay and Doppler-shift larger than coherent one.

In [59] it is also stated that in GNSS only the coherent type has a relevant role, because it introduces a systematic error in the time-delay estimation of the aggregate received signal. On the contrary, noncoherent multipath does not provide any information on the line-of-sight (LOS) time-delay.

Another possible characterization of multipath effect can be the division in terms of the distance—according to [19]:

- *close in* multipath: occurring when the reflecting surface is close to the satellite antenna direct line and the delay is small; best mitigation techniques rely on multipath-limiting antennas, such as choke rings or multibeam antennas; correlator-based multipath mitigation techniques have a bad performance for close in multipath rejection due to the fact that this type of multipath causes a small change in the ideal correlation function;
- *far* multipath: occurring when the reflecting surface is at larger distance from the antenna, such as a building, foliage, and a mountain; best mitigation techniques rely on correlator-based multipath mitigation techniques.

In Fig. 4.3, one can observe the example of degraded correlation peak shape caused by multipath, resulting in deterioration of the positioning accuracy. Multipath errors are relevant to detect and mitigate, because they cannot be eliminated by differencing the observations. What is more, they cannot be generally modeled because of being time- and location-dependent. As a result, the multipath errors in observation domain are not constant in time. According to [60] they show a sinusoidal behavior, due to the change of relative phase between direct and indirect signals during satellite vehicle continuous movement.

Two most important factors that influence the impact of the multipath on the time-delay estimation are the relative power compared to the direct signal and the distribution of the multipath components in time. More detailed description of the influence of the multipath effect on GNSS observables will be presented in the following sections.

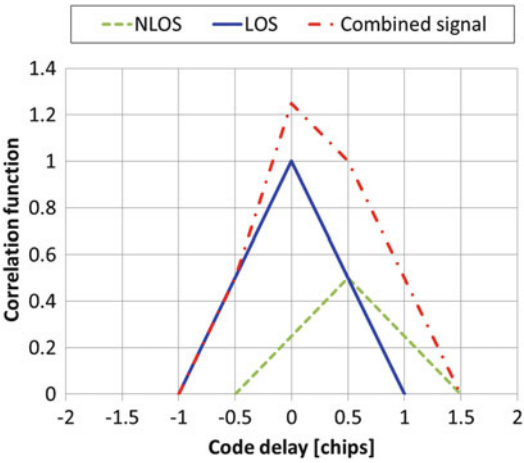

Fig. 4.3 Example of correlation function in two path statistic channel model, in-phase combination; based on [5]

4.3.1 Impact of Multipath on Code Measurements

Generally speaking, code measurements are being formed in the receiver by discrimination-based code tracking loops, usually DLL [5] in correlation process. This process, as shown in Fig. 4.3, can be severely affected by multipath. In one of the very early works on code multipath [48] it was shown that in vicinity of buildings, errors can reach values up to 100 m, and that code phase multipath has non-zero mean, and thus it cannot be simply eliminated by averaging in longer periods. Research in [21] shows that the possible error can reach up to a couple of hundreds of meters.

Most of the works, like [28], describe influence of only one reflector, and this situation is as easy to describe as not realistic, because of multiple reflecting surfaces surrounding the antenna and receiver. In [62] it was shown that the echoed reflections can arrive at the receiver with up to several hundreds of nanoseconds delay. Low-elevation satellites also amplify the multipath in general, the delay shortens and power increases with the increase of elevation. The combination of multiple time-delayed signals can result in significant distortion of correlation peak, and thus in erroneous correlation effect, resulting in error in calculated pseudorange up to 0.4 of chip length, as described in [67]. The multipath effect is less severe for longer time-delays, because it is easier to discriminate the reflected signals from the direct path.

4.3.2 Impact of Multipath on Carrier Phase Measurements

Carrier phase multipath is limited in value. As in [27], it might reach up to $\frac{1}{4}$ of cycle, resulting in approximately 5 cm of error. In [53] there are mentioned several influences of multipath on GPS carrier phase, such as reflectors placed distantly

from the antenna that can cause multipath of high frequency or changing in a fast way, in opposition to close reflectors causing slowly changing effect. It is also mentioned that L1 and L2 carriers are characterized by the same multipath amplitude but different phases, and L1 carrier when compared to L2 reveals the multipath of higher frequency.

What is more, it is mentioned that the far away sources of reflection cause weaker multipaths than those situated close by the antenna, and low-elevation satellites are more probable to cause carrier phase multipath caused by a bigger number of potential reflectors, but require reflecting objects of a greater surface because of large Fresnel zone for strong multipath. This is actually opposite to high elevation satellites, because not only they are less likely to generate multipath, but also they need smaller surface due to generate strong multipath. Code and data bits do not influence the signal power as they hardly ever change the phase; the carrier or Signal-to-Noise Ratio (SNR) with or without data bits remain the same.

What is interesting in the case of carrier phase multipath is that even though the value of the multipath effect is much lower and insignificant compared to code phase multipath, it is still large compared to potential precision of carrier phase observations, its combination in harsh environments with cycle slips can be potentially dangerous. The potential loss of lock in densely urbanized areas is much higher due to presence of obstacles. The combination of these two effects are worth examining and researching, along with the questions of how the number of constellations and visible satellite vehicles would improve the robustness against cycle slips and multipath effect.

4.3.3 Impact of Multipath on Signal Strength and Doppler Frequency Measurements

Information about the signal strength is useful in identifying multipath effect, because carrier-to-noise density (C/N_0) is the only GNSS observable where without any processing of the measurements, e.g., creating double differences of observations, the influence of multipath effect can be directly visible. The C/N_0 values are usually behave like a function of the satellite elevation, but only to level of mask elevation threshold. For angles below certain values, a strongly noised signal can be observed. This curve is mainly determined by antenna gain pattern, as it is described in [6, 56, 60], and this model can be used for identification and assignment of the detected deviations. The qualitative analysis is rather straightforward, whereas the quantitative becomes problematic because of the inconsistencies between the antenna manufacturers in terms of definitions of C/N_0 and algorithms applied, as each of the companies has different ways of calculating these values. Moreover, different receivers use different quantization levels, and very often manufacturers do not officially publish gain pattern of receiving antenna, which cause further problems. In favor of this method it was found out that C/N_0 measurements can help in reducing the influence of the effects of reflection and diffraction on observation level.

The research in the topic of Doppler frequency multipath is barely described in literature, but in general words, Doppler observations, contrary to carrier phase, will not be affected by cycle slips. They can be influenced by noise, but it will not affect them as much as loss of lock. Therefore, Doppler frequency observations can be described as more robust, as it is mentioned in [61].

4.3.4 System Sensitivity to Multipath

It was observed that the choice of modulation type for the signal is relevant for the possibility of mitigation. In the case of GPS L1 Coarse acquisition (C/A) signal, the BPSK modulation causes that the correlation function has only one tracking peak within ± 1 chip delay from the proper code delay. For the different type of modulation—cosine-phased BOC, used to modulate Galileo E1 signal—the correlation function has more peaks within ± 1 chip delay. As stated in [28], it can be noticed that the BOC modulation outperforms BPSK modulation, especially in the long-time delays. This helps in the detection of multipath signal, but still causes problems for the short-time delays. Also a small influence of front-end filter bandwidth can be observed on the shape of correlation function—it occurs that if the bandwidth is not chosen adequately high, it may influence the correlation peak by rounding it and compress the correlation function, like mentioned in [5].

4.3.5 Multipath Mitigation

Mitigation of multipath effect in GNSS can be applied on different levels and by several methods that have been developed in last years.

4.3.5.1 Correlation-Based Mitigation

Mitigation and detection on the receiver level are based on diversity of signal processing methods and strategies. The code tracking architecture in the receiver is classically based on correlating delayed signals, thanks to feedback loop. The first step in the evolution of the correlation-based mitigation techniques was development of the EML technique. It uses two correlators (Early and Late) arranged within a distance of one chip between each other to create a discriminator function, which zero-crossings describe the delay of signal. However, this method was found to be ineffective in mitigation of multipath effect and needed enhancement. Therefore, together with development of receiver signal processing techniques and feedback loops, a whole family of techniques basing on increasing the number and narrowing the distribution of correlators was developed. Other families of techniques focus on parameter estimation, some other on invariant properties of correlation function in the presence in multipath. More advanced techniques have also been developed, out

of which MEDLL applied in NovAtel receivers has been considered as significant milestone in multipath mitigation. It is based on Maximum Likelihood search, mitigates the long delays in an excellent way, but the level of complexity caused by correlators is quite high. Other modern, advanced techniques focus, e.g., on the estimation of the peak of correlation function.

Each mitigation concept can behave differently for specific types of multipath (short or long delays) and particular types of modulation. Very sophisticated methods cause high level of complexity of feedback loops in the receiver, making it less effective in implementation. Very good review of correlation-based techniques can be found in [4].

4.3.5.2 Measurement-Based Mitigation

The classical mitigation technique for code observations is smoothing the signal. The classic smoothing uses carrier phase observations, sometimes called as Hatch-filter, thanks to the R. Hatch and the algorithm developed by him [24]. The concept bases on the lesser influence of multipath on carrier phase observation than on code phase, and then can be used in weighing process to obtain code observations with reduced noise. The results depend on the time of smoothing [28], and the time of 100 s is used typically in aviation [60].

Another possible type of smoothing code phase observations in multipath mitigation is using Doppler frequency in that purpose. It is mentioned in [61] that Doppler observations are more robust to the multipath than carrier phase and they can be used for smoothing. However, the literature in this topic is still not very abundant.

It is also popular to use linear combinations of observations in reducing the effect of multipath. As it was mentioned in [66], the use of wide-lane, iono-free, and geometry-free combinations significantly reduced the multipath influence. However, it was used for the data collected by the regional network of GNSS reference stations, not for real-time processed urban data collected with usually cheap (not geodetic) and non-stationary receiver.

4.3.5.3 Hardware Mitigation

Hardware solutions to multipath problem could also be applied. One of them is the antenna design, and the application of firstly, the large ground plane, changing the set of signals to those from the upper hemisphere and reducing reflections, however resulting with diffraction at the edge of plane; then—introduction of choke ring antennas, attenuating negative signals. The problem in these solutions is the size of antenna. Another antenna design is based on the change of sensitivity for RHCP signals, as the GNSS signals are polarized in this way, and benefitting from the fact that the reflection changes the polarization.

Multi-antenna arrays also allow for mitigating multipath, as they work as a “spatial filtering” of the signal. The application of system of antennas in the first

attempts of implementation allowed for important reduction of multipath effect, as it was described in [52, 54]. However, use of antenna arrays also introduces hardware complexity, non-nominal effects connected with calibration, and coupling and also forces the estimation of additional unknown parameters.

Single GNSS antennas can also be used for determination of direction of arrival of signal, and thus detecting erroneous signals arrival, as it was described in [22].

4.3.5.4 Other Types of Mitigation

Another types of solution could be applied connected with ray-tracing, utilizing 3D city models, shadow matching, particle filtering, or visual methods in order to detect possible multipath signals [23, 63]. Their main concept is common to identify and eliminate the potentially erroneous signals based on the information on the received signal. However, this type of mitigation requires high complexity of algorithms and therefore is difficult to implement.

4.3.6 Multipath Detection Without Mitigation

For lower complexity alternatives and with the advent of four GNSS with many satellites in view, multipath detection schemes might give enough information to the user to benefit from being aware of the appearance of multipath propagation arriving at the antenna. Different ways of signal processing can be applied according to the result of the multipath detection process. In several cases only the information about the presence of the multipath in processed signal is relevant to the user. This information, in order to fulfill the requirements of liability or safety-of-life-critical applications, has to be provided instantly after the successful detection, therefore idea of the application of the theory of the Quickest Detection appeared, and was presented in [15, 18].

4.4 Radio Frequency Interference in GNSS

Since the GNSS receivers are relying on satellites broadcasting an RF signal, they are vulnerable to the RF interfering signals coming from radio transmitters. The RF interference problem is escalated by the fact that the GNSS satellite signals are received at very low power in the surface of the Earth (the received power ranges from −125 to −130 dBm depending on the GNSS constellation and signal [17, 45, 46]). The interfering signals that may appear in the GNSS bands can severely degrade the receiver performance. Therefore, in order to guarantee the integrity of the PVT solution, the receiver needs to detect an interference occurrence and mitigate its effects.

Interference can also arise from the signals transmitted by the other in-view satellites of the same constellation (intra-system interference) and from the signals transmitted by the other in-view satellites of other constellations (inter-system interference). However, these interference types are not addressed in this section.

4.4.1 Unintentional RF Interference

The unintentional RF interference to GNSS receivers is typically a result of a malfunctioning radio device. Because the spurious emissions from transmitters are typically located in the harmonics of their respective carrier frequencies, the most likely sources of unintentional RF interference are radios which are operating at the subharmonics of GNSS carrier frequencies. The first 15 subharmonics are shown in Tables 4.2 and 4.3 for GNSS L1 and L5 bands, respectively [36]. For GNSS L1 and L5 band, a reference receiver with a 20.46 and 24 MHz bandwidth was used to calculate the in-band interference frequency ranges.

4.4.1.1 Commercial Broadcasting Interference

Commercial broadcasting services that could cause RF interference to the GNSS L1 or L5 band are:

- AM Radio Broadcasting: AM transmitters are broadcasting in the ITU region 2 in frequencies between 535 and 1605 kHz. Even though the AM radio transmitters

Table 4.2 GNSS L1 frequencies subharmonics

Subharmonic	Frequency range (MHz)	Frequency allocation
2nd	782.595–792.825	UHF TV
3rd	521.730–528.550	UHF TV
4th	391.298–396.413	Defence systems
5th	313.038–317.130	Defence systems
6th	260.865–264.275	Defence systems
7th	223.599–226.521	Broadcasting
8th	195.649–198.206	VHF TV
9th	173.910–176.183	VHF TV
10th	156.519–158.565	VHF marine
11th	142.290–144.150	VHF military
12th	130.433–132.136	VHF
13th	120.399–121.973	VHF
14th	111.799–113.261	VOR/ILS
15th	104.346–105.710	FM

Table 4.3 GNSS L5 frequencies subharmonics

Subharmonic	Frequency range (MHz)	Frequency allocation
2nd	582.225–594.225	UHF TV
3rd	388.150–396.150	Defence systems
4th	291.113–297.113	Defence systems
5th	232.890–237.690	T-DAB
6th	194.075–198.075	VHF TV
7th	166.350–169.779	PMR
8th	145.556–148.556	VHF
9th	129.383–132.050	VOR/ILS/VHF
10th	116.445–118.845	VOR/ILS/VHF
11th	105.859–108.041	FM
12th	97.038–99.038	FM
13th	89.573–91.419	FM
14th	83.175–84.889	VHF TV
15th	77.630–79.230	PMR

can have a large transmission power (50 kW) it is very unlikely that they will cause any interference to the GNSS bands due to the minimum order of required harmonics to be aligned in-band with the GNSS bands. For L1 band, the minimum order of the harmonic is 985. For the GNSS L5 band, the minimum order is 735.

- FM Radio Broadcasting: FM radio transmitters are broadcasting in the ITU region 2 between 88 and 108 MHz. Even though there are several harmonics in the FM radio broadcasting service that fall into GNSS L1 or L5 bands that could possibly cause severe RF interference to a GNSS receiver operation, there has not been a single reported case of FM radio broadcasting service causing RF interference to GNSS receivers [57].
- Analog Television Broadcasting: The analog television broadcasting is operating in three different frequency bands. VHF transmission operates in two different frequency bands. The first band located between 54 and 88 MHz, and the second band located between 174 and 216 MHz. UHF band is located between 512 and 806 MHz. Each channel has a 6 MHz bandwidth. The maximum transmission power is limited to 5 MW for the UHF frequency band. Regulations state that the spurious emissions should be limited to −60 dBc. The total allowed spurious emissions for analog television broadcasting is 7 dBW. However, during field measurements it has been found that typically the spurious emissions from the analog television are suppressed more than 100 dB [57].
- Digital Television Broadcasting: The digital television broadcasting is operating in the same frequencies as its analog counterpart. The digital television broadcasting has more stringent power and spurious emissions requirements than

the analog television broadcasting. The maximum power level of the digital television broadcasting is limited to 1 MW, and the spurious emissions must be limited at least 110 dB with a measurement bandwidth of 500 kHz. This limits the maximum spurious emissions for the GNSS L1 band to −33.19 and −33.88 dBW for the GNSS L5 band, respectively [57].

- DME: The DME interrogators located in aircrafts operate between 1025 and 1150 MHz frequencies. The peak transmitting power is limited to 2 kW with 1 MHz channel bandwidth using pulse modulated signal. Even though interrogators are transmitting at a high power (up to 2 kW), the duty cycle is typically around 0.03%. Due to the low duty cycle, the interrogators should not pose a serious threat to GNSS receivers. The DME responders located in the ground operate between 960 and 1215 MHz frequencies. This frequency band is overlapping with the GNSS L5 band, which can result in severe RF interference conditions. The DME responders are limited to peak power of 15 kW with 1 MHz bandwidth. However, typical DME ground beacon installation is operating with 1 kW peak power. Additionally, if a TACAN beacon is co-located with the DME beacon, it is typically transmitting with a peak maximum power of 3.5 kW on the same frequencies. The DME responders are known to cause serious RF interference to the GNSS receivers. GNSS receivers installed in aircrafts can expect up to 6 dB SNR loss in the US DME hotspots when using pulse blanking mitigation technique [57].

4.4.2 Intentional RF Interference

The intentional RF interference is caused by devices that are designed to interfere with normal GNSS receiver operation and is typically divided into two categories: jamming and spoofing.

4.4.2.1 Jamming

GNSS jammers are devices which are designed to prevent GNSS receivers to provide user PVT solution. They accomplish this task by transmitting a signal with a higher power and close enough frequency of the GNSS signals to mask out any GNSS signal entering to the receiver. A comprehensive evaluation of different commercial GNSS jammers has been done [42]. In the study, it was found that all evaluated jammers were using swept tone modulation to produce jamming signal. Majority of the jammers were using chirp signal as the modulating tone, and the power of the jammers was varying greatly from 0.07 to 642 mW.

4.4.2.2 Spoofing

GNSS spoofers are radio transmitters which are trying to trick the GNSS receivers to produce corrupted PVT information. Several methods have been published on how to accomplish successful spoofing for a GNSS receiver. In this section, the spoofing is considered to be an RF signal attack against the GNSS receiver. There are other types of attacks, such as malware, that can be adopted to spoof GNSS receivers [58], but they are out of the scope of this document. The discussion is separated into three types of different spoofing attacks [29].

The simplest spoofing technique uses a GNSS signal simulator to produce false GNSS signals to the victim receiver. If the victim GNSS receiver acquires and tracks these false signals, it will eventually start to produce corrupted PVT solutions. Typically, this type of attack does not synchronize to real world GNSS signals, which makes it relatively easy to detect and counterattack. A GNSS signal simulator used for testing GNSS receivers can be seen as a typical example of this type of GNSS spoofer.

More advanced spoofer design is based on having a GNSS receiver synchronized with the spoofing signal transmitter. This enables the synchronization between the spoofer and real world GNSS signals. If the spoofer has a good knowledge on the 3D-pointing vector between its transmitting antenna and the victim receiver's antenna, it is able to synchronize its transmitted signals so that they are time and frequency aligned with the real world GNSS signal as seen by the victim receiver. After achieving this synchronization, the spoofer can lift the victim receivers' tracking loops into tracking the spoofing signals. This type of attack is much harder to counterattack, when the spoofer gradually changes the victim receiver off from its true PVT solution. However, this type of spoofer is much harder to build as it requires good knowledge of the 3D-pointing vector between the spoofers' transmitting antenna and victim GNSS receiver antenna.

The most advanced spoofer design is based on having multiple transmission antennas and synchronization between the real world GNSS signals and the spoofer. This type of attack requires extremely good knowledge of the victim receivers' antenna phase center. By having extremely good knowledge of the 3D-pointing vector, even the phases of the carrier signal can be matched between spoofing and real world GNSS signals. Having multiple transmitting antennas can effectively null the influence of having advanced spoofing mitigation techniques based on spatial signal processing. However, these types of spoofers are practically really hard to implement due to the precise position information that they require on the victim receiver.

4.4.3 Interference Countermeasures

Different countermeasures can be built in GNSS receivers in order to reduce the effect of RF interference to the GNSS receiver operation. These countermeasures can be divided into interference detection algorithms and interference mitigation

algorithms. The former group consists of algorithms that are designed to detect whether an interfering RF signal is entering the GNSS receiver antenna, and the latter group consists of algorithms that mitigate the effects caused by those interfering signals.

4.4.3.1 Interference Detection

Interference detection is very important in receivers in order to ensure the integrity of the PVT solution. Receiver should have the ability to detect the interfering signals entering into it in order to provide proper countermeasures. At minimum, GNSS receivers should be able to raise alarm flags indicating that the receiver is not working under the normal operating conditions. This information can be used to enhance the integrity of the output provided by the receivers.

The following section provides an overview of currently used interference detection algorithms.

- AGC: Today's GNSS receivers are using multi-bit ADC in the RF front end, as they offer better performance than single-bit ADCs. When multi-bit ADC is used in a GNSS receiver, an AGC needs to be implemented to keep the signal level optimum to the input of the ADC. In the normal operation of the AGC, the thermal noise floor overcomes the GNSS signals causing the AGC to follow the ambient noise environment [1]. Whenever an interfering RF signal is present at the victim receiver's antenna, a change can be observed at the AGC output due to excessive RF energy received by the receiver antenna [1]. This information of the gain level can be used to identify the presence of interfering RF signals. The difficulty of using AGC as an interference detection mechanism is that AGC operation is based on the receiver operational environment. The operational environment is changing constantly even without any interfering signals [1]. To overcome this problem, calibration and thermal isolation can be used to further increase the detector performance [1]. There are also some types of interfering signals such as pulsed interference where the performance of an AGC-based interference detector is degraded [47].
- Sample distribution: The distribution of the ADC samples has been proposed as an RF interference detection algorithm. RF interference detection based on the samples distribution is closely related to the AGC as the digital AGC is typically using sample distribution measurement to tune the AGC gain. To detect the RF interference from the sample distribution, a chi-square goodness-of-a-fit test can be used [3, 43]. The test compares distributions between observed and reference cases and produces a metric of the distribution similarity. The sample distribution algorithms work well when the RF interference has a distribution which is not similar to a Gaussian distribution. A good example of such distribution is Continuous Wave (CW) interference where the sample distribution is concentrated on the peaks of the continuous wave signal.

- Frequency domain: Frequency domain algorithms transform the pre-correlation samples into frequency domain by using the Fourier transform and try to identify the possible interfering source from the frequency content of the pre-correlation samples. Either an assumption is made that the frequency spectrum is flat in the receiver pass band as the incoming samples are dominated by the ambient white noise, or a calibration procedure is done at the start of detection procedure to determine the interference free spectral content. Frequency domain methods are extensively used in conjunction with adaptive filtering techniques to tune the filter to filter out interference signals [10, 14, 50]. Interference detection in frequency domain has a good performance detecting interfering signals that are occupying narrow bandwidth such as CW interference. For wideband interference, frequency domain analysis results in poor performance [41].

4.4.3.2 Interference Mitigation

The objective of interference mitigation is to decrease the effects of RF interfering signal on GNSS receiver performance. There are several approaches on how to perform the interference mitigation on GNSS receiver. Different RF interference sources require different algorithms to provide good performance.

The next section discusses the common approaches on RF interference mitigation.

- Blanking: Signal blanking methods are used for the pulsed interference in which the interfering signal is present only at a short period of time. Blanking methods try to detect when the interference signal is active and null the input samples during these periods. Theoretically, a perfect pulse blanking algorithm results in $10\log_{10}(1-\mathrm{PDC}_B)$ SNR loss [25], where PDC_B (pulse duty cycle of the blanker) is the net aggregate duty cycle of all pulses exceeding the blanker threshold. Any type of signal blanking algorithm is always a suboptimal interference mitigation technique, as it also removes the useful part of the signal [25].
- Adaptive filtering: Adaptive filter algorithms work by estimating the RF interference on the incoming signal and adapting the filter parameters appropriately to filter out all frequencies contaminated with the RF interference from the incoming signal. Many different approaches can be taken for the adaptive filtering methods. The LMS technique can be used to iteratively minimize the signal power at the filter output [8, 9]. The rationale of this approach is that the power of the interfering signal is typically much larger than the ambient noise, and minimizing the power at the signal output will effectively remove the RF interference signal. Adaptive filtering is most effective against narrowband RF interference. With wideband RF interference, adaptive filters are removing also a big part of the useful GNSS signal reducing the interference mitigation performance.
- Beamforming: Interference mitigation algorithms based on beamforming are dynamically controlling the antenna radiation pattern to direct the maximum

gain towards satellite locations, or minimum gain towards the direction of the RF interference. Algorithms based on beamforming have the best performance compared to other RF interference mitigation algorithms. In most cases, the source of the RF interference is located on the ground, which enables good protection from the RF interference as the DOA of the RF interference is well separated from the DOA of the GNSS satellite signals. Moreover, as the filtering is based on DOA, these algorithms work well against different types of RF interference [12]. The problem with all beamforming algorithms is that they are both hardware and computationally expensive. They require multiple antenna elements, RF front-end configurations, and adaptive weighting functions to be calculated for every antenna element.

4.5 Conclusions

This chapter presented three major error sources in GNSS positioning and different techniques to mitigate them.

Ionospheric mitigation techniques present advantages and limitations. Physical models reduce error in a certain amount with a small set of parameters. Ionospheric maps improve the accuracy of physical models, but require the user to access the detailed grids, and yet, do not provide high accuracy solutions for regions where density of receivers used to build the models is low. Data driven models provide a promising alternative for low latitudes, but require the generation of local models and their continuous update. Depending on the complexity level, available resources, and latency required, different solutions can be used.

The multipath phenomenon influences all the GNSS observables in a different extent and way. In some part, it is satellite system dependent, so utilizing more signals, especially with modernized modulation, can help in mitigation. The long delay reflections with modern mitigation techniques are possible to be detected; however, the short delay multipath is still challenging. Each mitigation technique can behave better in specific situations and can have drawbacks of the level of complexity and thus influence the level of effectiveness of implementation. Efficient integration of several techniques, especially in forthcoming multiple system scenario, seems to be nowadays the best solution for mitigating multipath influence. Still, in the situations where lower complexity is demanded, only detection of multipath could be applied with success.

Since the satellites signals are very weak when they reach the Earth surface, RF interference is another important GNSS vulnerability. Unwanted signals can disturb the receiver performance, causing a degradation in the PVT solution accuracy or even preventing the receiver from obtaining the solution. It is, therefore, vital for the receiver to have the capacity to detect an interference event and to mitigate its consequences. Depending on the interference type, some techniques are more suitable than others. Frequency domain-based approaches, for example, have good

performance in detecting narrowband interfering signals, but poor performance with wideband interferers. As for the mitigation, the blanking methods work very well for DME pulses, whereas adaptive filtering performs better with narrowband interference. Beamforming techniques have the best performance among the mitigation algorithms but they are hardware and computationally expensive.

Acknowledgements This work was financially supported by EU FP7 Marie Curie Initial Training Network MULTI-POS (Multi-technology Positioning Professionals) under grant nr. 316528.

References

1. D.M. Akos, Who's afraid of the spoofer? GPS/GNSS spoofing detection via automatic gain control (AGC). Navigation **59**(4), 281–290 (2012)
2. B. Arbesser-Rastburg, The Galileo single frequency ionospheric correction algorithm, in *Third European Space Weather Week*, vol. 13 (2006), p. 17
3. F. Bastide et al., Automatic gain control (AGC) as an interference assessment tool, in *ION GPS/GNSS 2003, 16th International Technical Meeting of the Satellite Division of The Institute of Navigation* (2003), pp. 2042–2053
4. M.Z.H. Bhuiyan, E.S. Lohan, Advanced multipath mitigation techniques for satellite-based positioning applications. Int. J. Navig. Observ. **2010**, 1–15 (2010)
5. M.H.Z. Bhuiyan, E.S. Lohan, Multipath mitigation techniques for satellite-based positioning applications, in *Global Navigation Satellite Systems-Signal, Theory and Applications* (InTech, Rijeka, 2012)
6. A. Bilich, K.M. Larson, P. Axelrad, Modeling GPS phase multipath with SNR: case study from the Salar de Uyuni, Boliva. J. Geophys. Res. Solid Earth **113**(B4), B04401 (2008)
7. D. Bilitza, IRI international reference ionosphere, http://iri.gsfc.nasa.gov/. Accessed 24 May 2016
8. D. Borio, L. Camoriano, P. Mulassano, Analysis of the one-pole notch filter for interference mitigation: Wiener solution and loss estimations, in *Proceedings of the 19th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS 2006)* (2006), pp. 1849–1860
9. D. Borio, L. Camoriano, L. Lo Presti, Two-pole and multipole notch filters: a computationally effective solution for GNSS interference detection and mitigation. Syst. J. IEEE **2**(1), 38–47 (2008)
10. D. Borio et al., Time-frequency excision for GNSS applications. Syst. J. IEEE **2**(1), 27–37 (2008)
11. C. Cesaroni et al., L-band scintillations and calibrated total electron content gradients over Brazil during the last solar maximum. J. Space Weather Space Clim. **5**, A36 (2015)
12. S. Daneshmand et al., Interference and multipath mitigation utilising a two-stage beamformer for global navigation satellite systems applications. IET Radar Sonar Navig. **7**(1), 55–66 (2013)
13. G. Di Giovanni, S.M. Radicella, An analytical model of the electron density profile in the ionosphere. Adv. Space Res. **10**(11), 27–30 (1990). doi:10.1016/0273-1177(90)90301-F
14. F. Dovis, L. Musumeci, Use of wavelet transforms for interference mitigation, in *2011 International Conference on Localization and GNSS (ICL-GNSS)* (IEEE, New York, 2011), pp. 116–121
15. D. Egea-Roca et al., Signal-level integrity and metrics based on the application of quickest detection theory to multipath detection, in *Proceedings of the 28th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS+ 2015)* (2015), pp. 2926–2938

16. EUREF Permanent Network EPN, www.epncb.oma.be. Accessed 24 May 2016
17. European Union 2015, Galileo open service signal in space interface control document. Technical Report 1.2. Accessed 24 May 2016
18. N.G. Ferrara et al., Combined architecture. Enhancing multi-dimensional signal quality in GNSS receivers, in *Inside GNSS Working Papers* (2016), pp. 54–62
19. L. Garin, F. van Diggelen, J.-M. Rousseau, Strobe and edge correlator multipath mitigation for code, in *Proceedings of 9th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION-GPS96)* (1996), pp. 657–664
20. Geodesia: Red de Estudios de Deformacion, geored.sgc.gov.co. Accessed 24 May 2016
21. S. Gleason, D. Gebre-Egziabher, *GNSS Applications and Methods* (Artech House, London, 2009)
22. D.E. Grimm, L. Steiner, R. Mautz, GNSS antenna orientation and detection of multipath signals from direction of arrival, in *Proceedings of 25th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS+ 2012)* (2012)
23. P.D. Groves et al., Intelligent urban positioning using multi-constellation GNSS with 3D mapping and NLOS signal detection, in *Proceedings of 25th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS+ 2012)* (2012)
24. R.R. Hatch, The synergism of code and carrier measurements, in *Proceedings of the Third International Symposium on Satellite Doppler Positioning* (1982)
25. C. Hegarty et al., Suppression of pulsed interference through blanking, in *Proceedings of the IAIN World Congress and the 56th Annual Meeting of The Institute of Navigation* (2000), pp. 399–408
26. M. Hernandez-Pajares et al., The IGS VTEC maps: a reliable source of ionospheric information since 1998. J. Geod. **83**(3–4), 263–275 (2009). doi:10.1007/s00190-008-0266-1
27. B. Hofmann-Wellenhof, H. Lichtenegger, E. Wasle, *GNSS - Global Navigation Satellite Systems: GPS, GLONASS, Galileo, and More* (Springer, Wien, 2008)
28. M. Irsigler, Multipath propagation, mitigation and monitoring in the light of Galileo and the modernized GPS. Ph.D. thesis, University FAF Munich, 2008
29. A. Jafarnia-Jahromi et al., GPS vulnerability to spoofing threats and a review of antispoofing techniques. J. Navig. Observ. **2012** (2012). doi:10.1155/2012/127072
30. C. Jeffrey, *An Introduction to GNSS GPS, GLONASS, Galileo and Other Global Navigation Satellite Systems* (NovAtel Inc., Calgary, 2010)
31. S.-P. Kao, Y.-C. Chen, F.-S. Ning, A MARS-based method for estimating regional 2-D ionospheric VTEC and receiver differential code bias. Adv. Space Res. **53**(2), 190–200 (2014). doi:10.1016/j.asr.2013.11.001
32. D.E. Kaplan, J.C. Hegarty, *Understanding GPS: Principles and Applications*, 2nd edn. (Springer/Artech House, Wien/New York, 2006)
33. P.M. Kintner Jr., T. Humphreys, J. Hinks, GNSS and ionospheric scintillation-how to survive the next solar maximum, in *Inside GNSS*, July–August 2009, pp. 22–30
34. J.A. Klobuchar, Ionospheric time-delay algorithm for single-frequency GPS users. IEEE Trans. Aerosp. Electron. Syst. **3**, 325–331 (1987)
35. J.A. Klobuchar, Ionospheric effects on GPS. GPS World **4**, 48–51 (1991)
36. R.J. Landry, A. Renard, Analysis of potential interference sources and assessment of present solutions for GPS/GNSS receivers, in *4th St. Petersburg International Conference on Integrated Navigation Systems* (1997), pp. 1–13
37. S.K. Leong et al., Assessment of ionosphere models at Banting: performance of IRI-2007, IRI-2012 and NeQuick 2 models during the ascending phase of Solar Cycle 24. Adv. Space Res. **55**(8), 1928–1940 (2015)
38. W.C. Machado, E.S. da Fonseca Jr., Artificial neural networks applied to VTEC prediction in Brazil. Redes Neurais Artificiais Aplicadas Na Previso Do VTEC No Brasil. Bol. Cienc. Geod. **19**(2), 227–246 (2013)
39. S. Magdaleno, M. Herraiz, B.A. de La Morena, Characterization of equatorial plasma depletions detected from derived GPS data in South America. J. Atmos. Sol. Terr. Phys. **74**, 136–144 (2012)

40. C.P Mantz, Q. Zhou, Y.T. Morton, Application of a neural network model to GPS ionosphere error correction, in *Position Location and Navigation Symposium, 2004. PLANS 2004* (IEEE, New York, 2004), pp. 538–542. doi:10.1109/PLANS.2004.1309039
41. L. Marti, F. van Graas, Interference detection by means of the software defined radio, in *Proceedings of the 17th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS 2004)* (2001), pp. 99–109
42. R.H. Mitch et al., Signal characteristics of civil GPS jammers, in *ION GNSS 2011 the 24th International Technical Meeting of The Satellite Division of the Institute of Navigation, Portland OR* (2011), pp. 1907–1919
43. B. Motella, M. Pini, L. Lo Presti, GNSS interference detector based on chi-square goodness-of-fit test, in *2012 6th ESA Workshop on Satellite Navigation Technologies and European Workshop on GNSS Signals and Signal Processing (NAVITEC)* (2012), pp. 1–6. doi:10.1109/NAVITEC.2012.6423070
44. NASA: Real Time Ionospheric Maps, http://ionojplnasagov/latest_rti_global.html. Accessed 23 May 2016
45. Navstar Global Positioning System: Interface Specification IS-GPS-705. Technical Report, GPS Joint Program Office, 2003
46. Navstar Global Positioning System: Interface Specification IS-GPS-200, Revision D. Technical Report, GPS Joint Program Office, 2006
47. A. Ndili, P. Enge, GPS receiver autonomous interference detection, in *Position Location and Navigation Symposium, IEEE 1998* (IEEE, New York, 1998), pp. 123–130
48. R.D.J. Nee, Multipath effects on GPS code phase measurements. Navigation **39**(2), 177–190 (1992)
49. O.J. Olwendo et al., Comparison of GPS TEC variations with IRI-2007 TEC prediction at equatorial latitudes during a low solar activity (2009–2011) phase over the Kenyan region. Adv. Space Res. **52**(10), 1770–1779 (2013)
50. M. Paonni et al., Innovative interference mitigation approaches: analytical analysis, implementation and validation, in *2010 5th ESA Workshop on Satellite Navigation Technologies and European Workshop on GNSS Signals and Signal Processing (NAVITEC)* (2010), pp. 1, 8, 8–10 doi:10.109/NAVITEC.2010.5708055
51. R. Prasad, M. Ruggieri, *Applied Satellite Navigation-Using GPS, GALILEO and Augmentation Systems* (ARTECH House Publishers, Boston, 2005)
52. J.K. Ray, Mitigation of GPS code and carrier phase multipath effects using a multi-antenna system. Ph.D. thesis, National Library of Canada, Bibliothque nationale du Canada, 2001
53. J.K. Ray, M.E. Cannon, Characterization of GPS carrier phase multipath, in *Proceedings of ION National Technical Meeting* (1999), pp. 243–252
54. J.K. Ray, M.E. Cannon, P.C. Fenton, Mitigation of static carrier-phase multipath effects using multiple closely spaced antennas. Navigation **46**(3), 193–201 (1999)
55. Red GNSS de Monitoreo Continuo del Ecuador REGME, http://www.geoportaligm.gob.ec. Accessed 24 May 2016
56. C. Rost, L. Wanninger, Carrier phase multipath mitigation based on GNSS signal quality measurements. J. Appl. Geod. **3**(2), 81–87 (2009)
57. RTCA (Firm). SC-159, Assessment of Radio Frequency Interference Relevant to the GNSS L5/E5A Frequency Band (2004). RTCA/DO. RTCA
58. L. Scott, Spoofing: upping the anti. Inside GNSS. Thought Leadership Series (2013), pp. 18–19
59. G. Seco-Granados et al., Challenges in indoor global navigation satellite systems: unveiling its core features in signal processing. IEEE Signal Process. Mag. **29**(2), 108–131 (2012)
60. M. Smyrnaios, S. Schon, M. Liso Nicolas, Multipath propagation, characterization and modeling in GNSS, in *Geodetic Sciences - Observations, Modeling and Applications*, ed. by S. Jin, Chap. 2 (InTech, Rijeka, 2013)
61. M. Spangenberg et al., Urban navigation system for automotive applications using HSGPS, inertial and wheel speed sensors, in *ENC-GNSS 2008, Conference Europeenne de la Navigation* (2008)

62. Er. Steingass, A. Lehner German, Measuring the navigation multipath channel-a statistical analysis, in *Proceedings of the ION GNSS 2004* (2004); Citeseer
63. T. Suzuki, N. Kubo, GNSS positioning with multipath simulation using 3D surface model in urban canyon, in *25th International Technical Meeting of the Satellite Division of the Institute of Navigation 2012, ION GNSS 2012* (2012)
64. H. Takahashi et al., Diagnostics of equatorial and low latitude ionosphere by TEC mapping over Brazil. Adv. Space Res. **54**(3), 385–394 (2014)
65. D.V. Ratnam et al., TEC prediction model using neural networks over a low latitude GPS station. Int. J. Soft Comput. Eng. **2**(2), 2231–2307 (2012)
66. L. Wanninger, M. May, Carrier-phase multipath calibration of GPS reference stations. Navigation **48**(2), 112–124 (2001)
67. R. Watson et al., Investigating GPS signals indoors with extreme high-sensitivity detection techniques. Navigation **52**(4), 199–213 (2005)

Chapter 5
GNSS Quality of Service in Urban Environment

Enik Shytermeja, Maciej Jerzy Paśnikowski, Olivier Julien, and Manuel Toledo López

5.1 Conventional GNSS Signal Tracking

Recently, Global Navigation Satellites Systems (GNSS) constitute an important position in the development of urban navigation applications and associated services. However, the major concern of the constant growth of GNSS-based urban applications is related to the quality of the positioning service, expressed in terms of accuracy, availability, and continuity not only of service but also of integrity provision, ensuring that the application requirements are met [28]. Prior to entering into the problematics of urban navigation, it is important to introduce the high-level representation of the GNSS signal processing and receiver architecture that is illustrated in Fig. 5.1.

The overall goal of any GNSS navigation receiver is the computation of the user's PVT solution based on the sequential processing of the received signals from the satellites in view. Contrary to CDMA in communications, GNSS receivers rely on accurate synchronization between their local time with the GNSS satellites time scale in order to make distance (pseudorange estimation) and velocity (Doppler estimation) measurements [27]. The receiver's macro blocks are:

E. Shytermeja (✉) • O. Julien
Telecom Lab, SIGNAV Research Group, École Nationale de l'Aviation Civile, 7 Avenue Edouard Belin, 31055 Toulouse cedex 4, France
e-mail: shytermeja@recherche.enac.fr

M.J. Paśnikowski • M.T. López
GMV Aerospace and Defence S.A.U., Calle Isaac Newton 11 P.T.M., 28760 Tres Cantos, Madrid, Spain

J. Nurmi et al. (eds.), *Multi-Technology Positioning*,
DOI 10.1007/978-3-319-50427-8_5

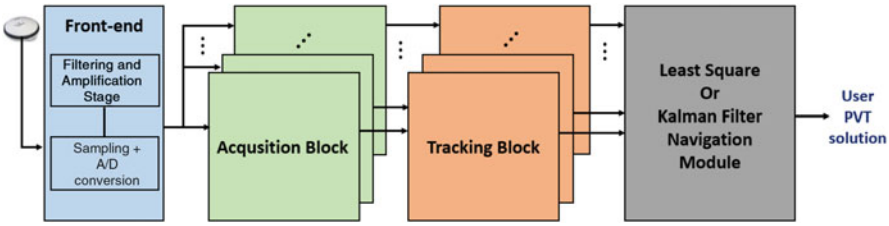

Fig. 5.1 High-level representation of a GNSS receiver architecture block diagram

- The analog **RF front-end** is the first stage of the signal processing chain. It includes the receiver antenna, the *Low Noise Amplifier (LNA)*, the *Intermediate Frequency (IF) down-converter*, the *IF band-pass filter*, and the *ADC*. The output of this block is an IF discrete version of the received Signal-in-Space (SIS) [14];
- An **acquisition strategy** per each satellite channel is employed to first detect the signal presence and moreover, to provide a rough estimation of the *delay of the received code* (τ) with respect to the receiver's generated PRN code and the *Doppler frequency shift* (f_D) of the received signal (comprising the Doppler term related to the satellite velocity and user's receiver dynamics) in the two-dimensional search space. The pair (τ, f_D) will be further fed to the following tracking stage to refine this estimation [4];
- **The tracking stage**, employed for each satellite in view, includes two sub-modules: Firstly, the DLL used for continuously tracking the code phase offset $(\delta\tau)$ with respect to the incoming signal's code and the carrier tracking loops used for estimating the residual Doppler shift (δf_D) and carrier phase offset $(\delta\phi)$. The carrier tracking module that estimate only (δf_D) are called FLL, while the carrier tracking loops which estimate both (δf_D) and $(\delta\phi)$ are called PLL [3, 4]. The generic code and carrier tracking loop (DLL/FLL/PLL) architecture of a single tracking channel is depicted in Fig. 5.2.
- The **navigation module** represents the final navigation processing block and includes the following steps:
  - *Bit synchronization*, following the code and carrier estimation procedure, is the phase where all the bit transition times are located based on zero crossing detection;
  - *Navigation data demodulation* performs the navigation data bits recovery from where the satellite ephemerides, GPS TOW and satellite clock correction parameters, the system state and time stamps on the received signal are extracted from the navigation message [4, 27];
  - *Satellite position computation*, based on the current date and time, plus an indication of the health of the satellite sending the signal, the orbital information about the satellite sending the signal, called the ephemeris, and the satellites information status included in the GNSS program, called the almanac;

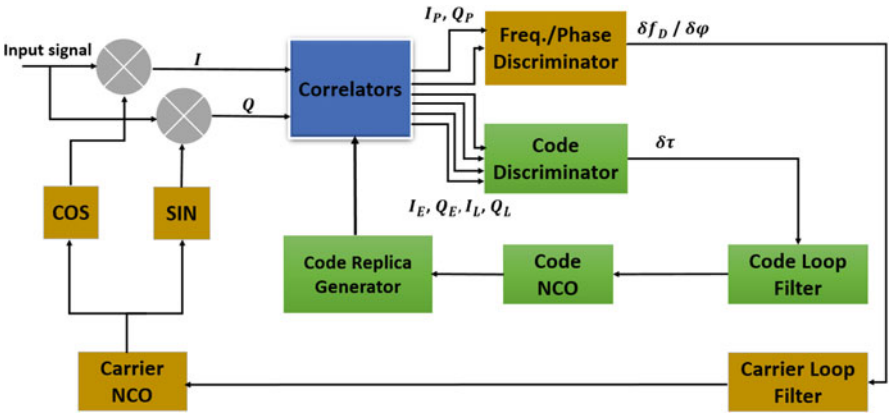

Fig. 5.2 Generic code and carrier tracking loop block diagram

– *The navigation solution computation* is the last step of the receiver signal processing unit. The final output of this block is the user position, velocity, and time estimation using the Weighted Least Square (WLS) algorithm or the Kalman Filter (KF) estimation technique, with the latter providing an increased position solution availability in signal-constrained environment, related to its state forward-prediction model. In both cases, the measurement inputs to this positioning algorithms are the satellite-to-user pseudoranges and Doppler measurements, obtained from the tracking level and taking use of the satellite position information from the navigation message decoding procedure.

The receiver actually obtains a pseudorange measurement for each tracked satellite that can be modeled as:

$$\rho_i = d_i + c\left(\delta t_u - \delta t_{s,i} + \delta t_{\text{iono},i} + \delta t_{\text{tropo},i} + \delta t_{\epsilon,i} + \delta t_{\text{rel},i} + \delta t_{n,i}\right) , \tag{5.1}$$

where d_i is the true geometric distance between the ith satellite and the receiver, δt_u denotes the offset between the receiver time and the GNSS time, $\delta t_{s,i}$ denotes the satellite clock bias, $\delta t_{\text{iono},i}$ and $\delta t_{\text{tropo},i}$ denote the ionospheric- and tropospheric-induced errors, respectively, $\delta t_{\epsilon,i}$ represents the satellite ephemerides errors that vary slowly in time (2 m root mean square (rms) in 1 h), $\delta t_{\text{rel},i}$ is the relativistic error, and $\delta t_{n,i}$ comprises both the noise and code multipath error terms.

It should be noted that the different error contributions are generally modeled as Gaussian distributions with zero mean and variances and they contribute to the total error that is the User Equivalent Range Error (UERE). In open sky conditions and for single frequency GNSS receivers, the major error source comes from the Ionosphere. For this matter, the most widely used ionosphere correction method is the Klobuchar model [16] that is capable of modeling about 50% of the ionospheric error. A more efficient technique performing the ionospheric error correction is

the IGS that publishes a global TEC map for users around the world. The IGS model requires the users to interpolate their own TEC and calculate the ionospheric delay from it. The IGS exact ionospheric model is proven to remove at least 80% of ionospheric error. The NeQuick is the ionospheric model proposed to be used by the Galileo, the European Global Navigation Satellite System (Galileo) single frequency receiver in order to compute ionospheric error correction [8]. However, when referring to the urban environment, the main error contribution is in fact constituted by the multipath errors that increase the variance of the total system UERE error. Moreover, multipath errors affect both the position accuracy and precision, introducing a positioning bias that is not accounted for in the pseudorange error budget.

Following the acquisition stage, the digitized IF signal is then fed into the signal tracking architecture that in the conventional receiver structure provides a simultaneous tracking of all the satellites in view. In this stage, the receiver generates a local signal replica that follows the parameters of the incoming signal. The objective of the tracking stage is the precise estimation of the code delay and carrier frequency/phase of the incoming signal based on the correlator outputs. The high-level block diagram representation of the scalar GNSS tracking architecture is illustrated in Fig. 5.2.

The tracking stage, employed for each satellite in view, includes two sub-modules: Firstly, the DLL used for continuously tracking the code phase offset $(\delta\tau)$ with respect to the incoming signal's code and the carrier tracking loops used for estimating the residual Doppler shift (δf_D) and carrier phase offset $(\delta\phi)$. The carrier tracking module that estimate only (δf_D) are called FLL, while the carrier tracking loops which estimate both (δf_D) and $(\delta\phi)$ are called PLL [3, 4]. The conventional scalar tracking architecture includes the following three main modules:

- **Code/Carrier Discriminator functions**: computing the code delay and carrier tracking errors by using combination of the correlator outputs;
- **Numerically Controlled Oscillator (NCO)**: is responsible of converting the discriminator output into a frequency that drives the local replica generation. It must be noted that different NCOs are used to generate the code and carrier replica;
- **Loop filters**: aiming at the noise reduction of the discriminator estimate at the NCO input. The choice of the filter parameters such as the bandwidth, the integration time, and the filter order is strictly related to the user dynamics.

In the GNSS receiver, the code delay and carrier frequency/phase lock loops are jointly used. However, for a better understanding of the tracking loops, the code and carrier tracking loops are separately analyzed in details in the following sections.

5.1.1 Code Tracking Loop

The code tracking loop is a feedback loop in charge of finely estimating the code phase delay of the incoming signal and to keep it aligned with the code delay

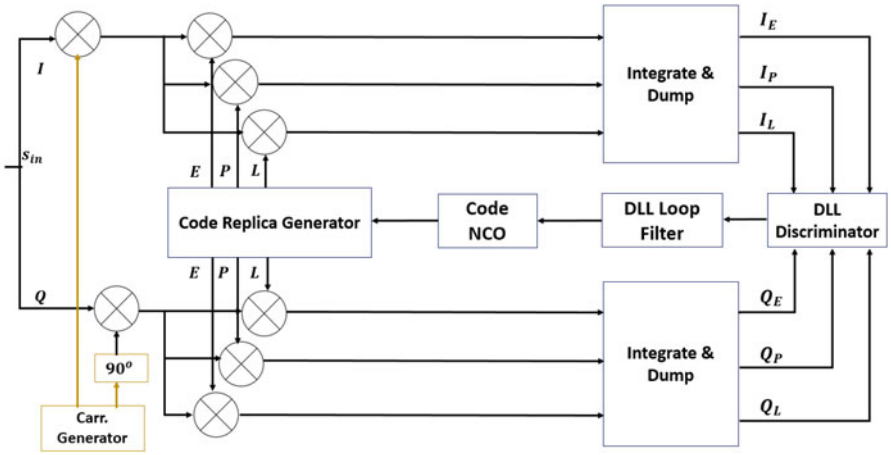

Fig. 5.3 GNSS Delay Lock Loop architecture. The *orange blocks* are related to the carrier tracking and are outside the DLL

of the local replica by means of DLL. It must be noted that code delay tracking provides more robust measurements than the carrier tracking ones due to the long chip duration and this is required especially for urban environment applications [13]. The general structure of the DLL starting from the digitized signal samples s_{in} is shown in Fig. 5.3.

The sine and cosine map functions produce the two signal branches, namely the I and Q signal component phases 90° apart from the carrier NCO and generator block that is given in orange in Fig. 5.3. The I and Q components are multiplied by the delayed code replicas (Early, Prompt, and Late) that are expressed as:

$$\begin{aligned} E &= R_c\left(t - \hat{\tau} - \frac{C_s}{2}\right) \\ P &= R_c\left(t - \hat{\tau}\right) \\ L &= R_c\left(t - \hat{\tau} + \frac{C_s}{2}\right) \end{aligned} \tag{5.2}$$

where: the Prompt code R_c replica is aligned with the incoming signal code, while the Early and Late code replica has a half code chip $\frac{C_s}{2}$ advance and delay, respectively, w.r.t. the incoming signal code.

Finally, the three correlator pairs, output of the correlator block for each parallel tracking channel, are expressed in Eq. (5.3):

$$
\begin{aligned}
I_P &= A \cdot N \cdot R_c(\delta t_u) \cdot \frac{\sin(\pi \cdot \delta f_D \cdot T)}{\pi \cdot \delta f_D \cdot T} \cdot \cos(\delta\phi_c) + n_{I_P} \\
I_E &= A \cdot N \cdot R_c\left(\delta t_u - \frac{C_s}{2}\right) \cdot \frac{\sin(\pi \cdot \delta f_D \cdot T)}{\pi \cdot \delta f_D \cdot T} \cdot \cos(\delta\phi_c) + n_{IE} \\
I_L &= A \cdot N \cdot R_c\left(\delta t_u + \frac{C_s}{2}\right) \cdot \frac{\sin(\pi \cdot \delta f_D \cdot T)}{\pi \cdot \delta f_D \cdot T} \cdot \cos(\delta\phi_c) + n_{IL} \\
Q_P &= A \cdot N \cdot R_c(\delta t_u) \cdot \frac{\sin(\pi \cdot \delta f_D \cdot T)}{\pi \cdot \delta f_D \cdot T} \cdot \sin(\delta\phi_c) + n_{QP} \\
Q_E &= A \cdot N \cdot R_c\left(\delta t_u - \frac{C_s}{2}\right) \cdot \frac{\sin(\pi \cdot \delta f_D \cdot T)}{\pi \cdot \delta f_D \cdot T} \cdot \sin(\delta\phi_c) + n_{QE} \\
Q_L &= A \cdot N \cdot R_c\left(\delta t_u + \frac{C_s}{2}\right) \cdot \frac{\sin(\pi \cdot \delta f_D \cdot T)}{\pi \cdot \delta f_D \cdot T} \cdot \sin(\delta\phi_c) + n_{QL}
\end{aligned}
\tag{5.3}
$$

where I and Q denote the in-phase and quadrature signal components, while E, P and L denote the early, prompt, and late correlators, A denotes the received signal amplitude computed from the signal's carrier-to-noise ratio C/N_0, N is the number of accumulated correlator samples that is typically set to 20 according to the integration time T of 20 ms, R_c is the autocorrelation function of the ranging code, $\delta\tau$ is the code phase delay error, C_s is the correlator offset expressed as $C_s = k_{Cs} \cdot T_c$ in [sec] where k_{Cs} is the correlator spacing in unit of code chips and T_c is the chip code period, δf_D is local carrier (or Doppler) frequency error, $n_{I/P}$ is the correlator noise being normally Gaussian distributed with zero mean and variance $\sigma_{IQ}^2 = \frac{C}{2 \cdot T \cdot 10^{\frac{(C/N_0)}{10}}}$, computed as a function of the carrier-to-noise density ratio C/N_0 in (dB-Hz) as in [23], and $\delta\phi_c$ is the average carrier phase error over the integration interval computed as [26]:

$$
\delta\phi_c = \delta\phi_0 + \frac{T}{2} \cdot \delta f_0 + \frac{T^2}{6} \cdot \delta a_0 \tag{5.4}
$$

where the zero subscript indicates the value at the start of the integration interval and δa_0 is the carrier frequency rate error (or phase acceleration error). The E and L correlator spacing is strictly related to the code autocorrelation function sharpness. Thus, typically for GPS L1 signals, the E and L code replicas are separated in phase by 0.5 chip while for the Galileo E1 Composite Binary Offset Carrier (CBOC) this spacing is reduced to 0.2 chip [13]. The code DLL is designed to track the relative delay between the incoming and the local signal code in the presence of AWGN. For this, a null-seeking strategy is used based on the code *discriminator function* that is zero (null) only in the case of synchronization between the incoming and replica code [14]. The most common code discriminators are the non-coherent EML and

the DP for their tracking robustness due to the insensitivity toward the carrier phase errors. They are defined as:

$$D_{\text{EMLP}} = \left(I_E^2 + Q_E^2\right) - \left(I_L^2 + Q_L^2\right) \tag{5.5}$$

$$D_{\text{DP}} = \left(I_E - I_L\right) \cdot IP + \left(Q_E - Q_L\right) \cdot QP \tag{5.6}$$

The code delay estimation τ_k at time epoch k is updated by using the code discriminator output $D_{\text{EMLP/DP}}$, the DLL loop filter gain K_{code}, and the code chipping rate f_{code} as follows [10]:

$$\tau_k = \tau_{k-1} + K_{\text{code}} \cdot \frac{D_{\text{EMLP/DP}}}{f_{\text{code}}} \tag{5.7}$$

The DLL performance depends on the following three main parameters [10]:

- **Accumulation time**: a high accumulation time is preferred for the DLL noise reduction, however, with an upper limit set to the data bit transition period (for instance, 20 ms for GPS L1 C/A signal tracking). While pilot signal tracking is performed, no limits on the accumulation time are imposed. Moreover, the integration time choice is proportional to the user dynamics;
- **DLL order**: has a great impact on the dynamic stress error. Therefore, a higher DLL order increases the code tracking loop robustness toward the dynamic error. In conventional GNSS receivers, a first order DLL is commonly used coupled with the employ of high order PLL to accurately face with the dynamic error;
- **Loop bandwidth**: where the narrower the DLL loop bandwidth is, the higher noise reduction is achieved.

5.1.2 Carrier Tracking Loop

The carrier tracking loop is a feedback loop responsible for finely estimating the Doppler frequency (f_D) of the incoming GNSS sinusoidal wave and its carrier phase (ϕ_c). For this purpose, two distinct carrier tracking loops can be employed in the GNSS receiver such as the FLL and the PLL. The FLL is capable of tracking the signal's Doppler frequency ignoring the carrier phase. Once the FLL provides a fine estimation of the carrier frequency, the PLL is initiated, working on a narrower band in order to recover the incoming signal phase [9]. Typically in GNSS receivers, a *Costas loop* implementation of the PLL is used for carrier tracking due to its lack of sensitivity to 180° phase reversal related to data bit transitions [4, 9]. The general structure of a carrier tracking loop is illustrated in Fig. 5.4.

Both the FLL and PLL tracking loops use only the prompt (aligned) code replica from the baseband signal processor with the only difference relying on the fact that the loop estimates three quantities for phase tracking and two for frequency tracking [10]. As illustrated in the block diagram of Fig. 5.4, two multiplications occur: Firstly, between the input signal s_{in} and the local replica to generate the

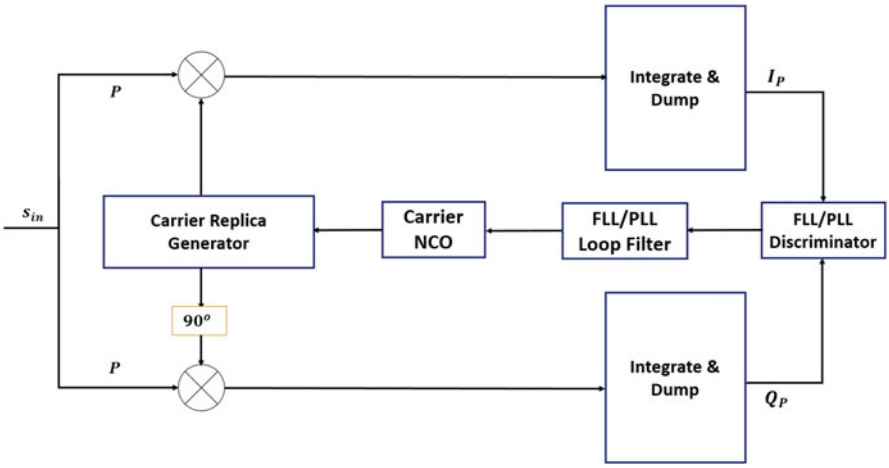

Fig. 5.4 General GNSS carrier tracking loop block diagram

In-phase branch I and secondly, between the incoming signal and the 90° phase-shifted carrier replica for the Quadrature (Q) branch.

Afterward, the two signals are low-passed filtered to eliminate the double intermediate frequency terms and fed to the carrier discriminator function. The most common used PLL discriminator used in the GNSS domain is the *atan* discriminator that is defined as:

$$\epsilon_{\phi_k} = \tan^{-1}\left(\frac{Q_P(k)}{I_P(k)}\right) \tag{5.8}$$

Finally, the carrier discriminator output is filtered to predict and estimate the Doppler frequency and the user dynamics [4]. Typically in a GNSS receiver, a third order PLL is used that provides estimation of the carrier phase error, Doppler frequency shift, and Doppler frequency rate. Whereas, the carrier frequency discriminators employ the correlator outputs coming from both the current and previous epoch. The commonly used FLL discriminators are the decision-directed cross-product (DDC), cross-over-dot product (COD), and the ATAN discriminators, computed as:

$$\begin{aligned} D_{\text{FLL-DDC}} &= \big(I_P(k-1)\cdot Q_P(k) - I_P(k)\cdot Q_P(k-1)\big) \\ &\quad \cdot \text{sign}\big(I_P(k-1)\cdot I_P(k) + Q_P(k-1)\cdot Q_P(k)\big) \\ D_{\text{FLL-COD}} &= \left(\frac{I_P(k-1)\cdot Q_P(k) - I_P(k)\cdot Q_P(k-1)}{I_P(k-1)\cdot I_P(k) + Q_P(k-1)\cdot Q_P(k)}\right) \\ D_{\text{FLL-ATAN}} &= \arctan\left(\frac{Q_P(k)}{I_P(k)}\right) - \arctan\left(\frac{Q_P(k-1)}{I_P(k-1)}\right) \end{aligned} \tag{5.9}$$

Details on the carrier frequency and phase error estimation along with their discriminators normalization relations are provided in [10, 14]. It must be noted that the carrier tracking loop performance depends on the integration time, the loop order, and its bandwidth. The higher the loop order and the integration time and the lower the loop bandwidth are, the lower is the carrier tracking loop variance with a drawback leading to signal loss-of-lock condition [13].

5.2 Problematic in Urban Environment

The GNSS signal reception is harshly affected by the urban environment conditions, that are the following [28]:

- *Multipath*: defined as the reception of the direct LOS and the NLOS reflected or diffracted echoes of the direct ray (from the ground, buildings, foliage, lampposts, and so on).
- *Attenuation or blockage* of the GNSS LOS signal: a phenomenon arising due to the partial or total obstruction of the GNSS LOS from the urban environment characteristics.
- *Interference*: occurring due to the overlapping frequency bands of the telecommunication networks, comprising the VHF and UHF television signals and Digital Video Broadcasting -Terrestrial (DVB-T) system and the GNSS system.

The consequences of the above-mentioned urban environment error sources either expressed in the signal level or position domain are given next:

I *Bias* of the receiver's correlation function: between the received multipath-contaminated signal and the receiver's locally generated replica. This bias of the null-point significantly affects the signal tracking accuracy that is later translated into large degradations to the pseudorange accuracy and lower position accuracy [28].

II *Only NLOS signals reception*: occurs when the direct LOS GNSS signal is blocked and thus, only reflected signals are received. This results in a pseudorange measurement error and consequently a decrease of the positioning accuracy.

III *Inconsistent GNSS measurements*: when a set of pseudorange measurements, provided from one or more GNSS satellites in view, largely differ from the other received observations. This event is the outcome of the three error sources mentioned above [28].

These severe urban conditions consequently lead to inaccurate positioning up to loss of navigation availability. Many navigation solutions in signal-constrained environments rely on the integration of the GNSS system with Micro-Machined Electro-Mechanical sensors (MEMS), constantly providing a position solution regardless of the urban propagation channels. The increased interest in the MEMS sensor utilization is related to their small size (in cm order) and low cost due to

the silicon fabrication process. In the most common configurations, MEMS inertial units comprise accelerometers that provide the user position by double integrating the specific force f of the vehicle along its sensitive axis; MEMS gyroscopes, measuring the body rotational motion across each sensitive axis, w.r.t. the body sensor frame and 2- or 3-axes accelerometers and gyroscopes along with the magnetometers measuring the heading of the vehicle. In land navigation where only horizontal positioning is of great interest, a standalone position from the dead-reckoning MEMS sensor can be provided from the use of two gyroscopes and one accelerometer [28]. MEMS low performance in standalone mode makes them unsuitable to be used autonomously as a primary mean for positioning and navigation. Therefore, GNSS/MEMS integration yields many benefits due to the complementary nature of these two systems, where MEMS represents an offline system immune toward interference, jamming, and the propagation channel. Detailed description of classical GNSS/INS hybridization techniques can be found in [20, 33].

5.3 Advanced Signal Processing

The focus of the following section is dedicated to the advanced GNSS signal processing techniques, necessary to mitigate the undesired urban environment effects in order to ensure the accuracy and availability of the position solution.

5.3.1 GNSS NLOS Rejection Technique

As previously stated, a frequently occurring phenomenon in urban environment is the only NLOS signals reception. As a consequence, a degraded position accuracy is observed due to the introduced error on the pseudorange measurements. An interesting strategy proposed to overcome this phenomenon consists in the use of vision-aided sensors not for navigation purpose but instead for erroneous measurement rejection. For this matter, the use of a large field-of-view (typically 180°) Fisheye camera is proposed and analyzed in details. This camera is mounted on the vehicle roof and oriented upwards to capture the sky images while the vehicle is moving along its trajectory [28]. The output of this approach is the identification of the GNSS satellites located in the non-sky region, which will further be excluded from the position calculation. The Fisheye camera technique, for GNSS satellite detection, was firstly proposed in [1] and is based on sequential image processing steps, as illustrated in Fig. 5.5.

The starting point of the Fisheye technique is the image data acquisition. Once the data have been acquired, the image processing phase is initiated consisting of four successive steps:

- *Image simplification*: aiming at simplifying the image color levels into the definition of buildings and sky areas color representation;

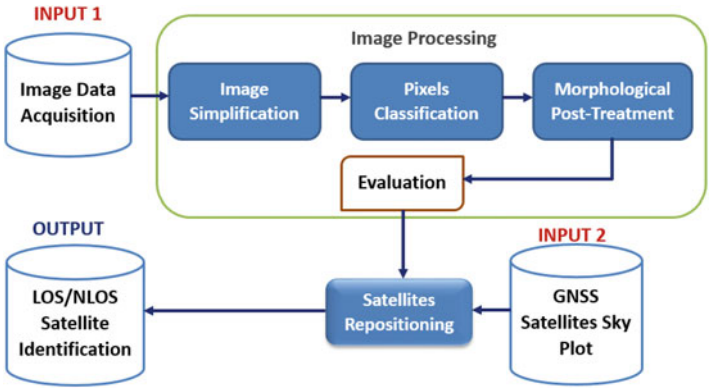

Fig. 5.5 Block diagram of GNSS LOS satellites detection using the Fisheye camera [28]

- *Pixels classification*: sky and not sky (including the buildings, vegetation, lampposts, etc.), by using image clustering classification algorithms;
- *Image post-treatment*: providing the reclassification of the uncertainty regions that do not directly fall into the two required categories. These doubt regions are mostly linked to the building reflections and tree shadowing;
- *Evaluation*: where the performance assessment of the three previous image processing techniques is carried out. In this step the image processing computational time is elapsed, as a crucial indicator for real-time applications [28].

The final stage of the Fisheye technique is the GNSS satellites repositioning, obtained from overlapping the satellites position in elevation and azimuth coordinates in the skyplot output of the GNSS receiver on board of the vehicle and the simplified Fisheye image. Concluding, only the measurements coming from unblocked or LOS satellites are fed to the navigation module. An alternative approach that aims at an increase of the position availability is de-weighting the measurements obtained from the satellites in the non-sky area. The detailed Fisheye image processing outputs from the measurement campaign carried out in Toulouse city center are illustrated in Fig. 5.6.

5.3.2 VT Technique

In Scalar Tracking (ST) configurations in the presence of weak signals or significant signal power drops, loss of lock of the affected satellite occurs and therefore, its estimated pseudoranges are not passed to the navigation processor due to their lack of accuracy. A promising approach for reducing the effect of multipath interference and NLOS reception is VT, first introduced in [23] where the signal tracking and navigation solution tasks are accomplished by the central navigation filter. In comparison to conventional or ST, where each visible satellite channel is being

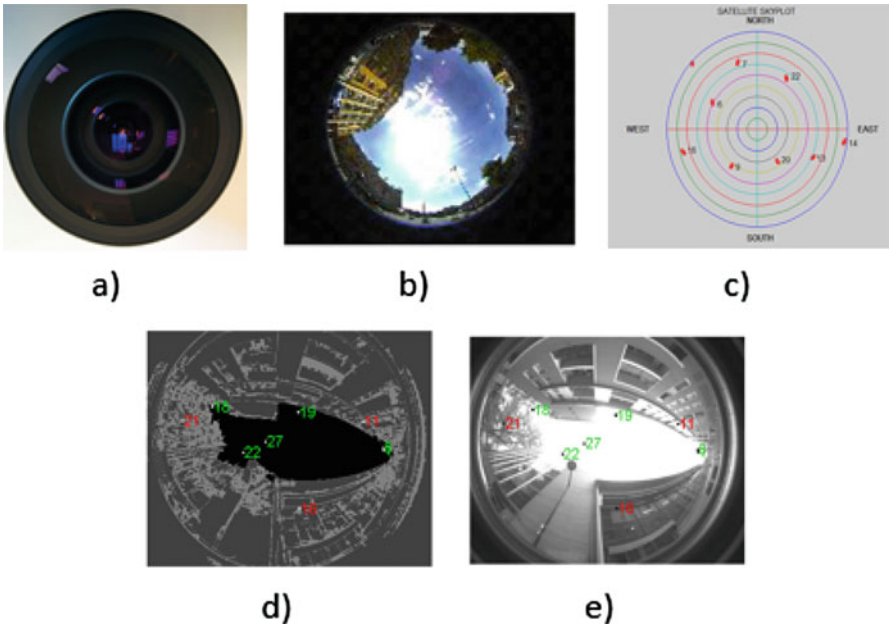

Fig. 5.6 Fisheye camera image processing outputs: (**a**) illustration of the Eye UI-3240CP camera used in the tests; (**b**) the image output of the Fisheye camera in the initial stage (Input 1) in the processing algorithm; (**c**) the GNSS skyplot of the satellites in view; (**d**) the image output of the pixels classification and morphological post-treatment steps; (**e**) the final output of the LOS detection algorithm [28]

tracked individually and independently, VT performs a joint signal tracking of all the satellite channels. Moreover, vector tracking exploits the knowledge of the receiver's position and velocity that can be further used to aid the receiver's tracking performance.

In [23], the Vector Delay Lock Loop (VDLL) architecture is exploited in details, for which the DLL filter loop is with an Extended Kalman Filter (EKF). In this configuration, the updated navigation solution from the EKF estimation will generated in the feedback loop the code NCO of each tracking channel while the carrier frequency/phase estimation is still achieved scalarly by the FLL/PLL. The GPS L1 VDLL tracking performance in weak signal reception and robustness against signal interference and attenuation has been demonstrated in [18, 22, 25, 32]. The objective of this section is to assess the performance of the Vector Delay Frequency Lock Loop (VDFLL) architecture, seen as a combination of the VDLL and Vector Frequency Lock Loop (VFLL) loops, in signal-constrained environment. From the navigation point of view, VDFLL can be seen as a fusion technique, since all the tracking channels NCO are controlled by the same navigation solution filter.

In particular, a dual architecture is presented since this type of receiver can significantly improve the availability of a navigation solution in urban canyons and heavily shadowed area due to an overdetermined number of satellite observations

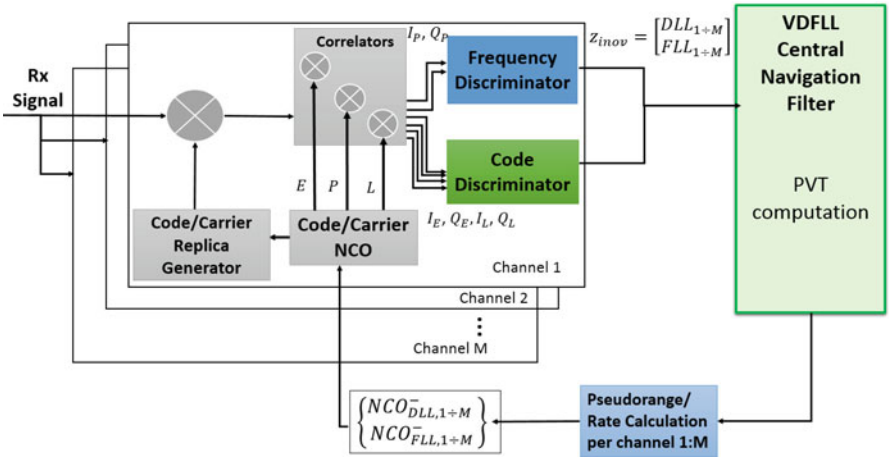

Fig. 5.7 The detailed block diagram of the proposed L1/E1 VDFLL architecture [29]

that also improves the position reliability [29]. The proposed VDFLL architecture comprises three sub-modules such as: the code/carrier tracking loops including the DLL/FLL discriminators, the EKF navigation filter, and the code/carrier NCO update. In this section, a dual constellation single frequency band L1/E1 VDFLL architecture is presented, wherein the code DLL and frequency FLL tracking loops are coupled through the navigation solution computed by the central EKF. The detailed architecture of the proposed L1/E1 VDFLL configuration is sketched in Fig. 5.7.

Kalman filter estimation equations employed in the navigation domain fall into two groups:

- *Time update (Prediction)* equations, performing the forward projection in time of the current vector state X_k^+ and the state error covariance matrix P_k^+ "a priori" estimates for the next time epoch, noted as X_{k+1}^- and P_{k+1}^-, where k indicates the current time epoch;
- *Measurement update (Correction)* equations, responsible for the feedback that is achieved by feeding the current epoch measurement vector, denoted as $z_i n$ into the a priori estimate, X_{k+1}^- and P_{k+1}^-, to obtain an improved "a posteriori" estimate, X_{k+1}^+ and P_{k+1}^+.

5.3.2.1 VDFLL EKF State Model

The chosen state vector model in the proposed EKF navigation filter implementation is the absolute PVT representation, containing the following states:

$$X_k = [x\ \dot{x}\ y\ \dot{y}\ z\ \dot{z}\ c \cdot t_{\text{GPS-clk}}\ c \cdot t_{\text{GAL-clk}}\ c \cdot \dot{t}_{\text{clk}}\ ]_k^T\ , \tag{5.10}$$

containing both the receiver's position vector $[x(k), y(k), z(k)]^T$ and the receiver's velocity vector $[\dot{x}(k), \dot{y}(k), \dot{z}(k)]^T$ expressed in Earth Centered Earth Fixed (ECEF) coordinates; the receiver's clock dynamics comprising the receiver clock bias w.r.t. the GPS and Galileo time and drift components $[c \cdot t_{\text{GPS-clk}}, c \cdot t_{\text{GAL-clk}}, c \cdot \dot{t}_{\text{clk}}]_k$, where c is the speed of light and therefore the clock biases and drift are expressed in unit of [m] and [m/s], respectively. Assuming that satellite clock biases are perfectly corrected, the bias increase (and thus the rate) only depends on the receiver's clock. Therefore, there is the same clock drift for both constellation.

The system model of the EKF filter in the continuous time domain may be expressed as:

$$\dot{X}(t) = \phi \cdot X(t) + B \cdot w(t) \ , \tag{5.11}$$

where $\dot{X}(t)$ denotes the derivative of the state vector $X(t)$, $w(t)$ is the centered gaussian white noise affecting the state vector, ϕ is the system matrix, and B is the colored noise transition matrix, both in the continuous time domain. Passing to the discrete time domain, the system or dynamic model of the VDFLL navigation filter can be detailed as follows:

$$X_k = \Phi \cdot X_{k-1} + w_k \ , \tag{5.12}$$

where: X_k denotes the state vector forward projection from the $k-1$th to the kth time epoch and Φ represents the dynamics of the user platform and clock, expressed as follows:

$$\Phi = \begin{bmatrix} C & 0_{2\times2} & 0_{2\times2} & 0_{2\times3} \\ 0_{2\times2} & C & 0_{2\times2} & 0_{2\times3} \\ 0_{2\times2} & 0_{2\times2} & C & 0_{2\times3} \\ 0_{3\times2} & 0_{3\times2} & 0_{3\times2} & C_{\text{clk}} \end{bmatrix}_{9\times9} \ , \tag{5.13}$$

where:

$$C = \begin{bmatrix} 1 & \triangle T \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad C_{\text{clk}} = \begin{bmatrix} 1 & 0 & \triangle T \\ 0 & 1 & \triangle T \\ 0 & 0 & 1 \end{bmatrix} \ , \tag{5.14}$$

and $\triangle T$ denotes the time interval between two consecutive estimations, representing the measurement update time of the central filter. The discrete process noise vector w_k is modeled as a white Gaussian noise vector with zero mean and discrete covariance matrix Q_k. The process noise w_k comes from two sources namely, the user dynamic noise $[w_x \ w_{\dot{x}} \ w_y \ w_{\dot{y}} \ w_z \ w_{\dot{z}}]_k^T$ (constituted by the user's position and velocity terms) and the receiver's clock noise (local oscillator NCO noise) $[w_b \ w_d]_k^T$ grouped in a single vector representation as:

$$w_k = [w_x \ w_{\dot{x}} \ w_y \ w_{\dot{y}} \ w_z \ w_{\dot{z}} \ w_{b\text{-GPS}} \ w_{b\text{-GAL}} \ w_d]_k^T \ , \tag{5.15}$$

Based on their nature, the five tuning factors of process noise Q covariance matrix can be grouped in two main categories, such as:

- *User's dynamics sensitive*: including the velocity error variance terms along the ECEF axes $(\sigma_{\dot{x}}^2, \sigma_{\dot{y}}^2, \sigma_{\dot{z}}^2)$ that will be projected in the position domain through the state transition matrix Φ and the colored noise transition matrix B from Eq. (5.11);
- *Receiver's oscillator noise PSD*: including the oscillator's phase noise PSD, σ_b, and the oscillator's frequency noise PSD, σ_d, which by themselves depend on the oscillator's Allan variance parameters h_0 and h_{-2} [26].

The process noise covariance matrix $Q_k = \text{diag}[Q_{x,k}, Q_{y,k}, Q_{z,k}, Q_{c,k}]$ in the discrete domain per each entry is given in details in [29].

5.3.2.2 VDFLL EKF Observation Model

The non-linear relation between the state and the measurement vector is expressed as follows:

$$z_k = h(X_k) + v_k \ , \tag{5.16}$$

where h is the *non-linear function* relating the measurement z_k to the state X_k and v_k is the measurement noise vector that is modeled as a zero-mean uncorrelated Gaussian noise process and independent to the process noise w_k. The measurement vector z_k comprises the pseudoranges ρ_j and Doppler measurements $\dot{\rho}_j$, output from the code/carrier tracking process for the $j = 1 \cdots M$ L1/E1 tracking channels:

$$z_k = [(\rho_1, \rho_2, \ldots, \rho_M), (\dot{\rho}_1, \dot{\rho}_2, \ldots, \dot{\rho}_M)] \ , \tag{5.17}$$

In the Cartesian ECEF-frame implementation, the pseudoranges $\rho_{j,k}$ per each tracked satellite j are computed as:

$$\begin{aligned} \rho_{j,k} = & \sqrt{(x_{\text{sat}j,k} - X_k(1))^2 + (y_{\text{sat}j,k} - X_k(3))^2 + (z_{\text{sat}j,k} - X_k(5))^2} \\ & + X_k(7)(X_k(8)) + n_{\rho_j,k} \ , \end{aligned} \tag{5.18}$$

While the remaining M-entries of the measurement vector, constituted by the Doppler measurements, are computed as:

$$\begin{aligned} \dot{\rho}_{j,k} = & (\dot{x}_{\text{sat}j,k} - X_k(2)) \cdot a_{x,j} + (\dot{y}_{\text{sat}j,k} - X_k(4)) \cdot a_{y,j} \\ & + (\dot{z}_{\text{sat}j,k} - X_k(6)) \cdot a_{z,j} + X_k(9) + n_{\dot{\rho}_j,k} \ , \end{aligned} \tag{5.19}$$

where $(a_{x,j}, a_{y,j}, a_{z,j})$ denote the LOS unit vectors from the receiver to the jth satellite along the X, Y, and Z axes and $(n_{\rho_j,k}, n_{\dot{\rho}_j,k})$ denote the zero-mean Gaussian-distributed noise affecting the pseudorange and Doppler measurements,

respectively. The measurement noise vector v_k is modeled as a zero-mean uncorrelated Gaussian noise process and independent to the process noise w_k and with a measurement covariance matrix R_k having in the main diagonal the following entries:

$$R_{jj} = \begin{cases} \sigma^2_{\mathrm{DLL},j} \text{ for } j = 1 \ldots M, \\ \sigma^2_{\mathrm{FLL},j} \text{ for } j = 1 \ldots M. \end{cases} \tag{5.20}$$

where the first entry refers to the pseudorange error variance terms for the tracked GPS and Galileo satellites, while the second ones are a common term for the pseudorange rate error variances for all tracked satellites.

5.3.2.3 VDFLL Estimation Workflow

The sequential steps followed by the vectorized architecture in terms of state vector update are depicted in Fig. 5.8.

Following the VDFLL estimation workflow of Fig. 5.8, the successive step after the state propagation or prediction is the computation of the Kalman gain in Step 2.1. For this matter, the measurement prediction z_k and observation matrix H_k shall be calculated. Afterward, the state vector update is computed from the measurement innovation vector input to the EKF navigation filter, which comprises the code and carrier discriminator outputs from the tracking loops. Finally, the code and carrier NCO update, computed from the EKF filter prediction states, closes the feedback loop to the tracking module. Extensive description of the L1/E1 VDFLL estimation process is detailed in [29].

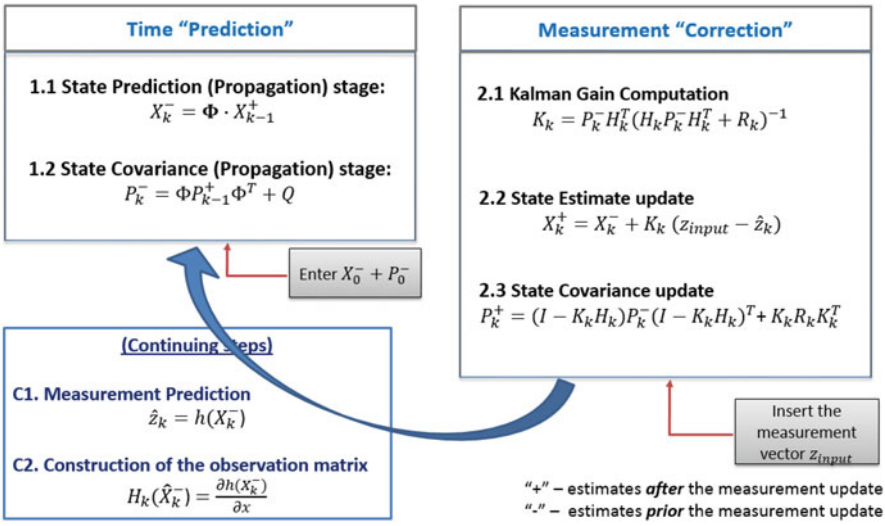

Fig. 5.8 The L1/E1 VDFLL estimation filter workflow [29]

5.3.2.4 Performed Tests in Urban Conditions

In order to test the performance of the proposed L1/E1 VDFLL architecture, a GNSS signal emulator compiled in C language owned by Ecole Nationale de l'Aviation Civile (ENAC), able to generate GPS L1 and Galileo E1 signals up to 54 channels simultaneously, was used. Moreover, the vector tracking algorithm is software implemented in C language, driven by the faster execution time of EKF algorithm at high rates (set equal to the tracking outputs at 50 Hz, or $T_{\mathrm{EKF}} = 20$ ms). Three distinctive GNSS receiver architectures will be analyzed with the scope of performance comparison:

- Scalar tracking employing a 3rd order loop PLL and a DLL, with an EKF positioning module at 1 Hz for the PVT computation, where the pseudorange and Doppler measurements are included in the observation vector;
- The same scalar tracking architecture but now integrated with an EKF positioning module at 50 Hz, similar to the VDFLL algorithm update rate;
- The proposed VDFLL EKF architecture working at $T_{\mathrm{EKF}} = 20$ ms integration time and thus providing 50 Hz code and carrier frequency updates.

It must be noted the KF positioning module is similar to the EKF filter of the vectorized solution, with the differences that a closed-loop measurement covariance matrix is used in the former and moreover, the KF filter operates on locked satellites only whereas the VDFLL uses all satellites in view. The simulations performed are related to a real car trajectory in Toulouse urban area. The simulated reception conditions will consist in several signal outages and significant power drops simulated in different satellite channels in order to observe the tracking performance of the proposed VDFLL architecture with respect to conventional tracking. In both test scenarios, there is a maximum of 13 simultaneously tracked GPS L1 and Galileo E1 channels during 200 GPS epochs. A detailed performance comparison between the scalar and vectorized configurations will be assessed in the system level. This is expressed in terms of user's position and velocity estimation accuracies, position and velocity errors statistics, and resistance to degraded signal reception conditions.

An RF front-end with a 24 MHz bandwidth (double-sided) is assumed. Multiple outages on different tracked satellites have been simulated, by generating a sudden drop of the C/N_0 ratio down to 20 dB-Hz that coincides with the C/N_0 level in quasi-indoor environment. Therefore, the GNSS signals exhibiting this low C/N_0 level will cause the tracking loops to experience a loss of lock condition. The outage conditions were simulated in three time epochs as depicted in Fig. 5.9:

1. Outage 1 from the 2nd - 12th time epoch (10 s);
2. Outage 2 from the 60th - 80th time epoch (20 s);
3. Outage 3 from the 140th - 160th time epoch (20 s).

In particular, a realistic car trajectory in high dynamic condition is generated based on the reference trajectory computed by the NovAtel's SPAN receiver mounted on car during a 40 min measurement campaign conducted in Toulouse

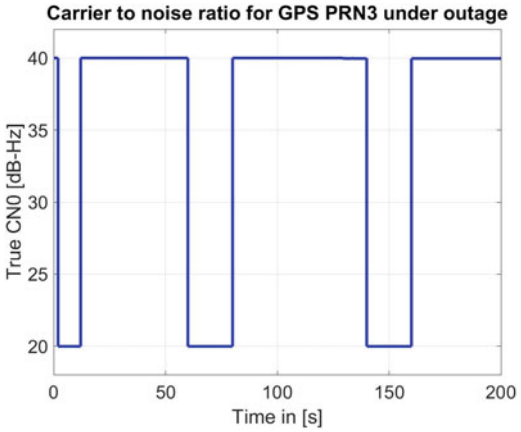

Fig. 5.9 Illustration of C/N_0 ratio drops to 20 dB-Hz applied to four satellite namely, GPS PRN 3 and 4 and GAL PRN 51 and 52 during three outages intervals

Fig. 5.10 The simulated car trajectory in Toulouse area [29]

in April 2016. It must be noted that the simulated car path of 200 s duration is a representative of a car trajectory but not of urban signal reception conditions since the simulated received conditions are generated from an open sky environment plus the forced drops of received signal C/N_0 values. The simulated car path in Google Earth format is shown in Fig. 5.10.

Moreover, the position domain comparison is extended to the scalar tracking architecture with the KF module working at the same rate as the vectorized architecture that is 50 Hz. The position error plots in the ECEF frame illustrated

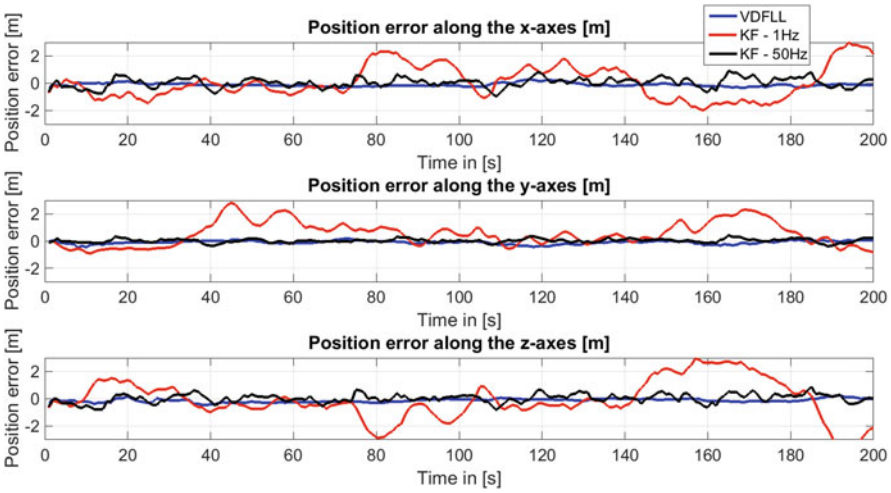

Fig. 5.11 Position error comparison between the scalar tracking + KF positioning module at 1 Hz (in *red*), the VDFLL algorithm (in *blue*), and the scalar tracking + KF positioning module at 50 Hz (in *black*) for a car trajectory under signal outages along the: (**a**) *X*-axis in [m]; (**b**) *Y*-axis in [m]; (**c**) *Z*-axis in [m] [29]

in Fig. 5.11 demonstrate a clear convergence of the VDFLL-computed navigation solution to the reference trajectory within a standard deviation of 0.2 m during the outages. The positioning RMSE per each ECEF axes illustrated in Fig. 5.11 highlight the positioning robustness of the vectorized architecture during both outages and high C/N_0 reception conditions.

It is obvious the positioning error variance increase of the scalar tracking + KF module operating at 1 Hz during the outages intervals, with an error peak up to 2.75 m at the 80th time epoch. Moreover, a sudden navigation error increase is also observed at the 190th epoch coming from the last car turn prior to arriving at the end of the path in Fig. 5.10. Concerning the proposed VDFLL algorithm, a smoother position error in the ECEF frame is marked along the overall car trajectory, with significantly smaller errors. An interesting result is observed for the scalar architecture employing the KF positioning module at 50 Hz rate that clearly outperforms its counterpart running at 1 Hz, in terms of position error bounding. The explanation lies on the higher rate of the state propagation and measurement innovation process, providing a faster convergence of the position estimations to the reference trajectory. Nevertheless, quasi similar positioning errors are observed w.r.t. to the VDFLL architecture and this comes due the overdetermined number of observables (13 tracked satellite channels in total). The tracking robustness of the VDFLL architecture under sudden signal power drops is fully detailed in [29, 32]. It must be highlighted the fact that the vector tracking technique keep continuous track of the received signal with the signal reacquisition requirement of the previously "loss of lock" channel.

5.4 Carrier Phase Measurements in Urban Environments

Previous considerations referred to the most typical GNSS observables: the code pseudorange and Doppler frequency shift (Doppler). The observable that can provide the best accuracy is the carrier phase. There is a growing popularity of carrier phase observation analysis in urban environments, and the Commercial off-the-shelf (COTS) receiver manufacturers have started to provide access to such measurements. The processing of carrier phase to obtain pseudoranges is complex. In fact, this observable is vulnerable to any obstructions of the signal, resulting in the so-called *cycle slips*.

Carrier phase measurements are generated in the receiver along to code pseudorange and Doppler frequency shift after the signal lock on a particular satellite. This observable is generated in the carrier tracking loop of the receiver, and is determined during the process of integrating carrier Doppler frequency offset over the epoch interval. The fractional carrier phase in the end of the integration epoch is recorded as a measurement. It might be described mathematically by Eq. [14]:

$$\Phi_{L_n} = \Phi_{L_{n-1}} + \int_{t_{n-1}}^{t_n} f_D(\tau) d\tau + \Phi_{r_n}, \tag{5.21}$$

where Φ_L is the accumulated carrier phase for particular L frequency, n and $n-1$ are the current and previous epoch, f_D is the carrier Doppler frequency offset as a function of time τ, and Φ_r is the fractional carrier phase output. It is worth noting that for $n = 0$, the receiver determines its own arbitrary cycle count A_L for a particular frequency.

The measurement resolution in most modern receivers is typically better than 0.1 mm [12]. However, this measurements are biased due to the fact that the whole integer number of cycles between the satellite and the receiver is unknown and has to be determined. This value is called the integer carrier phase ambiguity, and it has to be resolved to fully benefit from the high precision of this type of measurements.

The ambiguity resolution process has therefore a crucial influence on the accuracy of the solution. When the ambiguity is estimated as an integer value, the convergence time for the solution shortens. The description of techniques providing Integer Ambiguity Resolution (IAR) is presented in [15]. The IAR process usually consists of three main steps. The first one is the generation of potential integer ambiguity combinations that would be evaluated by the algorithm. This is done in the search space, the volume of uncertainty surrounding the antenna location. In static positioning, search space can be constructed from float ambiguities, while for kinematic positioning it is realized by code range solution. Decreasing the dimension of the search space would increase efficiency of calculations, which is important for kinematic solutions. The second step is the identification of correct integer ambiguity combinations. The key aspect in it is having enough redundant satellites, as the criterion used by many of the techniques is based on the minimization of the sum of the squared residuals in a least squares adjustment

approach. The third step is the validation of the ambiguities. It can be done by the ambiguity success rate, depending on following factors: observation equations, precision of observables, and the method of integer ambiguity estimation itself.

Potential difficulties in IAR process can be placed in several issues:

- Assumption of normal distribution of the residuals; errors like multipath, orbit errors, atmospheric errors influence it, and this is the main reason of failure of solution for long baselines;
- Decision on statistical significance of ambiguity decision, as the integer ambiguity combination fitting the measurements in the best way should be significantly better than all the others.

In urban scenarios, the IAR process will be therefore much harder, because of presence of multipath and NLOS observations caused by high power signal reflections, reduced LOS signal visibility and frequent geometry changes due to signal blockage. The urban and suburban propagation channels were described, e.g., in [19, 30]. Furthermore, the ambiguity resolution process can be done only for float ambiguities, sacrificing slightly the measurement precision and benefitting from faster solution, as in [17].

According to [24], the availability of carrier phase signal in urban environment is limited: in open sky conditions, a two-constellation receiver phase locks usually to 15 satellites, while for urban it is usually 8 satellites (Fig. 5.12). What is more, the mode duration of observation epoch in urban scenarios is 1 s, median reaches up

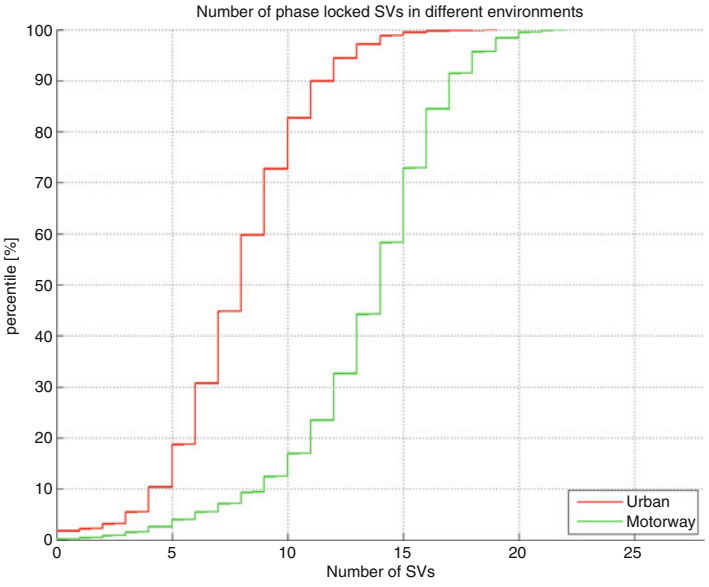

Fig. 5.12 Cumulative distribution function of the number of visible satellites in different environments

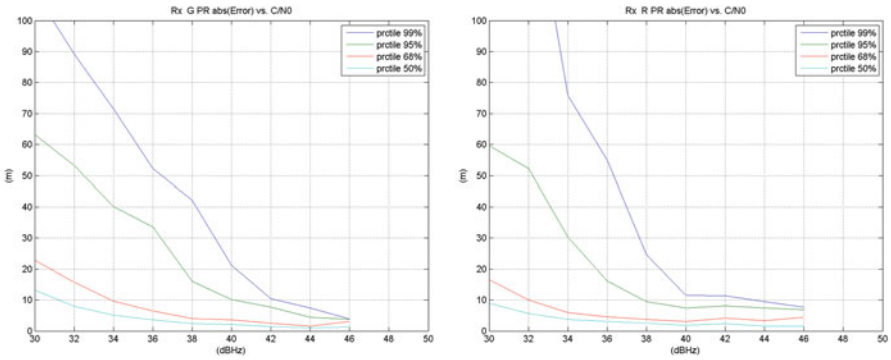

Fig. 5.13 Receiver code pseudorange error vs. C/N_0 values: GPS (*left*) and GLONASS (*right*)

to 7 s, where the lack of observation periods are actually longer: the mode equals to 3 s, and median value up to 11 s. These results show the lack of availability of carrier phase observations in urban environments, which causes the problem even in the beginning of the processing of these observables. This fact combined with the problem of degradation of code pseudoranges quality in urban environments for the observations with C/N_0 lower than 40 dB-Hz ([6, 24], Fig. 5.13) and the occurrence of cycle slips cause the problems of processing the carrier phase observations in urban environments.

A cycle slip is a discontinuity in the carrier phase measurement that is the consequence of a loss of lock in the carrier tracking loop of the GNSS receiver. Usually, there are four causes of cycle slips mentioned in the literature, according to [12, 13]:

- signal obstruction,
- low signal-to-noise ratio,
- failure in the receiver software,
- receiver dynamics.

When one of these happens, the cycle count re-initializes, causing a jump in the fractional carrier phase output Φ_r by an integer number of cycles. This results in the change of ambiguity of tracked carrier phase, causing a new estimation of this parameter. When the cycle slip remains undetected, it degrades the accuracy of the precise positioning inducing uncontrolled error propagating into the position estimation.

Cycle slip detection process is crucial for precise positioning methods, and varies greatly depending on whether the receiver is single- or multi-frequency. Multiple frequencies allow for constructing combination of observations. A specific combination of carrier phase measurements called *geometry-free combination* removes the influence of geometry, including clocks, and all non-dispersive effects from the signal. This results in very precise and smooth signal, suffering mainly from the ionosphere changes that, without occurrence of scintillations, can be

removed by a low degree polynomial fit. Another possibility for the detection is the use of code and carrier phase observations in Melbourne-Wübenna combination, which, however, suffers from code pseudorange multipath influence. For more than two frequency receivers, a combination of these two methods can successfully be applied.

Single frequency approach is more complicated because of lack of any possible source of reference. According to [5], they can be divided into two categories:

- statistical testing, utilizing an assumption of certain distribution of carrier phase residuals, and
- phase prediction method, where the measured carrier phase is compared to the predicted one, computed using Doppler and inertial data.

However, single frequency cycle slip detection and repair is still challenging. Especially, the movement of the receiver may cause problems in precise Doppler shift estimation [7] or might be challenging in integration of high precision positioning engine and INS measurements [31]. For a moving receiver, it is recommended to solve ambiguities every single epoch [2]. Another attitude is to estimate the cycle slips as integer values in a similar way that ambiguities are solved [5].

Carrier phase measurements are mainly used in two high precision positioning techniques in urban environments: Real Time Kinematics (RTK) and Precise Point Positioning (PPP). RTK is a type of kinematic relative positioning method, in which differencing observables between receivers, between satellites, and between time can result in eliminating common errors between them, e.g., relating to satellite and receivers' clocks. In the principle of the kinematic method, one receiver remains fixed, while the other one moves, and its position is to be determined for arbitrary epochs. In real-time situations, the ambiguity solution has to be solved immediately, thanks to observations from the fixed receiver and knowing its position. Decorrelation of error sources limits this method to about 20 km baselines. In Wide Area RTK (WARTK) [11] implementing ionospheric corrections allows a fast decorrelation and preserving integer nature of ambiguities. RTK and WARTK need time for convergence, up to several minutes, allowing for solutions of baselines up to 20 and 400 km, respectively.

The horizontal accuracy of 1σ level of RTK solution in good observation conditions and with properly solved ambiguities, for five tracked satellites in open sky conditions, can be given as 5 cm + 5 ppm, depending on the length of the baseline [12]. Unfortunately, this is not usually true in urban environments. The degradation of solution caused by urban canyons can obtain values of several tens of meters, completely degrading the precision provided by the carrier phase measurements.

PPP is the method of absolute positioning using accurate orbital and clock data provided by external sources, along with dual frequency code pseudoranges and/or carrier phase observations. The determined parameters are the position, receiver clock error, tropospheric delay, and ambiguities. Taking into account more effects like Sagnac effect, solid earth tides, ocean loading atmospheric loading, polar motion earth orientation effects, crustal motion, antenna phase center models,

and antenna phase wind-up allow for refining the model and improving resulting accuracy. Weighting of observations allows for improving further the accuracy.

PPP theoretically allows an easier detection of anomalous behaviors. The key issue influencing the solution is the quality of reference data (orbits and clocks) and dual frequency observations. It is, however, possible to utilize single frequency observations, but the ionosphere information have to be provided together with orbit and clock data. Problematic aspects in PPP technology are:

- poor observability of some parameters causing high correlations, that are undesirable in convergence process,
- systematic errors in orbits and clocks,
- environmental obstruction influencing convergence process,
- communication losses in real-time scenarios can lead to degradation of solution (while RTK needs a real-time communication link with the base station, PPP can withstand much longer communication losses).

As stated in [17] the main indicators for PPP integrity issues can be stated as the following:

- definition of the terrestrial reference frame,
- necessary covariance indicators from PPP estimation filter,
- residuals from phase measurements, especially in urban or poor visibility scenarios.
- quality of the orbits and clocks,
- convergence period has to be taken into account to compensate for the strong initial correlations between different parameters.

Taking into account the factors mentioned above, carrier phase positioning techniques have to be treated differently in terms of the measures of trust that the user can have in the estimated position. Relative positioning technique like RTK provides the solution with respect to the base station, and thanks to that it eliminates to a great extent the common errors for both receivers. Absolute technique like PPP provides the solution usually in the global reference frame. It is much more vulnerable therefore to the quality of the data utilized in the estimation process, but is independent from the availability of the reference data. Both techniques have to implement the resolution of ambiguity. PPP uses undifferenced observations, and therefore also the ambiguity resolution is performed without any reference, utilizing wide and narrowlane combinations of data, often together with the information from the network of local stations (PPP-RTK concept, [20]). For RTK technique, however, an important parameter that has to be included into the integrity algorithm is the differenced relative ambiguity. For both methods utilized in real time, the communication channel is crucial and has to be accounted for as a potential source of failure in estimating the measure of trust to the position provided to the user. We can see therefore that both of these techniques have similar problems, but have to be treated differently in the process of determining the trust to the outputs.

Preliminary results of integrity and reliability algorithm were presented in [17]. Very interesting results on providing integrity for PPP method were showed in

[21], where the statistical method based on different multivariate t-distributions called Kalman Integrated Protection Level method, an extension of Isotropy Based Protection Level method, has been applied to PPP, showing high accuracy PPP solutions and protection levels of few decimeters. Similar attitude might in future also be applied to the RTK method.

5.5 Conclusions

In this chapter, the urban environment-induced challenges to GNSS signal reception, that severely deteriorate the positioning accuracy, were shortly presented. Moreover, great emphasis was given to the description and understanding of the GNSS receiver architecture focusing also on the signal tracking process up to the measurement generation, required for the computation of the navigation solution. To overcome the urban environment constraints, strongly affecting the positioning performance, three distinct approaches were proposed and later analyzed in details.

The first proposed strategy consisted in the GNSS/Video sensor coupling, through the use of a Video Fisheye camera that is mounted on the vehicle roof and oriented upwards to capture the sky images. This approach is a measurement rejection technique, aiming at discarding the NLOS received signals prior to their use in the navigation filter for the PVT computation.

Secondly, the spotlight was directed to the analysis of the vector tracking mechanism, proposed as an advanced signal processing and estimation algorithm in the GNSS domain. In specifics, according to this approach, a joint channel's signals tracking is performed by a single KF filter that is in charge of performing two main tasks such as: first, computation of the navigation solution and second, the simultaneous signal tracking estimation process. In details, the dual constellation single frequency GPS L1/Galileo E1 VDFLL algorithm was implemented, aiming at an increase of observation and in the same time, conserving the simplicity of the signal's estimation process. In this architecture, the code and carrier filter loops are abolished and instead replaced by the central EKF filter performing both the code and Doppler estimations in the feedback loop to all the tracked channels. The L1/E1 VDFLL positioning robustness in urban car trajectory under week signal power reception and sudden C/N_0 drops simulated in different satellite channels was assessed. Contrary to conventional scalar tracking technique, the vectorized algorithm is able to recover the frequency and code delay estimation at the end of the simulated outages without the requirement of signal reacquisition within smaller position error limits. The likely reason for this behavior is linked to the inter-channel aiding through the update process based on the forward position/velocity projection in the vectorized architecture.

Finally, the description of carrier phase measurements' behavior in urban scenarios was described, showing their vulnerability and availability problems. These measurements, however, can provide the most precise measurements, therefore they shouldn't be rejected, but taken into account in navigation solution in harsh

environments. Two main techniques utilizing them were briefly presented, showing strengths and weaknesses of RTK and PPP techniques. Furthermore, a short discussion of the requirements for the estimation of the measures of trust—or integrity—that the user can place in the estimated position was presented, showing the weak spots that have to be taken into account. However, in the time of the preparation of this book, the research in this topic is still developing.

Acknowledgements This work was financially supported by EU FP7 Marie Curie Initial Training Network MULTI-POS (Multi-technology Positioning Professionals) under grant nr. 316528.

References

1. D. Attia et al., Image analysis based real time detection of satellites reception state, in *2010 13th International IEEE Conference on Intelligent Transportation Systems (ITSC)* (IEEE, 2010), pp. 1651–1656
2. M. Bahrami, M. Ziebart, Instantaneous Doppler-aided RTK positioning with single frequency receivers, in *Position Location and Navigation Symposium (PLANS), 2010 IEEE/ION* (IEEE, New York, 2010), pp. 70–78
3. S. Bhattacharyya, D. Gebre-Egziabher, Development and validation of parametric models for vector tracking loops. Navigation **57**(4), 275–295 (2010)
4. K. Borre et al., *A Software-Defined GPS and Galileo Receiver: A Single-Frequency Approach* (Springer Science & Business Media, Berlin, 2007)
5. S. Carcanague, Real-time geometry-based cycle slip resolution technique for single-frequency PPP and RTK, in *Proceedings of the 25th International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS 2012)* (2012), pp. 1136–1148
6. S. Carcanague, Low-cost global positioning system (GPS)/Globalnaya Navigazionnaya Sputnikovaya Sistema (GLONASS) precise positioning algorithm in constrained environment. Ph.D. thesis, Universite de Toulouse, 2013
7. P. Cederholm, D. Plausinaitis, Cycle slip detection in single frequency glsGPS carrier phase observations using expected Doppler shift. Nordic J. Surv. Real Estate Res. **10**(1), 63–79 (2014)
8. G. Di Giovanni, S.M. Radicella, An analytical model of the electron density profile in the ionosphere. Adv. Space Res. **10**(11), 27–30 (1990)
9. F. Dovis, P. Mulassano, Introduction to Global Navigation Satellite Systems. Politecnico I Torio (2009)
10. P.D. Groves, *Principles of GNSS, Inertial, and Multisensor Integrated Navigation Systems* (Artech House, London, 2013)
11. M. Hernandez-Pajares et al., Wide area real time kinematics with Galileo and GPS signals, in *Proceedings of the 17th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS 2004)* (2004), pp. 2541–2554
12. B. Hofmann-Wellenhof, H. Lichtenegger, E. Wasle, *GNSS - Global Navigation Satellite Systems: GPS, GLONASS, Galileo, and More* (Springer, Wien, 2008)
13. O. Julien, *Design of Galileo L1F Receiver Tracking Loops*, Library and Archives Canada, Bibliothque et Archives Canada, 2006
14. D.E. Kaplan, J.C. Hegarty, *Understanding GPS: Principles and Applications*, 2nd edn. (Springer/Artech House, Wien/New York, 2006)
15. D. Kim, R.B. Langley, GPS ambiguity resolution and validation: methodologies, trends and issues, in *Proceedings of the 7th GNSS Workshop-International Symposium on GPS/GNSS*, Seoul, Korea, 30, No. 2.12 (2000)
16. J.A. Klobuchar, Ionospheric time-delay algorithm for single-frequency GPS users. IEEE Trans. Aerosp. Electron. Syst. **3**, 325–331 (1987)

17. M.D. Lainez Samper, M.M. Romay Merino, In-the-field trials for real-time precise positioning and integrity in advanced applications, in *Proceedings of the ION 2013 Pacific PNT Meeting* (Apr. 2013), pp. 146–167
18. M. Lashley, D.M. Bevly, J.Y. Hung, A valid comparison of vector and scalar tracking loops, in *Position Location and Navigation Symposium (PLANS), 2010 IEEE/ION* (IEEE, 2010), pp. 464–474
19. A. Lehner, A. Steingass, A novel channel model for land mobile satellite navigation, in *Proceedings of the 18th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS 2005)* (2005), pp. 2132–2138
20. L. Mervart et al., Precise point positioning with ambiguity resolution in real-time, in *Proceedings of the 21st International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS 2008)* (2008), pp. 397–405
21. P.F. Navarro Madrid, M.D. Lainez Samper, M.M. Romay Merino, New approach for integrity bounds computation applied to advanced precise positioning applications, in *Proceedings of the 28th International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS+ 2015)* (2015), pp. 2821–2834
22. T. Pany, R. Kaniuth, B. Eissfeller, Deep integration of navigation solution and signal processing, in *Proceedings of the 18th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS 2005)* (2001), pp. 1095–1102
23. B.W. Parkinson, *Progress in Astronautics and Aeronautics: Global Positioning System: Theory and Applications*, vol. 2 (AIAA, Washington, DC, 1996)
24. M.J. Pasnikowski et al., Challenges for integrity in navigation of high precision, in *Proceedings of the 28th International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS+ 2015)* (2015), pp. 2983–2994
25. M.G. Petovello, G. Lachapelle, Comparison of vector-based software receiver implementations with application to ultra-tight GPS/INS integration, in *Proceedings of Institute of Navigation GPS/GNSS Conference Fort Worth, TX. Institute of Navigation* (2006)
26. M.L. Psiaki, Smoother-based GPS signal tracking in a software receiver, in *14th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GPS 2001)* (2001), pp. 2900–2913
27. G. Seco-Granados et al., Challenges in indoor global navigation satellite systems: unveiling its core features in signal processing. IEEE Signal Process. Mag. **29**(2), 108–131 (2012)
28. E. Shytermeja, A. Garcia-Pena, O. Julien, Proposed architecture for integrity monitoring of a GNSS/MEMS system with a Fisheye camera in urban environment, in *2014 International Conference on Localization and GNSS (ICL-GNSS)* (IEEE, New York, 2014), pp. 1–6
29. E. Shytermeja, A.G. Pena, O. Julien, Performance comparison of a proposed vector tracking architecture versus the scalar configuration for a L1/E1 GPS/Galileo receiver, in *Proceedings of European Navigation Conference Helsinki*, Finland (2016)
30. A. Steingass, A. Lehner, Differences in multipath propagation between urban and suburban environments, in *Proceedings of the 21st International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS 2008)* (2008), pp. 602–611
31. T. Takasu, A. Yasuda, Cycle slip detection and fixing by MEMS-IMU/GPS integration for mobile environment RTK - GPS, in *Proceedings of the 21st International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS 2008)* (2008), pp. 64–71
32. X. Tang et al., Theoretical analysis and tuning criteria of the Kalman filter-based tracking loop. GPS Solutions **19**(3), 489–503 (2015)
33. D. Titterton, J.L. Weston, *Strapdown Inertial Navigation Technology*, vol. 17 (IET, London, 2004)

Chapter 6
Multi-GNSS: Facts and Issues

Nunzia Giorgia Ferrara, Ondrej Daniel, Pedro Figueiredo e Silva, Jari Nurmi, and Elena-Simona Lohan

6.1 Introduction

GNSS satellites continuously transmit RF signals in the L band, allowing passive receivers to determine PVT. Such signals have three main components: carrier, ranging code and navigation data. The carrier is an RF sine wave at a given frequency modulated with the combination of the ranging and navigation components. The ranging code is a binary sequence which allows the computation of the distance to the satellite. These sequences are known as PRN codes. The navigation data, a binary-coded message, contains essential information for positioning computation, such as satellite ephemeris, clock bias parameters, almanac, satellite health status and other complementary information. A detailed description of the GNSS signal and its processing can be found in Chap. 3.

Satellite-based positioning and timing are well-established fields, which are also significantly growing, with more satellites, more signals and more applications on the way. The United States of America's GPS was the first globally available GNSS. However, positioning and timing are no longer just about GPS. Other GNSS are fully operational today, such as the Russian GLONASS, or under development, such as the European Galileo and the Chinese BeiDou. When their full operational capability will be reached, as expected by 2020, more than one hundred GNSS satellites will be available for positioning and timing purposes, offering different types of services on a worldwide basis. Therefore, the concept of GNSS receiver is evolving from the old conception of a stand-alone receiver to a receiver that is able

N.G. Ferrara (✉) • O. Daniel • P. Figueiredo e Silva • J. Nurmi • E.-S. Lohan
Tampere University of Technology, Korkeakoulunkatu 10, 33720 Tampere, Finland
e-mail: nunzia.ferrara@tut.fi; jari.nurmi@tut.fi; elena-simona.lohan@tut.fi

J. Nurmi et al. (eds.), *Multi-Technology Positioning*,
DOI 10.1007/978-3-319-50427-8_6

to process modern signals coming from more than one system and with advanced structures and properties. This multi-GNSS scenario will bring several advantages, such as increase in accuracy, availability and reliability of positioning solutions. The modernisation of GPS and GLONASS and the introduction of new systems will help reduce the GNSS limitations in challenging environments, such as urban areas, where the signals might be contaminated by non-line-of-sight reception or multipath propagation. Moreover, the multi-constellation diversity will be an advantage in dealing with RFinterference, which represents one of the major threats to GNSS, since, due to the long travelled distance, the satellite signals are very weak when they reach the Earth. On the other hand, the multi-constellation integration presents several issues and challenges to be overcome.

This chapter describes the current and future GNSS, and provides an overview of the advantages and the issues of the multi-constellation environment.

6.2 Global Navigation Satellite Systems

Every GNSS consists of three components: space, ground and user segment. The space segment is the constellation of satellites transmitting radio signals from which the ranging measurements are made, the ground network is responsible for maintaining the satellites and ensuring their correct functioning, and the user segment consists of the GNSS receivers which process the satellite signals to obtain a PVTsolution.

This section provides the reader with an overview of the GNSS, with a focus on constellation, transmitted signals, and time and coordinates reference frames of each system.

6.2.1 GPS

GPS is the first globally available and most widely used GNSS. It reached its full operational capability in 1995, and since then it has been continuously operating, providing reliable timing, positioning and navigation services to civilian and military users on a worldwide scale. To meet growing demands and remain competitive with other systems, the US government has started a modernisation program of the GPS space and control segments with the introduction of new features to enhance the system performance, including new signals, both military and civilian.

6.2.1.1 Space Segment

GPS original constellation consisted of 24 satellites flying in Medium Earth Orbit (MEO) at an altitude of approximately 20,200 km and with a nominal period of 11 h 58 min 2 s, repeating the geometry every sidereal day. They are distributed in six equally spaced orbital planes with an inclination of 55° with respect to the equator. Each plane contains four slots occupied by baseline satellites. This configuration allows users to view at least four satellites from virtually any point on Earth. However, in order to ensure coverage in case of baseline satellites malfunctioning, the US Air Force, responsible for developing, maintaining and operating the space segment, has been adding extra satellites that may improve the system performance but are not acknowledged as part of the core constellation.

The space segment is constantly subject to maintenance and evolutions, and, as of May 2016, 31 satellites are operational in the GPS constellation.

6.2.1.2 Current and Planned Signals

The carrier frequencies of the legacy GPS signals are called L1 and L2. They are obtained from a fundamental frequency $f_0 = 10.23\,\text{MHz}$, generated by atomic clocks onboard the satellites.

$$\text{L1} = 154 \times 10.23\,\text{MHz} = 1575.42\,\text{MHz}, \tag{6.1}$$

$$\text{L2} = 120 \times 10.23\,\text{MHz} = 1227.60\,\text{MHz}. \tag{6.2}$$

The carriers are Direct Sequence Spread Spectrum (DSSS) modulated by spreading codes, with each Space Vehicle (SV) using a different PRN sequence. This means that all SVs transmit at the same carrier frequencies in a CDMA fashion. In particular, every satellite has two unique spreading codes: the coarse-acquisition code (C/A), also known as civilian code, which is modulated only on L1, and the precision code (P(Y)), which is reserved for authorised civilian users and military use only and is modulated over both carrier L1 and L2. The polarisation adopted for all the GPS signals is the Right Hand Circular one.

The GPS frequency plan, with the occupied bandwidth and the carrier frequencies for the various GPS bands, is shown in Fig. 6.1.

Currently, civilian GPS receivers use the C/A code on the L1 frequency (1575.42 MHz), a Gold code of 1023 bits that repeats itself periodically every 1 ms (i.e. with a chipping rate of 1.023 Mchip/s). The code is modulo-2 added to the navigation data, that is sent at a bit rate of 50 bps (hence, 20 complete C/A codes are contained in each navigation bit). The code$\oplus$data, where $\oplus$ represents the modulo-2 addition, is BPSK modulated onto the carrier signal [1].

In the near future, the legacy civil signal will not be the only one broadcast by the GPS satellites. A crucial aspect of the system modernisation process is, indeed, the introduction of new navigation signals for civilian use. They are referred to as L2C, L5 and L1C.

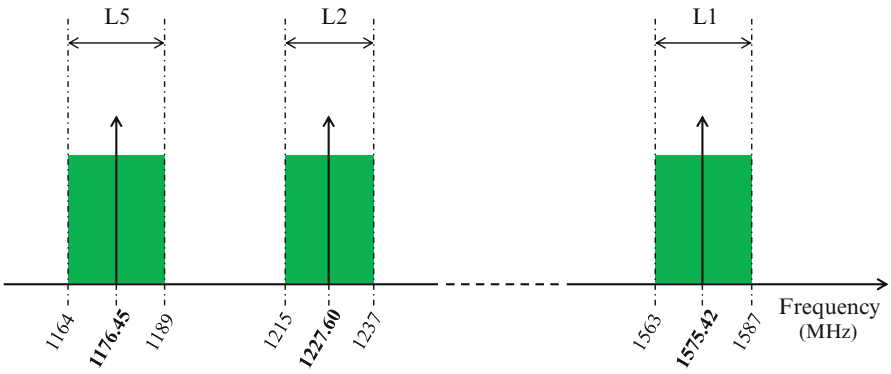

Fig. 6.1 GPS frequency plan

A second civilian signal is foreseen by the GPS modernisation program: L2C, designed specifically to meet commercial needs, is modulated onto the L2 frequency (1227.6 MHz) and uses two different PRN codes per satellite: CM (Civil Moderate) code, with a sequence length of 10,230 bits, a chipping rate of 511.5 kchip/s and a repetition period of 20 ms, and CL (Civil Long) code with a sequence length of 767,250 bits, a chipping rate of 511.5 kchip/s and a repetition period of 1.5 s. A 25-bps navigation signal is encoded into a 50-baud stream with a rate 1/2 convolutional coder, and the resulting data stream modulates the CM code. The L2C carrier signal is BPSK modulated by a chip-by-chip time multiplex combination of bit trains consisting of the L2 CM code with data and the L2 CL code with no data (alternating between L2 CM$\oplus$data and L2 CL chips at 1.023 MHz rate). The L2C civil signal has a similar power spectrum to the C/A code, i.e. 2.046 MHz null-to-null bandwidth. In both cases, as well as for any other GNSS signal, since the ranging code is periodic, it follows from the Fourier analysis that its frequency spectrum is discrete. However, since the C/A code is much shorter than both CM and CL codes, the L2C maximum spectral lines are far lower than the C/A maximum spectral lines. A smaller maximum in the spectrum represents an advantage in the presence of narrowband interference. In the worst-case scenario of interfering signal located at or near the maximum of the GNSS signal spectrum, a smaller maximum in the spectrum causes the signal to be less affected by the interference [2].

With the purpose of meeting the stringent requirements for safety-of-life and other high-performance applications, a third civilian GPS signal, called L5, has been designed. The L5 signal is contained within a 24 MHz band centred around the L5 nominal frequency.

$$\mathrm{L5} = 115 \times 10.23\,\mathrm{MHz} = 1176.45\,\mathrm{MHz} \tag{6.3}$$

Two carrier components in phase quadrature to each other, namely the in-phase signal component (I5) and the quadrature-phase signal component (Q5), are used for the L5 signal, and BPSK scheme is employed to modulate each carrier component

by separate bit trains. One bit train is the modulo-2 sum of the I5 code, the navigation data and a synchronisation sequence, whereas the other is built with the Q5 code and a different synchronisation sequence, without navigation data. A 50-bps navigation signal is used to modulate the I5 component. The overall symbol rate is, however, of 100 baud, since Forward Error Correction (FEC) with the same convolutional encoding as L2C is used. The PRN ranging codes are independent but time synchronised. They both have a length of 10,230 chips, a chipping rate of 10.23 MHz and a period of 1 ms. Quadrature Phase Shift Keying (QPSK) is used to combine the I5 and Q5 components.

Finally, a fourth civilian signal will be broadcast by the GPS satellites with the intent to facilitate interoperability between GPS and the other GNSS. This signal is called L1C, since it is modulated onto the L1 frequency (1575.42 MHz) and it is for civilian use. The L1C signal comprises two main parts: a pilot signal ($L1C_P$), which is not modulated by any data message and it is only spread by a ranging code, and a second component ($L1C_D$), which is spread by a ranging code and also modulated by a data message. $L1C_P$ and $L1C_D$ codes are independent and time synchronised. They are 10,230-chips long, and have a rate of 1.023 MHz and a period of 1 ms. The pilot component is also modulated by an SV unique overlay code, called $L1C_O$, which is 1800-bits long, and has a rate of 100 bps and a period of 18 s. The data-less component is given by the modulo-2 addition of $L1C_P$ code and $L1C_O$ code, and the Time-Multiplexed Binary Offset Carrier (TMBOC) modulation scheme is used to modulate the bitstream on L1 carrier. This technique in turn uses a mixture of BOC(1,1) spreading symbols and BOC(6,1) spreading symbols, where each BOC(6,1) spreading symbol consists of 6 cycles of a 6×1.023 MHz square wave, defined as binary 101010101010, with total duration 1/1.023 ms. The data component is given by the modulo-2 addition of $L1C_D$ code and the L1C message symbol train. It is modulated on L1 carrier using BOC(1,1) modulation, with a subcarrier frequency of 1.023 MHz and a chipping rate of 1.023 MHz.

L1C will be transmitted on the same frequency as the L1 C/A signal, and its design will improve mobile GPS reception in challenging environment.

In addition to the civil signals, a new military signal, the M-code, is planned to be transmitted in L1 and L2 frequencies [3]. However, the description of such signal is not addressed in this chapter.

A summary of the present and planned GPS signals and their main characteristics is given in Table 6.1.

6.2.1.3 Time and Geodetic Reference Frame

The GPS Control Segment is responsible for defining the GPS Time (GPST) based on the atomic clocks onboard the satellites and the ones of the monitor stations located throughout the world. The time scale started at 0 h UTC (midnight) of January 5th, 1980, and it differs from UTC because unlike it, which is periodically corrected with an integer number of leap seconds, GPST is a continuous time scale. The information about the offset between GPST and UTC is contained in

Table 6.1 GPS signals summary

Band	Carrier freq. (MHz)	PRN code	Modulation	Code rate (Mcps)	Data rate (bps)	Service
L1	1575.42	C/A	BPSK(1)	1.023	50	Civil
		P	BPSK(10)	10.23	50	Military
		M	$BOC_{sin}(10,5)$	5.115	N/A	Military
		L1C-I data	TMBOC(6,1,1/11)	1.023	50	Civil
		L1C-Q pilot		10.23	–	Military
L2	1227.6	P	BPSK(10)	10.23	50	Military
		L2C M	BPSK(1)	0.5115	25	Civil
		L2C L		0.5115	–	Civil
		M	$BOC_{sin}(10,5)$	1.023	N/A	Military
L5	1176.45	L5-I data	BPSK(10)	10.23	50	Civil
		L5-Q data	BPSK(10)	10.23	–	Civil

Table 6.2 Ellipsoidal parameters of WGS-84

Parameter	Symbol	Value
Semi-major axis of the ellipse	a	6,378,137 m
Flattening factor	f	1/298.257223563
Earth angular velocity	ω_E	$7292115 \cdot 10^{-11}$ rad/s
Gravitational constant	μ	$3986004.418 \cdot 10^8$ m^3/s^2

the navigation message, and the Control Segment is responsible for maintaining the offset error within 50 ns (with 95% probability).

The geodetic reference frame used by GPS is the World Geodetic System 1984 (WGS-84), developed by the US Department of Defence [4]. The most recent WGS-84 frame agrees with International Terrestrial Reference Frame (ITRF) at the centimetre level. The WGS-84 ellipsoid parameters are shown in Table 6.2.

6.2.2 GLONASS

Russia has also developed its satellite-based navigation system and its name is GLONASS. The system was formally declared operational in 1993 and brought to its optimal status of 24 operational satellites in 1995. After a period of difficulties in finding the funds for maintaining GLONASS, subsequent to the dissolution of the Soviet Union, the full constellation was restored in 2011. Differently from the other GNSS, GLONASS uses FDMA technique, which means that each satellite transmits on its own carrier frequency. A fewer number of carrier frequencies are actually needed, since two satellites in the same orbit that differ by 180° in latitude, the so-called antipodal satellites, can transmit on the same frequency. The reason behind this is that an operational receiver on the Earth's surface will never be able

to view both of them at the same time. Therefore, for the entire constellation of 24 satellites, 12 channels are enough.

6.2.2.1 Space Segment

The GLONASS constellation comprises 24 satellites distributed over three orbital planes, with eight satellites per plane. The circular orbits have an altitude of about 19,100 km and an inclination of 64.8°, and each SV has an orbital period of 11 h 15 min.

6.2.2.2 Current and Planned Signals

Two types of navigation signals are transmitted by each GLONASS satellite in L1 and L2 bands: high accuracy and standard accuracy signal.

The FDMA L1 carrier frequencies values are obtained as follows:

$$f_{k1} = 1602\,\text{MHz} + K \times 0.5625\,\text{MHz} \tag{6.4}$$

where K represents the frequency channel and is provided, for any particular satellite, in almanac. The used channels are $K = -7, \ldots, 6$. Therefore, the frequencies in L1 band used by GLONASS satellites are in the interval [1598.0625 MHz, 1605.375 MHz]. The standard and high accuracy services use a navigation signal with clock rates of 0.511 and 5.11 MHz, respectively. The modulo-2 sum of navigation message, ranging code and an auxiliary meander sequence is used to BPSK modulate the L1 carrier. The 511-bit long PRN ranging code has a period of 1 ms and a chipping rate of 0.511 MHz, and it is identical for all satellites. Navigation data is transmitted at a rate of 50 bps and the auxiliary sequence at a rate of 100 Hz.

The FDMA L2 carrier frequencies values are obtained as follows:

$$f_{k2} = 1246\,\text{MHz} + K \times 0.4375\,\text{MHz} \tag{6.5}$$

The used frequency channels are $K = -7, \ldots, 6$, so the frequencies in L2 band used by GLONASS satellites are in the interval [1242.9375 MHz, 1248.625 MHz]. Even in the L2 band, the adopted modulation scheme is BPSK, and the ranging code is the same as the one used in L1 band.

GLONASS is also evolving with plans to introduce additional CDMA signals in the third generation of its satellites, namely GLONASS-K satellites, the first of which was launched on February 26, 2011. The first step of this upgrade programme, which will enable greater compatibility with other GNSS, introduces a third signal in the L3 band centred at 1202.025 MHz. QPSK scheme is used to modulate the

Table 6.3 GLONASS signals summary

Band	Carrier freq. (MHz)	PRN code	Modulation	Code rate (Mcps)	Data rate (bps)	Service
L1	1602 + 0.5625k	C/A	BPSK(0.511)	0.511	50	Civil
		P	BPSK(5.11)	5.11	N/A	Military
L2	1246 + 0.4375k	C/A	BPSK(0.511)	0.511	50	Civil
		P	BPSK(5.11)	5.11	N/A	Military
L3	1202.025	L3-I	BPSK(4)	4.092	100	Civil
		L3-Q	BPSK(4)	4.092	–	Civil

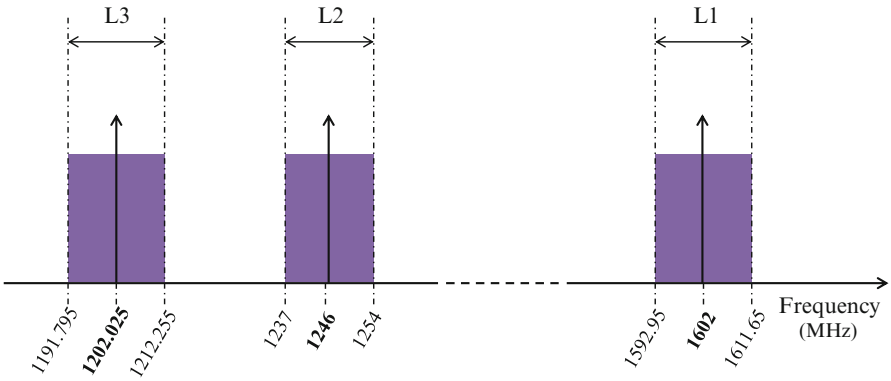

Fig. 6.2 GLONASS frequency plan

carrier, with an in-phase data channel and a quadrature data-less channel. Thirty one truncated Kasami sequences, which have good cross-correlation properties, constitute the set of ranging codes. These sequences are binary sequences of length $2^m - 1$ where m is an even integer. Their full length is $2^{14} - 1 = 16{,}383$ symbols, but the ranging code is truncated to a length of $N = 10{,}230$ chips with a period of 1 ms.

A summary of the current and planned open GLONASS signals and their main characteristics is given in Table 6.3, and the GLONASS frequency plan is shown in Fig. 6.2.

6.2.2.3 Time Scale and Geodetic Reference Frame

Just like GPS, GLONASS employs its own reference systems for time and coordinates. Whereas GPST represents a continuous time scale, GLONASS system time is strictly connected to Moscow time $\mathrm{UTC_{SU}}$, with periodic introduction of leap seconds.

Table 6.4 Ellipsoidal parameters of PZ-90 and PZ-90.11

Parameter	Symbol	Value
Semi-major axis of the ellipse	a	6,378,136 m
Flattening factor	f	$1/298.257839303$
Earth angular velocity	ω_E	$7292115 \cdot 10^{-11}$ rad/s
Gravitational constant	μ	$3986004.4 \cdot 10^8$ m^3/s^2

GLONASS coordinates (of satellites and user) are expressed in the Parametry Zemli 1990 (Parameters of the Earth 1990) (PZ-90) reference frame. Since December 2013, the new version of PZ-90, namely PZ-90.11, has been adopted. The parameters associated with the PZ-90 and PZ-90.11 are shown in Table 6.4.

6.2.3 Galileo

Galileo is the European-controlled satellite-based navigation system currently under development and expected to be completed by 2020. When Fully Operational Capability (FOC) will be reached, four frequency bands, namely E1, E6, E5a and E5b, will be used by every Galileo satellite. The different broadcast signals will support different services, based on various user needs as follows:

- Open Service (OS): it is free of charge and worldwide available. Signals transmitted on three different frequencies will provide the service, with similar performance to GPS L1C/A when using single-frequency receivers. In general, OS Galileo signals used in combination with GPS signals will enhance the performance in harsh environments, such as urban areas.
- Public Regulated Service (PRS): two encrypted signals will be used by the security authorities (police, military, etc.) under governmental control. Since high robustness and continuity will be required by this type of service, improved modulation/encryption schemes are introduced to enhance the capabilities to cope with jamming and spoofing. Two PRS navigation signals will be available.
- Commercial Service (CS): it will be provided through the introduction of two additional signals with higher data rates (up to 500 bps) and protected by commercial encryption.
- Galileo, in addition, will be part of MEOSAR, the Medium Earth Orbit Search and Rescue component of the International Cospas-Sarsat Programme. Equipped with Search and Rescue transponder, Galileo satellites will be able to capture signals coming from emergency beacons in the 406–406.1 MHz band and to send them back to dedicated ground stations called Medium Earth Orbit Local User Terminals (MEOLUT)s, which then localise the beacon, decode the distress message and send the information to Cospas-Sarsat Mission Control Centres.

Similarly to GPS, Galileo employs CDMA technique to differentiate the satellites, namely all of them share the same frequencies but use different ranging codes. These signals can contain data and pilot channels.

6.2.3.1 Space Segment

When completely operational, Galileo constellation will consist of 30 satellites (27 operational and 3 spares) in three orbital planes, with 10 satellites evenly distributed per plane. The orbits will have an altitude of 23,222 km, the orbital period of about 14 h and an inclination of 56° to the equator. This inclination will provide better coverage of polar latitudes compared to GPS system.

6.2.3.2 Current and Planned Signals

Galileo E1 will support OS, CS and PRS. Three navigation signal components will be broadcast by Galileo satellites in the L1 band, centred at 1575.42 MHz. E1-A is encrypted and only accessible to authorised PRS users. The other two components, E1-B and E1-C, respectively, a data channel and a pilot channel, are open access signals with unencrypted ranging codes accessible to all users. These two latter signals will use Multiplexed Binary Offset Carrier (MBOC) modulation implemented by the CBOC(6,1,6/11) scheme. The E1 signal is defined in baseband as:

$$s_{\mathrm{E1}}(t) = \frac{1}{\sqrt{2}}(s_{\mathrm{E1\text{-}B}}(t)(\alpha \cdot \mathrm{sc}_A(t) + \beta \cdot \mathrm{sc}_B(t)) - s_{\mathrm{E1\text{-}C}}(t)(\alpha \cdot \mathrm{sc}_A(t) - \beta \cdot \mathrm{sc}_B(t))) \tag{6.6}$$

where the subcarriers $\mathrm{sc}_A(t)$ and $\mathrm{sc}_B(t)$ are defined as

$$\begin{aligned} \mathrm{sc}_A(t) &= \mathrm{sgn}(\sin(2f_{s,\mathrm{E1A}}t)), \\ \mathrm{sc}_B(t) &= \mathrm{sgn}(\sin(2f_{s,\mathrm{E1B}}t)) \end{aligned} \tag{6.7}$$

with $f_{s,\mathrm{E1A}} = 1.023\,\mathrm{MHz}$, $f_{s,\mathrm{E1B}} = 6.138\,\mathrm{MHz}$ the subcarriers rates, $\alpha = \sqrt{\frac{10}{11}}$ and $\beta = \sqrt{\frac{1}{11}}$.

The E1-B component is the modulo-2 addition of the Integrity Navigation Message, transmitted at 250 sps, and a PRN code, $\mathrm{C_{E1B}}$. The E1-C component is a pilot signal, generated by using a secondary code sequence, $\mathrm{C_{E1Cp}}$, to modify successive repetitions of a primary code ($\mathrm{C_{E1Cs}}$) periods. The $\mathrm{C_{E1B}}$ and $\mathrm{C_{E1Cp}}$ primary codes are pseudorandom memory code sequences of 4092 chips at a rate of 1.023 MHz and they are defined in [5]. The binary sequence of the secondary code $\mathrm{C_{E1Cs}}$ is 001110000000101011011 0010.

The E5 signal comprises four components: E5a-I, that consists of the navigation data stream $D_{E5a\text{-}I}$ modulated with the unencrypted ranging code $C_{E5a\text{-}I}$, E5a-Q (pilot component), from the unencrypted ranging code $C_{E5a\text{-}Q}$, E5b-I, that consists of the navigation data stream DE5b-I modulated with the unencrypted ranging code $C_{E5b\text{-}I}$, and E5b-Q (pilot component), from the unencrypted ranging code $C_{E5b\text{-}Q}$. E5a and E5b signals together form the E5 signal in its full bandwidth. The carrier frequencies are 1176.45 MHz, 1207.140 MHz and 1191.795 MHz for E5a, E5b and E5, respectively. Data channels $D_{E5a\text{-}I}$ and $D_{E5b\text{-}I}$ will transmit at a rate of 50 sps and 250 sps, respectively. A modified version of a BOC with code rate of 10.23 MHz and a subcarrier frequency of 15.345 MHz, called Alternate Binary Offset Carrier (AltBOC), is used in Galileo E5, a wideband signal transmitted at 1191.795 MHz. Its two components, E5a and E5b, are separately QPSK signals with a carrier frequency of $f_{s,\mathrm{E5a}} = 1176.45$ MHz and $f_{s,\mathrm{E5b}} = 1207.14$ MHz, respectively, and can, therefore, be processed independently by the user receiver. The E5 baseband AltBOC signal is defined as:

$$\begin{aligned} s_{\mathrm{E5}}(t) = {} & \text{E5b-}I(t)\mathrm{sgn}(e^{j2\pi f_{s,\mathrm{E5b}}t}) \\ & + \text{E5a|-}I(t)\mathrm{sgn}(e^{-j2\pi f_{s,\mathrm{E5a}}t}) \\ & + \text{E5b-}Q(t)\mathrm{sgn}(e^{j(2\pi f_{s,\mathrm{E5b}}t+\pi/2)}) \\ & + \text{E5a-}Q(t)\mathrm{sgn}(e^{j2\pi f_{s,\mathrm{E5a}}t-\pi/2)}) \end{aligned} \tag{6.8}$$

E6 signal will support the CS and PRS services. Three navigation signal components will be broadcast in the E6 band, centred at 1278.75 MHz, by the Galileo satellites. Similarly to E1, one component, namely E6-A, is encrypted and only accessible to authorised PRS users. The other two, E6-B and E6-C, respectively, a data channel and a pilot channel, are commercial access signals. The E6 signal is generated according to:

$$s_{\mathrm{E6}}(t) = \frac{1}{\sqrt{2}}[e_{\text{E6-B}(t)} - e_{\text{E6-C}(t)}] \tag{6.9}$$

The E6-B component is the modulo-2 addition of the ranging code C_{E6B} with the navigation data stream, and the pilot component E6-C is generated by a tiered code construction, without data. The E6 ranging codes are not disclosed in [5] and their rate is 5.115 MHz.

A summary of the planned Galileo signals and their main characteristics is provided in Table 6.5, and the Galileo frequency plan is shown in Fig. 6.3.

6.2.3.3 Time Scale and Geodetic Reference Frame

Galileo System Time (GST) is a continuous time scale under responsibility of the Galileo Mission Segment (GMS) and its start epoch is 0 h on August 22nd, 1999.

Table 6.5 Galileo signals summary

Band	Carrier freq. (MHz)	PRN code	Modulation	Code rate (Mcps)	Data rate (sps)	Service
E1	1575.42	E1-A data	BOCcos(15,2.5)	2.5575	N/A	PRS
		E1-B data	CBOC(+)(6,1,1/11)	1.023	250	OS, CS
		E1-C pilot	CBOC(−)(6,1,1/11)	1.023	–	OS,CS
E6	1278.75	E6-A data	BOCcos(10,5)	5.115	N/A	PRS
		E6-B data	BPSK(5)	5.115	1000	CS
		E6-C pilot	BPSK(5)	5.115	–	CS
E5a	1176.45	E5a-I data	AltBOC(15,10)	10.23	50	OS
		E5a-Q pilot		10.23	–	OS
E5b	1207.14	E5b-I data		10.23	250	OS, CS
		E5b-Q pilot		10.23	–	OS,CS

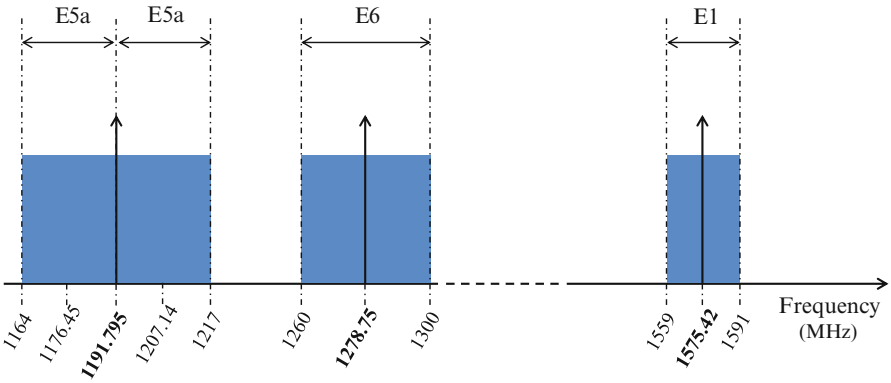

Fig. 6.3 Galileo frequency plan

Both the Galileo to GPS Time Offset (GGTO) and the GST-UTC offset will be broadcast by the Galileo satellites [5].

Like GPS and GLONASS, Galileo will use a dedicated terrestrial reference frame, GTRF, which is an independent realisation of the ITRF. In practice, GPS and Galileo frames will be almost the same, with only a few centimetres difference.

6.2.4 BeiDou

BeiDou is the Chinese GNSS, currently under development and expected to be completed by 2020. As in GPS and Galileo, the CDMA principle is the one chosen for differentiating the satellites, and three sub-bands of the L-band, referred to as B1, B2 and B3, will be used. BeiDou satellites will transmit Right Hand Circularly Polarized (RHCP) signals, and in line with the other GNSS, there will be two levels of positioning service: open and restricted. The open service signals are denoted as B1I and B2I. B1Q, B2Q and B3 will, instead, be used for authorised services.

6.2.4.1 Space Segment

Once completely deployed, BeiDou constellation will consist of 35 satellites: 5 Geostationary Earth Orbit (GEO) satellites operating at an altitude of 35,786 km and positioned at 58.75°E, 80°E, 110°E, 140°E and 160°E, respectively, 3 Inclined Geosynchronous Orbit (IGSO) satellites operating in orbit at an altitude of 35,786 km and an inclination of 55° to the equatorial plane, and 27 MEO operating in orbit at an altitude of 21,528 km and an inclination of 55° to the equatorial plane.

6.2.4.2 Current and Planned Signals

This section focuses on the open service signals only, for which the Interface Control Document (ICD) was released in 2012. The nominal frequency of B1I signal is 1561.098 MHz. This carrier is BPSK modulated by the modulo-2 addition of ranging code (CB1I) and navigation message (D1). The B1I ranging code is a balanced Gold code truncated with the last one chip. Its length is 2046 chips and the chipping rate is 2.046 MHz. The 50 bps D1 navigation message contains basic navigation information (time offsets from other GNSS, almanac information for all the satellites and fundamental NAV information of the broadcasting satellites) and is broadcast by MEO/IGSO satellites.

The Phase II B1 open service signal is a QPSK signal centred at 1561.098 MHz and with 4.092 MHz bandwidth, whereas in the BeiDou Phase III, it is planned to shift the signal to 1575.42 MHz and to use an MBOC(6,1,1/11) modulation similar to the Galileo L1 OS signal and the modernised GPS civil signal (L1C).

The nominal frequency of B2I signal is 1207.140 MHz. This carrier is BPSK modulated by the modulo-2 addition of ranging code (CB2I) and navigation message (D2). For each satellite, the B2I ranging code is the same as B1I ranging code. D2 navigation message contains basic navigation and augmentation service information, it is broadcast by GEO satellites and has a rate of 500 bps.

A summary of the BeiDou Phase II open service signals and their main characteristics is given in Table 6.6, and the corresponding frequency plan is shown in Fig. 6.4.

Table 6.6 BeiDou Phase II signals summary

Band	Carrier freq. (MHz)	PRN code	Modulation	Code rate (Mcps)	Data rate (bps)	Service
B1	1561.098	B1-I	BPSK(2)	2.046	500 (GEO) 50 (MEO/ISGO)	Open
B2	1207.14	B2-I	BPSK(2)	2.046	500 (GEO) 50 (MEO/ISGO)	Open

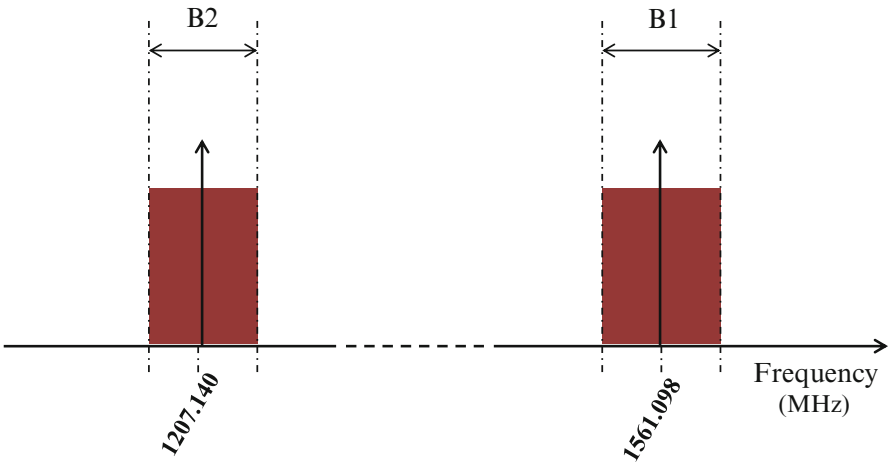

Fig. 6.4 BeiDou frequency plan

Table 6.7 Ellipsoidal parameters of CGCS2000

Parameter	Symbol	Value
Semi-major axis of the ellipse	a	6,378,137 m
Flattening factor	f	1/298.257222101
Earth angular velocity	ω_E	$7292115 \cdot 10^{-11}$ rad/s
Gravitational constant	μ	$3986004.418 \cdot 10^8$ m^3/s^2

6.2.4.3 Time Scale and Geodetic Reference Frame

BeiDou uses a continuous time scale, BeiDou Time (BDT), which started at 0 h UTC on January 1st, 2006, and is synchronised with UTC. The navigation message contains the offset with respect to UTC, which is controlled to be within 100 ns (modulo 1 s).

The coordinate system adopted by BeiDou is the China Geodetic Coordinate System 2000 (CGCS2000) with a reference ellipsoid defined by the parameters shown in Table 6.7.

6.3 Multi-GNSS Benefits

The advantages of a multi-GNSS solution with respect to a single-system receiver are described in this section. Receiving more signals from different systems will offer significant benefits in terms of reliability, continuity, accuracy and availability:

- **Accuracy improvement**: more received signals means that more measurements will be available at the receiver for obtaining the PVTsolution. Results showing

that a multi-GNSS solution outperforms single-system ones in terms of positioning accuracy can be found in [6, 7]. Moreover, the effects of multipath and interference can be more easily mitigated. By taking, for example, into account the diversity provided by various signals and by implementing a multi-GNSS signal selection mechanism, the multi-GNSS performance in the presence of intentional interference would be significantly improved with respect to the single-system solutions [8]. Another consequence of the joint utilisation of four systems will be the Geometric Dilution of Precision (GDOP) improvement, as proved in [9]. This higher geometrical strength will further enhance accuracy.

- **Reliability increase**: multi-frequency multi-system reception will provide an additional protection against intentional and unintentional interference. Even when one signal is severely affected by the interferer, it will still be possible to obtain a good PVTsolution by selecting and relying on the other, less affected, signals. Furthermore, integrating multiple carriers provides more redundancy in the measurements, and this fact can be exploited for ionospheric and multipath errors estimation [10]. Spoofing detection would also benefit from the existence of several systems, as it requires more resources to spoof all the systems at the same time.
- **Continuity improvement**: the fact that GPS, GLONASS, Galileo and BeiDou are independent systems means that the major system problems are a very remote possibility of occurring simultaneously. Provided that the time scales offsets among the systems are known or calibrated, at least four satellites are required to maintain a valid position fix. The chances of this being the case are higher with a receiver looking for satellites from more than one constellation.
- **Availability improvement**: one of the main impacts of multi-constellation in PVTsolutions is the increased number of available satellites [9]. This is particularly important for navigation in harsh environments, as urban canyons, where a clear path to the sky can be difficult to obtain. Obstacles such as high buildings can obscure satellite signals, but, with much more satellites in the sky, the success rate of positioning in such challenging scenarios increases.
- **Integrity monitoring**: the increase in available satellites brings major benefits also to integrity monitoring. With more available satellites, the system can more easily understand if some of the satellites are misbehaving, when considerable differences appear in satellite to satellite range computations [2, 11, 12].

6.4 Multi-GNSS Issues

If, on the one hand, the diversity of multiple systems will bring several advantages, on the other hand, when using multi-constellation solutions, several problems arise as well.

The modernisation of the two existing systems and the development of the other two GNSS include the transmission of new and enhanced signals, with more complex characteristics and specifications. New frequency, modulation schemes and navigation message structures will be, for example, used. Consequently, the receiver

has to be able to handle the new and upgraded GNSS signals. The analog parts of the receiver must operate with multiple frequency bands, multiple systems and larger signal bandwidths, and, in the position computation block, new algorithms must be implemented. These requirements undoubtedly lead to a higher complexity and therefore a higher cost of the receiver. Regarding baseband processing, innovative methods have to be developed to properly exploit the enhancements derived by the use of the new advanced signals. This increases the computational complexity, the resource capability demand and eventually the cost of the receiver. For these reasons, the software receiver approach, where the hardware requirements are minimised, implemented on programmable processors, whose computational capability keeps increasing rapidly, represents a suitable solution for multi-GNSS receivers.

Besides more complex hardware to deal with the multitude of signals, for the computation of a PVTsolution, challenges appear also due to different time scales and reference frames. Despite all systems having a reference to UTC time, the time keeping in the ground control is independent for each system, leading to time differences with respect to UTC time and among systems. Therefore, the time offset between each system needs to be taken into account and estimated by the user or provided through the ground control [13]. BeiDou has been designed to allow the users to easily compute the time bias to any other GNSS system [14]. Regardless of that, other systems will also broadcast parameters to ease the determination of time bias among multi-constellation users.

Reference frames also play an important role in the position computation, since each system uses a particular realisation of the Earth's geoid. Regarding GPS and Galileo, the coordinate frames WGS-84 and ITRF, respectively, have the same realisation up to a centimetre level. For these systems, the recommendation is to assume that the coordinates are the same. GLONASS, on the contrary, has considerable deviations that must be taken into account. Currently, the transformation between PZ-90 and WGS-84 is available through official sources [15]. Recommendations and advances in these topics are available through the International Committee on GNSS [16].

Another significant consequence of a multi-GNSS environment will be the inter-system interference. The four constellations will transmit signals in shared frequency bands. From the receiver design point of view, this aspect could be advantageous, but, on the other hand, interference from signals broadcast by other GNSS increases. Results on the inter-system interference levels to each GNSS in the presence of all the four systems can be found in [9].

6.5 Conclusions

This chapter presented the four systems that, in the near future, will make a reality of multi-GNSS positioning. Two of them, namely GPS and GLONASS, are already in fully operation, whereas BeiDou and Galileo are still under development and expected to be completed by 2020.

Therefore, more than a hundred satellites transmitting different signals in different frequency bands will be soon available. Such diversity and redundancy will bring advantages in terms of availability, accuracy, continuity and reliability of the service. However, several issues also arise from having different systems: more GNSS sharing the same frequency bands will increase the inter-system interference, and more unknowns will need to be solved due to the different time and coordinates frames adopted. Collaboration is then required between the agencies controlling each GNSS not only to mitigate possible interference issues, but also to enhance each other systems and provide better service to the end users. International cooperation to assure compatibility and achieve interoperability is essential.

On the hardware side, designing receivers for several GNSS is a challenge. Having multiple systems means an increase in the number of tracked satellites, handling different frequencies, time bias, codes, among others. The increase of hardware components means an increase of the errors introduced by the receiver hardware. While this can be accounted for, development of software receivers is also seen as an approach to deal with multi-system receivers.

Acknowledgements This work was financially supported by EU FP7 Marie Curie Initial Training Network MULTI-POS (Multi-technology Positioning Professionals) under grant nr. 316528.

References

1. Navstar Global Positioning System Interface Specification is-GPS-200, Revision D. Technical report, GPS Joint Program Office (2006)
2. E.D. Kaplan et al., *Understanding GPS - Principles and Applications*, 2nd edn. (Artech House, Boston, 2006)
3. Navipedia, *GPS Future and Evolutions*. Accessed 19 Dec 2016. http://www.navipedia.net/index.php/GPS_Future_and_Evolutions
4. W.M. Mularie, World geodetic system 1984–its definition and relationships with local geodetic systems. Technical report (2000)
5. European Union 2015, Galileo open service signal in space interface control document. Technical report, 1.2. Accessed 24 May 2016
6. D. Minh Truong, T. Hai Ta, Development of real multi-GNSS positioning solutions and performance analyses, in *2013 International Conference on Advanced Technologies for Communications (ATC)* (IEEE, Ho Chi Minh City, 2013), pp. 158–163
7. S. Soderholm et al., A multi-GNSS software-defined receiver: design, implementation, and performance benefits. Ann. Telecommun. **71**(7), 399–410 (2016)
8. H. Bhuiyan et al., Performance analysis of a multi-GNSS receiver in the presence of a commercial jammer, in *2015 International Association of Institutes of Navigation World Congress (IAIN)* (IEEE, Prague, 2015), pp. 1–6
9. N.G. Ferrara, E.S. Lohan, J. Nurmi, Multi-GNSS analysis based on full constellations simulated data, in *2016 International Conference on Localization and GNSS (ICL-GNSS)* (Barcelona, 2016)
10. M. Sahmoudi, R. Landry Jr., F. Gagnon, Robust mitigation of multipath and ionospheric delays in multi-GNSS real-time kinematic (RTK) receivers, in *IEEE/SP 15th Workshop on Statistical Signal Processing, SSP'09* (IEEE, Cardiff, 2009), pp. 149–152

11. K. Borre et al., *A Software-Defined GPS and Galileo Receiver: A Single-Frequency Approach* (Springer, Dordrecht, 2007)
12. GMV, Navipedia, *Integrity*. Accessed 19 Dec 2016. http://www.navipedia.net/index.php/Integrity
13. A. Druzhin, A. Tyulyakov, A. Pokhaznikov, Broadcasting system time scales offsets in navigation messages. Assessment of feasibility. Technical report (2013)
14. China Satellite Navigation Office, BeiDou navigation satellite system signal in space interface control document - open service signal. Technical report, 2.0 (2013)
15. Russian Institute of Space Device Engineering, GLONASS interface control document. Technical report, 5 Jan 2008
16. Working Group D International Committee on GNSS, Report of working group: reference frames, timing and applications. Technical report (2010)

Chapter 7
Towards Seamless Navigation

Pekka Peltola and Terry Moore

7.1 Introduction and Motivation

Navigation in urban environment and indoor spaces is yet a challenge. A GNSS receiver cannot cope with the heavy multipath and attenuation of the GNSS signals within these areas. Research on alternative technologies like WiFi or Bluetooth may produce a more efficient solution for successful indoor navigation.

This chapter provides a short review of the available sensors and methods that have been tested to fill this gap. It is divided into four sections. The first section discusses the adaptability features that are necessary to be adopted in order to have a seamless navigation system. Seamlessness is a term used in this chapter for a navigation system that is able to give a solution regardless of the environment. The second section deals with different location contexts and behavioural contexts. At different locations the navigation filter setup will be different. When walking, running or taking a bus the behavioural context changes. The third section looks through the sensors that are available and that have been previously tested for localisation. The final fourth section uses the results from the previous sections and discusses the common methods in combining the sensor data into a final navigation solution. Multi-sensor approach, if the design is good, benefits of the complementary properties of different sensors and can be more useful than using a single technology solution.

P. Peltola (✉) • T. Moore
The Nottingham Geospatial Institute, The University of Nottingham, Triumph Rd, Nottingham NG7 2TU, UK
e-mail: pekka.peltola@nottingham.ac.uk; terry.moore@nottingham.ac.uk

J. Nurmi et al. (eds.), *Multi-Technology Positioning*,
DOI 10.1007/978-3-319-50427-8_7

The general aims for indoor positioning are stated in [24] by Rainer Mautz. The goal is to find matches between quantified user requirements and assessed sensor technologies and fusion methods. In addition, four important aspects should be achieved in order to be able to call the system seamless.

1. sub-meter level horizontal positioning accuracy
2. floor identification
3. >99% availability
4. minimal installation costs

Continuing in [9], Paul Groves mentions four challenges within the multi-sensor navigation systems. These are shortly interpreted here. Complexity should be kept at the optimal level. Context detection enables optimal navigation filter settings. Ambiguous solutions should be tackled correctly. Data storage, retrieval and handling is often application specific and there is often a compromise between positioning accuracy and computation time or system cost.

7.2 Adaptability

So far, adaptability in navigation has solely been the user's responsibility. Simple example of this is the use of GPS while driving and then checking the billboard map within the commercial centre for finding the wanted shop. In order to increase the level of seamlessness, the mobile phone that was used for car navigation should be able to offer the same information as the billboard map. Even further increase could be possible if the application on the phone would be able to assist in navigating to the correct shop while walking.

Two kinds of adaptive device behaviour can be distinguished in the previous example. The device that is capable of navigating the user to the shop has the configurational flexibility to do this. In addition, it has the environmental adaptability to adapt to car navigation as well as to indoor navigation.

7.2.1 Configurational Flexibility

Configuration for navigation purposes consists of the infrastructure, mobile device sensors, mobile device hardware and of the navigation software. Infrastructure and sensors define the technology choices that the navigation solution derivation can use. This affects strongly the availability of the navigation solution. That is if only GNSS is used and available there will not be a solution for indoor environment. Hardware aspects, such as the type and capability of the battery and processor, define the power configuration which also limits the software methods that can be used. Software methods that will be implemented will have an effect on the efficiency and accuracy of the final navigation solution. Figure 7.1 depicts this division and lists the features that the corresponding level affects.

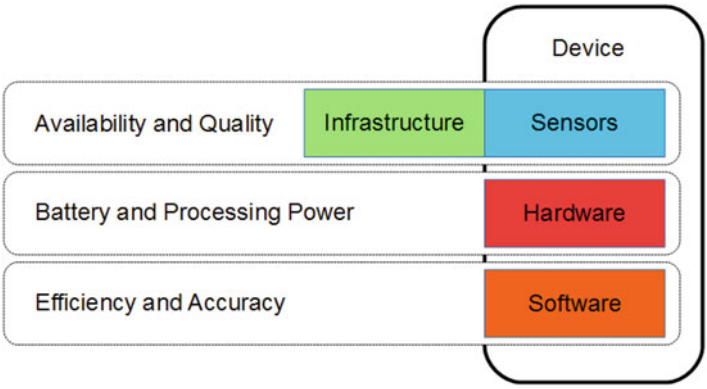

Fig. 7.1 Configuration can be adjusted at the different levels. Features affected are highlighted on the *left side*

The diversity of sensors ranges from temperature to acceleration and deformation sensors. Plug-and-play sensor modules all using a well-defined unified serial interface would ease hardware designer's workload. Smart sensors could be set up and used in altering ways depending on the final application. To save space the sensor hardware could be implemented on a reconfigurable Field Programmable Gate Array (FPGA). Then the sensor hardware could be able to reprogram its FPGA accordingly.

7.2.2 Environmental Adaptability

For a truly seamless navigation system the changing conditions in the environment should not affect the availability of the position solution. Seamless system has an adequate number of sensors, the processing hardware and software for deriving the correct good quality position solution. This adequate sensor configuration is able to offer enough positioning information for the navigation application. The application is also smart enough in selecting the best quality measurements to use, independent of the environment. In other words the system adapts to the surrounding environment.

The application uses the available hardware resources and possible parallel application information (e.g. calendar). Three resources are listed in [18]. Sensor measurements are the first obvious source of positioning information. Second, inference based positioning information from other applications and position history help in deriving the position. Third, interaction with the user can be taken along to the solution derivation. The combination of information happens in the Context Engine as shown in Fig. 7.2.

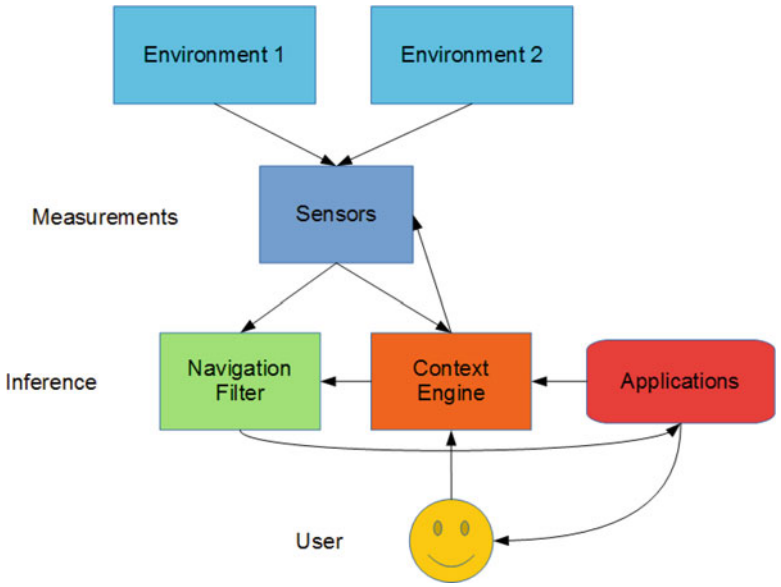

Fig. 7.2 Context Engine fuses the information available from the user, parallel applications and sensors. Context Engine sets the navigation filter settings in order to have the best possible position solution

7.3 Context

Context can be determined using physical sensors, virtual sensors (applications) and logical sensors (inferred from physical and virtual) [34]. Contextual information changes by locations. Context also differs according to different behavioural features. As well, user interaction is involved in the determination of the context as stated in the previous section. Moreover, navigation target and route preferences define the context in part. These will be examined in the following subsections.

Mobile applications use the context detection information. Context inference algorithms try to figure out the user context. Either for commercial use or user's personal chores, in any case, defining the correct context offers enhanced information for predicting user patterns and adapting to these possible upcoming needs. The navigation filter can use this information to use the most fit sensors and corresponding data to offer the best possible positioning solution.

7.3.1 Location

In the study in [34], five classification methods were tested for location detection. Of the Naive Bayes, Decision Tree, Bagged Tree, Neural Network and K-Nearest

Neighbour methods the Bagged Tree method gave the best results. In the Bagged Tree method multiple decision trees are trained using different subsets of the same larger dataset. Each tree is assigned with a weight. Using a voting system, more weight is given to a tree in the region where classification rate is better. The locations used in the study were in the abstraction level of home, work and so on.

Perhaps the first and the most relevant detection is between the indoor and outdoor environments. In [5], a method was developed for simple indoor/outdoor detection. An indicator derived from GNSS receiver properties managed to differentiate whether a person was walking outside or inside.

Bluetooth low energy (BLE) tags send an advertisement message to the environment at fixed intervals. If GNSS receiver indicator is not available, these alternative proximity sensors may be used to indicate the current location context. Camera based image matching using online image database is another alternative.

Map is the most obvious presentation of location. Most of the navigation applications need a preloaded or online map database in order to present the location to the user. For the blind or the deaf touch and audio based navigation databases are the better choice. In both cases, the location context is very much defined using the map, depending on how successful the acquired positioning solution has been.

Map updates are relevant. Especially if the sensor infrastructure, like in the case of BLE anchor deployment, is changing, the changes need to be informed for the user's map at latest when arriving to the proximity of the system. GNSS signals contain information for deriving satellite positions within the navigation message structure. Similar approach for the BLE devices could ease the map update process. BLE anchors capable of sending their locations would free the map database of this responsibility.

7.3.2 Behaviour

Indoor spaces consist of stairs, hallways, corridors, offices, rooms, lifts and so on. User movement in these differing spaces changes. Restricted spaces usually require the user to walk. In open areas, especially if the calendar alerts for a missed meeting, the user might make a dash and run for a certain amount of time. In taxis and trains the vehicle vibration can be sensed.

Especially the inertial sensor can detect changes in the behavioural context. Khalifa et al. [16] examined this and classified the pedestrian motion into activities relative to the location. Walk, run and climbing stairs are examples of the changing behavioural context. This behavioural activity information is related to the vehicle type as well [9]. Vehicles have their individual features that can be sensed. For example, if a smart shoe detects the round motion of the foot while cycling, this is a clear indication of the user's vehicle. This is depicted in Fig. 7.3. Inertial trace for climbing the stairs and when cycling have distinctive characteristics. The Context Engine can detect these differences and give instruction to, for example, the navigation filter or another application.

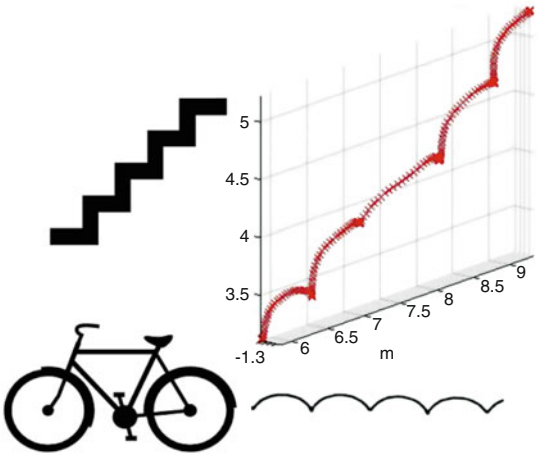

Fig. 7.3 Foot-mounted inertial measurement unit mechanisation results when climbing stairs and cycling

The behaviour detection can be implemented at different levels of abstraction. For example, if the user is wearing smart boots which can sense inertial movement, the sensor can process the measured data in two ways. Either it can forward the raw data or alternatively process it locally and forward formatted, high abstraction level sensor data. This saves the processes on the mobile device side for something more relevant, like the navigation solution derivation. Interface design between the sensor and the mobile device has then an impact on the overall performance of the system.

7.4 Sensors

A sensor measures physical phenomena. A virtual sensor measures virtual events, like calendar events or user interaction. The data that a sensor provides ranges from raw accelerometer values to a scheduled meeting with a friend. The abstraction level of the measurement information can be thought in four levels [35]. Figure 7.4 depicts this vertical abstraction division into raw, feature, pattern and decision levels. The physical and virtual sensors can be located into the table by their distinctive features. Adequate horizontal heterogeneity is necessary for the system to be able to function in different environments. Horizontal abstraction spectrum consists of a set of physical phenomena that the sensors can sense. Different sensor is suited for different phenomena.

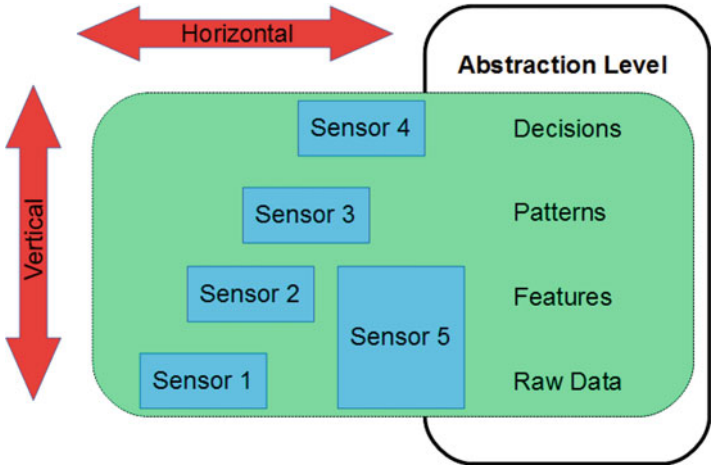

Fig. 7.4 Horizontal and vertical information abstraction applied here to sensor output information [35]

7.4.1 Maps and Infrastructure

Maps represent location relationships between places and infrastructure. WiFi and Bluetooth anchor location information being one example. Maps are located at the abstraction level that is understandable by most people. Apart from architectural CAD drawings and construction designs there are not yet many widespread detailed floor plan standards for navigation purposes. Open Geospatial Consortium initiative aims to provide a common schematic framework for wider use in indoor navigation applications [28]. Other existing related standards and examples include CityGML, IFC, X3D, ESRI BISDM, Google, Bing and Here maps.

The information coming from different sensors is combined with map data. The pedestrian dynamics model and moving patterns are different from the way the car sensor data is combined. Since dimensions indoors are smaller the requirement for positioning accuracy is higher indoors. The walls, streetways and park or other open zones function as constraints for possible user locations. These constraints help the application to snap the location solution onto sensible map locations. This method of using constraints on maps is called map matching. The general procedure is shown in Fig. 7.5 where map database is combined in the last stage when the positioning results are derived using the Bluetooth Low Energy (BLE), Ultra-Wideband (UWB) and Inertial Measurement Unit (IMU) Sensor Positioning technologies.

Building corridors are usually symmetrical consisting of only a few main directions. Pinchin et al. [29] used these cardinal headings of the building for aligning the user track towards these main directions.

The GNSS signals are blocked by tall buildings. The building height information can be exploited in the computation of signal blockage. Shadow matching is a

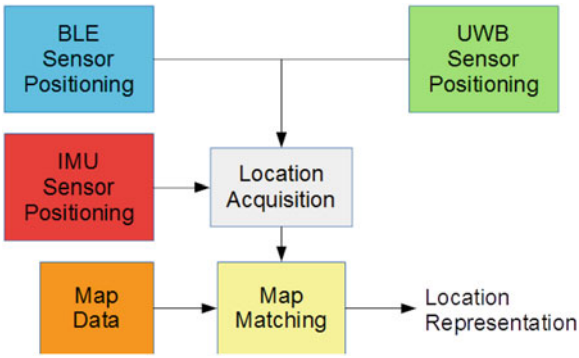

Fig. 7.5 Map matching infers the final presentable solution combining the relationships between infrastructure and sensor input [15]

method to overcome the troubles with GNSS signal multipath. The processing power needed for using 3D maps with many details is high. Alternative virtual modeling and more effective representation techniques are under research [8].

The access to floormap data of certain buildings is restricted. Thus crowdsourcing the floormap information would concentrate on the available public areas. Closed-circuit television (CCTV) and surveillance cameras could be used for positioning purposes, although privacy issues will have to be agreed.

7.4.2 Odometers and Tactile Sensors

Tactile localisation is based on a sensor or probe measuring or contacting a surface or an obstruction directly. Odometer sensor contact is continuous. Physical contact sensing or a measure of a direct physical distance or speed describes the tactile and odometer sensor outputs. Simple tactile sensor outputs binary data, either on or off. Force and odometer sensors output either transduced force measures or velocity information of the user.

A small puppy dog robot, in [32], uses pressure sensors and joint movement sensors for gait cycle recording. The odometry models for cars and four-legged robots are obviously different. The study in [14] examines the car heading filter measurement model which changes according to car movement characteristics. The odometer measures the car speed and this is used with gyroscope angle information for the dead reckoning of the system when GPS is not available. This acts as a small context aware filtering selection system offering movement model decision. Using multiple odometers and differential analysis on the odometer position in the body and on the circumference of the wheels the heading and distance traveled can be tracked more effectively [4].

Coordinate measuring machines are very accurate and expensive devices. They offer an exact way to position a product on a work bench but are not applicable for pedestrian or car navigation schemes. Piezoelectric or capacitive touch surfaces, levers or buttons can be used in recognising activity at certain locations. Load cells

on the floor may be used to sense the position of a user within a building. As a heal strike detection sensor an LDT0 piezoelectric PVDF sensor was used in [2].

Sensory overload is discussed in [3] on the user's perspective. When driving in the car, audible information coming from radio, environment, navigator and passengers might sometimes be too much. The navigation is thus performed through the use of a vibrating belt that tells the driver in which direction to turn next. The overall information channel content given to the driver in this method is less in comparison with audio-only guiding.

7.4.3 Sound and Pressure

Sound speed in air is approximately 346 m/s. This is much less than the RF signal speed. In order to travel from source to target the sound needs to travel through a medium like air. The properties of the medium need to be taken into account when designing a localisation system based on sound waves. These properties are, for example, temperature and pressure.

Sound positioning system construction consists of sound sources and recorders. Depending on the approach the anchors located in the environment can be either. Time of Arrival (ToA) or Time Difference of Arrival (TDoA) are two possible methods for sound system positioning architecture. Although noise is troublesome in sound based positioning, if possible line of sight detection methods help in the correct timing measurement derivation.

The two widely known sound based positioning systems are DOLPHIN and CRICKET [12]. DOLPHIN ultrasound system consists of transducers that consist of both transmitting and receiving ultrasonic sensors. This system can be set up in two ways. Either the mobile transducer transmits and the anchors receive and return time-of-flight information to the mobile unit or anchors transmit and the mobile unit receives and calculates its position. CRICKET system has the mobile tag as receiver and in addition an RF pulsing scheme is used for keeping the system synchronised. Chirp signals on mobile phones with signal amplitude envelope detection [13], barely audible environmental sound sources [33], multiple microphone setup and distance estimation between feet [19] are few examples on how sound has been utilised for positioning. Figure 7.6 shows a commercial off the shelf ultrasound sensor.

Fig. 7.6 Ultrasound sensor

In comparison to camera systems the audio data throughput and amount is much smaller. The audio signal composition and strength should be designed regarding all the external sound sources, noise and properties of the target environment.

Atmospheric pressure sensors can be used to detect the floor in the building. Pressure decreases as altitude increases. In [1], a real-time calibration method is used with a barometer to detect the current floor where the user is. The hardware system that was built takes into account the three most relevant parameters, temperature, reference pressure and the measured pressure in solving the correct floor.

7.4.4 Inertial Sensors

Inertial sensors measure the movement of the sensor, with a typical refresh rate of 100 Hz. If the sensor is strapped down onto an object like a boot, the movement of the boot in this case can be followed. How well it can be tracked depends on the accuracy of the IMU. Very accurate IMUs can cost tens of thousands of pounds. These are found in airplanes or in military vehicles. On mobile phones the cost of an IMU must be kept very low so correspondingly the accuracy of these mobile device sensors is poor. Inertial sensors include gyroscopes for attitude change sensing and accelerometers for movement tracking.

Inertial tracking is a dead reckoning tracking scheme. This means that in the beginning the guess of the location is best and gets worse over time, unless supporting measurements are introduced from external, usually non-dead reckoning systems. Harle did a literature review on inertial pedestrian positioning systems in [10]. The main topics and techniques are shown in Fig. 7.7.

The cold atom interferometry and chip-scale atomic clocks are still under research [8]. Before more accurate inertial measurement is possible the pedestrian navigation systems will have to rely on help from external absolute positioning technologies like GNSS, WiFi and Bluetooth. Similarly for car navigation constraints like snapping to a road is necessary before the IMU starts to drift too much.

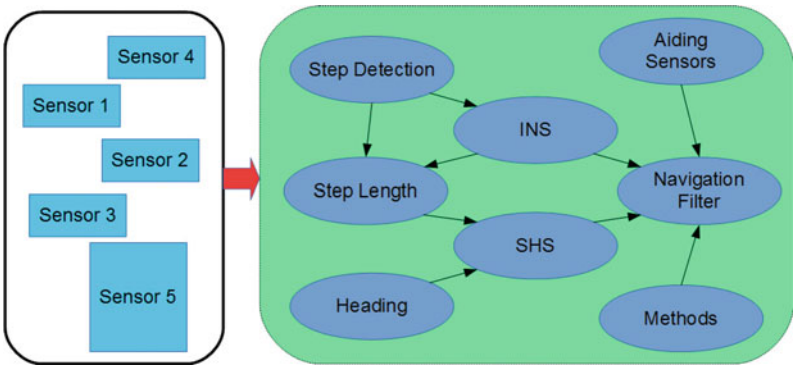

Fig. 7.7 Main topics on inertial measurement unit research for pedestrian positioning [10]

The challenge in inertial dead reckoning is the drift. Double integration of acceleration data into position information is challenging to stabilise. Drift means that the measured and derived position has an error deviation from the true position that increases over time. It is often described as the deviation from true value over elapsed time or covered distance. This is why constraints like zero velocity updates, map matching and assisting systems are needed that offer additional location information. The drift is hundreds of metres in several minutes even for high quality tactical grade inertial navigation systems.

IMU error modeling is important. Quinchia et al. [30] divides the inertial sensor errors into deterministic and stochastic errors. Deterministic errors, as errors in manufacturing and mounting, can be calibrated out, but stochastic errors need modeling. These can be bias errors, orthogonal or scale factor errors. Thermal changes while in operation also affect error characteristics and inherent noise. Error state Kalman Filter approach takes into account these non-ideal sources that come within the IMU measurements.

Figure 7.7 shows ways to produce a position solution from IMU data. In step and heading systems (SHS), the step length and heading are estimated. Peak-detection, zero crossing, moving variance, magnitude, angular energy rate detection, template matching and spectral frequency analysis are ways to detect steps. Step or gait can be divided into stance and swing periods. During stance phases foot is usually on the ground and has a zero velocity, unless on slippery ground. Zero velocity updates can be applied in order to compensate the Inertial Measurement Units (IMU) drift. Detecting gait phases is easier if the sensor is located on the foot. This gets more difficult if the sensor is placed, for example, inside a pocket.

Initialisation of an IMU when starting to navigate is another challenge. If the starting position and heading are slightly wrong in the beginning these errors will accumulate when navigating.

GNSS is most commonly the system that is integrated with INS. In loose coupling the separate GNSS and Inertial Navigation System (INS) systems produce their respective solutions which are then combined in a Kalman filter. In ultratight coupling, the INS observations are brought into the GNSS receiver tracking feedback loops.

Alternative ways for tracking the position change from IMU data are, for example, by using learning methods like statistical model comparisons, artificial neural networks and regression forests. This means that IMU values are mapped into tracks by using location specific learning methods [27].

7.4.5 GNSS, Pseudolite and Broadcast Signals

Mobile devices come with cheap GNSS receivers and cellphone communications chips with appropriate antennas. The position output rate is approximately 1–10 Hz for the GNSS receivers. The time for the first position fix is a few minutes depending on the quality of the receiver. Time to first fix is the time from device start to the

moment the GNSS receiver has first position solution available. GNSS receivers are capable of collecting GNSS and pseudolite signals while the communications chip can interpret the broadcast signals.

GNSS receivers don't work well indoors or in urban areas. Metals and water reflect RF signals as wood or concrete merely attenuate. Radio Frequency (RF) signals reflect and scatter indoors. This is called multipath. The multipath environment can be modeled by mapping the materials in the environment and by estimating the attenuation.

GNSS signal strength indoors (−170 to −190 dBW) is 5–30 dB weaker than outdoors (−158 dBW). Some high sensitivity receivers can cope even with such a weak signal by using parallel correlation and longer correlation times. GPS and GLONASS are fully operational navigation satellite systems. In addition Galileo and Beidou are being set up. Multi-constellation receivers offer more robust navigation solution. In addition integrity will get better when there are more satellites. Integrity is the receiver ability to give the correct position solution all the time without gaps. The best satellite signals that are in view can be selected more easily [24].

GNSS receiver compares a locally generated pseudorandom code with the ones sent from the satellites. Software defined radio receivers implement the GNSS signal processing in software (hardware-accelerated).

Pseudolites are transmitters on the ground and buildings that imitate GNSS signals. GNSS receivers with pulsed pseudolites cope better with the near-far problem that arises when weaker signals cannot be detected because of the presence of a very strong signal. Locata is a pseudolite system using GNSS-like signals where the nodes use accurate clock synchronisation signal. Slaves in the locatanet synchronise with a master transceiver which makes Locata reasonably accurate [6].

Television broadcast and cellular signals penetrate indoors better than GNSS signals. The positioning accuracy that can be achieved with these signals though is not very accurate (>50 m). To enhance positioning coverage TCOFDM systems are being designed. These signals share the cellular communication network spectrum capabilities but contain navigation message along with communication data. This helps the mobile receiver to act more seamlessly and have a faster time to first position fix [6].

7.4.6 WiFi, Bluetooth and Ultra-Wideband

WiFi access point scan on a mobile device takes a few seconds. Received signal strength can be read and used for positioning purposes. Traditional Bluetooth similarly needs a two way connection before signal strength can be measured. Bluetooth low energy, on the other hand, has an advertising mode which allows one way signal strength measurements even at every 20 ms. Ultra-wideband transceivers use timing approach in measuring the ranges between each other. The measurement rate is approximately few Hz, although power consumption is higher compared to Bluetooth.

New RF communication standards take into account the needs for positioning. Previously localisation had to be implemented using protocols that were designed only for communication purposes. Groves [8] expects that modeling the signal propagation medium will be a bigger challenge in the future than the new standards. Multi-wall model is a path loss model depicting the environment effect. Ray tracing analysis is another way to model the RF environment effects.

There are many existing WiFi access points. Signal strength and flight time are usually the wanted attributes. The standard 802.11v consists also of a positioning protocol. The signal strength from an access point depends on the distance and obstacles in between the access point and the mobile unit. Other method for ranging is to use accurate timing between the access points and the mobile device. Crowdsourcing WiFi access points and signal strength fingerprints is a promising way to implement extension for GPS within the near future. WiFi access point locations and signal strength maps are already commercially available. In addition cooperative position offers another dimension into real-time tracking. Sharing the location of nearby vehicles could, for example, smoothen the traffic flow [37].

Yang [37] raises the issue of the reduction of excess information. In collaborative ranging and in localisation the amount of sensors and ranging information should be kept in an optimal level.

UWB characteristics offer advantages for coping with multipath. Especially UWB impulse radio short pulses makes it easier to detect the multipath components. Repeatability is a strong advantage for ultra-wideband approach. This means that the positioning result stays consistent over a time period [25]. Figure 7.8 shows a Decawave UWB testing module. UWB tag was placed on a shoe and helmet in [38]. The tag measurements on the shoe had much more outliers due to non-line of sight conditions. Although high time resolution of UWB signals makes it easier to distinguish between original and multipath signals, the non-line of sight condition is still a challenge.

ZigBee and Radio Frequency Identification (RFID) are alternative radio protocols that can be used for positioning. Signal strengths can be converted to distances using models. In addition timing and angle information can be used in deriving position using these RF technologies. Fingerprinting the signal strengths and other distinctive features of the access points is maybe the most attractive of the current approaches. One of the major drawbacks of fingerprinting is the instability of the environment. Fingerprinting consists of an offline phase where the radiomaps are collected and of an online phase where these maps are used to derive the current position. If the radio environment changes so will the online phase positioning [22].

7.4.7 Magnetic Sensors

Honeywell 3-axis magnetoresistive sensor HMC5843 can output data at the rate of 116 Hz. Environmental magnetic anomalies and perturbations indoors, like metallic objects or radio devices, often make orientation determination difficult with magnetometers. Thus also fingerprinting the indoor magnetic environment is not

Fig. 7.8 Decawave UWB module for ranging with decimetre level accuracy and tens of metres reach

very reliable although it can be used for tracking [20]. Magnetic north measurements indoors do not work well because of this. This same can happen inside a vehicle where the changing magnetic fields disturb the measurements.

An alternative way to take advantage of the magnetic field measurements is described in [39]. Here the stable magnetic field during stance phase was measured. If there was any angular rate detected during the stance, it was used to correct the yaw drift and gyroscope bias. The magnetometer was also used to detect the swing period of the foot. As previously mentioned dead reckoning with IMU needs additional techniques to overcome the problem with drift.

7.4.8 Visible and Infrared Light Sensors

Cheap digital cameras are nowadays easy to find anywhere. These can be used for navigation in at least three different ways. Camera images can be compared and matched with a database of crowdsourced images. Camera pose can be resolved through analysis of multiple features and by comparing feature descriptors of known objects. In addition, using visual odometry, the location can be tracked from the

image flow by comparing patterns in sequential images. Stereo vision setup can be applied for more accurate camera movement estimation. Laser scanners offer even more high resolution ranging data.

Mulloni et al. [26] placed markers around the conference venue. Mobile phone camera was used to detect unique markers and to solve the positions according to the acquired marker image. The artificial markers work well under different lighting conditions and are flexible for small local scale. Still it requires preparatory work in mapping the marker locations to the area map application. For the detection to be successful, an element in the marker image on the camera image should be at least two pixels. Furthermore the power consumption of these snapshot images is not too high. CLIPS [24] is a similar system where laser spots are projected onto a wall to form a pattern that the camera recognises. It can then derive its position using this projected laser marker pattern.

Microbolometer and Golay cell based IR cameras are very expensive and are not applicable for pedestrian navigation applications. Thermopiles and pyroelectric sensors, although less accurate, are very affordable. These could be preferred in low lighting conditions where conventional image processing is impossible [17].

Image processing and computer vision literature contain many useful examples and ideas for camera based navigation. The invariance of vanishing points in images or the epipolar constraint can be used in image comparison. Patterns in images can be stored (offline) into a database and compared afterwards (online). This needs a priori work. The database collection phase is called the offline phase. The actual navigation with the help of this collected database is called the online phase. In stable lighting environments optical flow measurements of pixel movements can be tracked with optical flow algorithms. Feature tracking seems to be a feasible alternative, where consecutive images are recorded and features in them compared. It is then important to optimise the feature detection and matching algorithm effectiveness on mobile platforms. The less computing power these take the less power and time it takes to calculate the motion from images. Rapid camera movements produce blurred images and make the detection of features difficult.

Torres-Solis et al. [36] describe location detection from images. The first phase is the image acquisition. The second step is segmentation and feature extraction (SIFT/SURF, Canny edge detector, FAST or Harris corner detector). The third is the search for matches where features are compared with pre-existing information. Bundle adjustment, RANSAC, GOODSaC and BaySaC are algorithms that can be used in the refinement and the outlier rejection of the matched features. This step also involves scale rotation and luminance analysis. The last step is to compute the camera location and pose (for example, using homography).

Time-of-flight cameras consist of a pattern projector and a camera detector. This setup produces then depth information in addition to the image of the environment. In the Google Tango tablet, an infrared pattern is projected to the environment which is then registered using an RGB camera. Figure 7.9 shows this tablet in action.

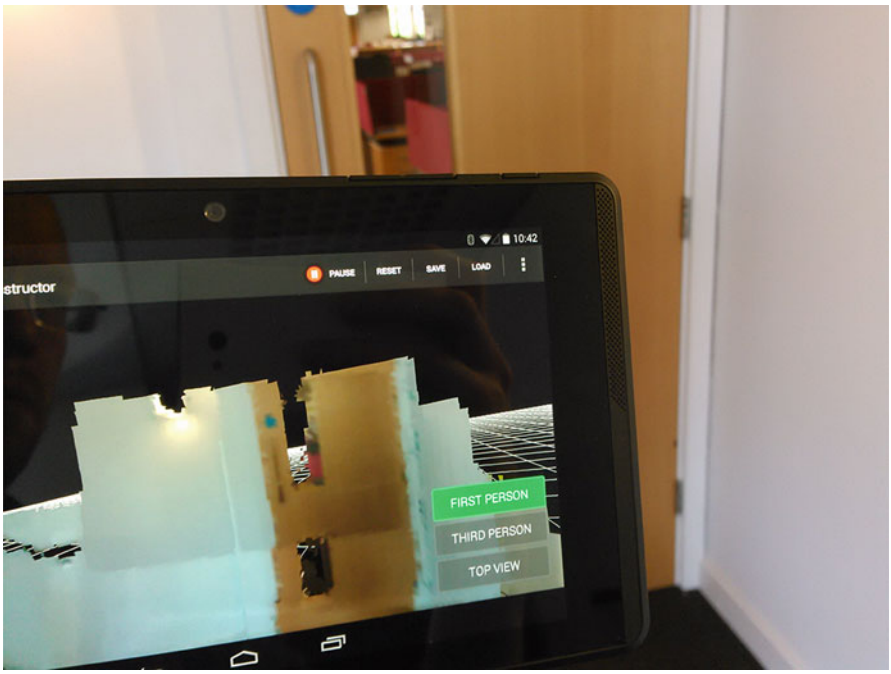

Fig. 7.9 Google Tango tablet uses a time-of-flight camera to create additional depth information of the surroundings

7.5 Sensor Fusion

Accuracy and efficiency tend to be the two characteristics between which compromises need to be made depending on the navigation application. Separate positioning technologies produce varying spectrum of data to be fused. The typical abstraction levels of the sensor data differ between technologies. Moreover depending on the sensors involved and the application, the level where the fusion of sensor data is implemented should be designed accordingly. Good example of these are the different options for integrating INS and GNSS systems. In loosely coupled integration the GNSS and INS systems are fused in the position domain, meaning that GNSS positions aid the INS solution. In tightly coupled the fusion happens within the range domain and in ultratight coupling the fusion happens inside the receiver correlation channel loops [7].

Fusion can happen in a centralised or decentralised way. In a centralised model the fusion happens in one single node while in a decentralised, multiple smaller fusion centres fuse parts of sensor data before final fusion centre. Different fusion strategies are discussed further in [31]. Joint Directories of Laboratories (JDL) process model is perhaps the best known fusion model.

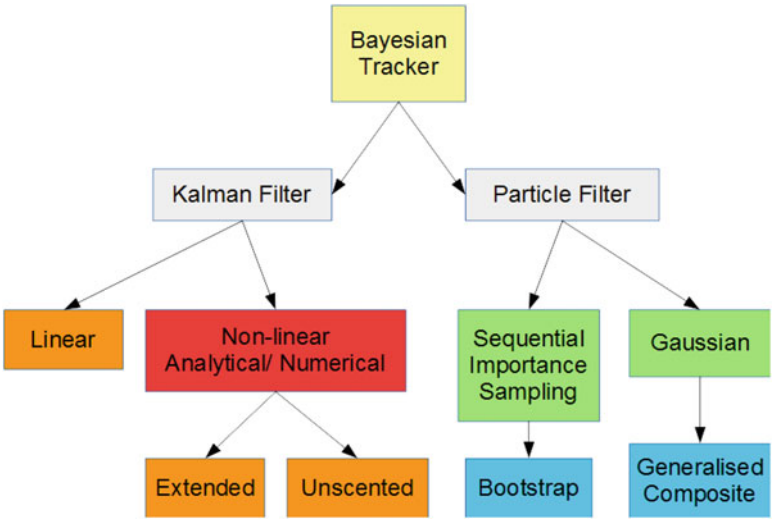

Fig. 7.10 The division of estimation and inference filters [11]

Different integration methods exist for all the abstraction levels of sensor data. Figure 7.10 shows the Kalman and particle filter variants. The gaussian trackers are considered in [23] as estimation methods. Kalman variants and covariance methods work well with low level abstraction data. This raw data is often described using dynamic quantities.

For snapshot processing or non-continuous measurement sources a particle filter might be the better choice. The inference methods include the particle filters, fuzzy logic and Bayesian inference. These methods are closer to human language and function as inference processes where logical conclusions are derived using predefined models. In between the two ends, classification methods use threshold values for classifying sensor measurements into categories [11].

7.5.1 Estimation

Common factor for estimation methods is the use of a model for approximating the position from the measurement data. The raw input data has noise or an error distribution depending on the measured parameter. In estimation methods the sensory data input is often raw data. That is, for example accelerometer readings. Least squares and Kalman filter variants belong to the estimation methods.

Bayesian filter consists of two steps: prediction and correction. Both steps use a model. Prediction model estimates the next time step. State vector and its covariance are put through the model.

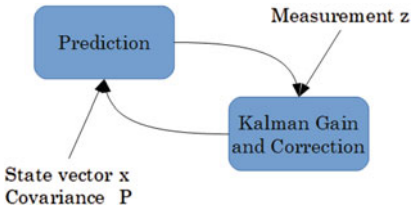

Fig. 7.11 Kalman filter loop. State and covariance go through prediction and measurement models

Measurement model dictates how the measurements affect the states. The prediction result is compared with the measurements and is weighted accordingly.

Kalman filter can handle situations where the tracking object behaves in a linear fashion. Extended Kalman filter is able to follow an object which movement can be non-linear. Unscented Kalman filter resembles particle filter slightly. It uses a small set of sample points called sigma points which propagate through the unscented model.

Not all the measurements are correct. Hard data association although assumes this. In soft data association multiple measurements are analysed. Nearest or strongest of these measurements can be picked or multiple hypothesis weighting can be used to select good quality measurements. Kalman gain itself adjusts the weight between the prediction and measurements. Figure 7.11 shows the loop of a simple Kalman filter.

7.5.2 Classification

In classification methods inputs are mapped into outputs. When the output is continuous, the mapping can be called regression as for discrete it is called classification. Machine learning methods offer alternative ways for navigation.

Positioning systems are characterised by a set of parameters like position and velocity, etc. These parameters can be tracked and optimised through learning methods. For example, if the same person walks at the same speed through checkpoint every day, this information can be used to identify that person. There are two ways to learn, by prediction using example data or by description acquiring knowledge of the data and measurements. Machine learning has often a large space and time complexity. That is why efficiency and predictive accuracy compete between applications. In the study by Nguyen et al. [27], the sensor readings were mapped into displacements using a statistical model.

Classification methods include, for example, Gaussian Processes, Principal Component Analysis, Support Vector Machines, Decision Trees and Neural Networks.

7.5.3 *Inference*

Logical deduction of higher level abstractions is typical for inference processes. Fuzzy logic is an example of this type of process. In fuzzy logic the output variable is divided into weighted regions. The combinations of weights and different input variables define the inference output. Fuzzy logic or Dempster–Shafer theory can be used in describing the uncertainty. In fuzzy logic systems the parameter values can be divided into sets. Using these sets a logical inference can be conducted. Dempster–Shafer theory further extends the manipulation of these sets or beliefs into a more complex belief choice ranges. The Dempster–Shafer is considered a robust alternative for noisy or uncertain measurement and prior data processes. Instead of a probability distribution the Dempster–Shafer theory represents a degree of belief as a belief function in the context of the transferable belief model [21].

Sequential importance sampling (SIS) defines the core of the particle filter. Filter is initialised by generating samples and weights. The samples are propagated through the system model and the weights are recalculated and normalised. State vector mean and covariance can be calculated from the progressed particles. Additionally, measurement likelihoods can be used to obtain weights. Particle filter approach is well suited for deriving multi-modal or multiple hypothesis density functions. Here if the state vector is large, then the number of samples required is large as well. The computation becomes lighter if some of the states can be modeled without sampling and instead by using a Kalman filter. This is called as Rao-Blackwellisation [11]. Figure 7.12 shows the processes of a simple particle filter. Particle propagation in time is the final step.

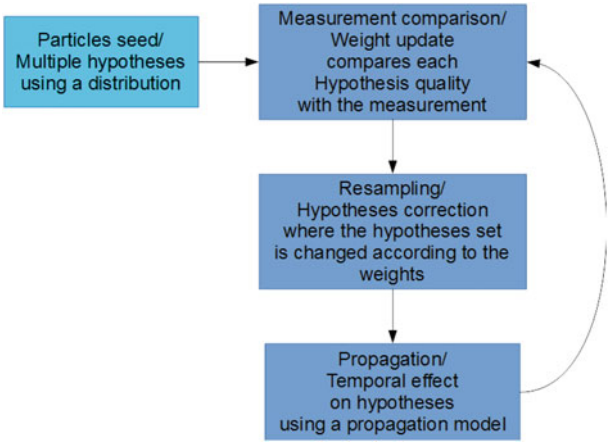

Fig. 7.12 Particle filter initiates with particle seed. Measurement comparison weights each particle. Good quality particles continue after the resampling step. Propagation of the particles in time is the final step before new iteration

Bootstrap particle filter resamples the particle set so that the weaker weighted particles die and stronger ones are duplicated. Different resampling strategies have been implemented. Regularisation is made after resampling. This is to prevent sample impoverishment. Often the stronger duplicates concentrate on certain values and are repeated. If the sequential measurements do not contain characteristic differences for location estimation the particles may diverge. Regularisation moves the particles slightly separate in a random fashion so that the sample impoverishment is prevented. In an optimal SIS particle filter care is taken in choosing the importance density that is the weights. Importance weights can be calculated before the particles are propagated. The optimal choice is to make it to minimise the variance of weights [11].

7.6 Conclusions

Different sensors produce positioning information that is heterogeneous. The heterogeneity of the sensor data is complementary. A sensor that fails to offer a necessary characteristic can be aided by a supporting sensor that is more suitable for the task. This task depends on the application and the environment. The sensors mentioned in the previous sections are listed in Fig. 7.13 below with notes on main usage on pedestrian, vehicle and building related applications.

An adaptive system is able to analyse its own performance and dynamically adjust its behaviour. Feedback loop for the context detection and filter settings is needed for these adaptive processes. Sensor management is the other important task for an adaptive system. Relations between sensor data are divided into cooperative, redundant and complementary relationships. Input data can be examined by sensor data types and the nature of the data flow. Moreover it can be viewed by the abstraction level or through the system architecture demands (centralised, decentralised and distributed). Resource distribution between tasks and scheduling

	Maps	Tactile	Air Pressure	Inertial	Broadcast RF	Local RF	Magnetic	Light
On Pedestrian	detail amount increases map database size	step detection	barometer for height detection	stepwise mechanisation	outdoor GNSS and indoor context detection	timing or signal strength based positioning	outdoor heading aiding	crowd-sourced image comparison
Onboard vehicle	additional highlighting of paths/ weather and traffic jam notification	additional odometry information from wheel	sound noise context detection/ proximity detection	dead reckoning	Outdoor GNSS	co-operative location information exchange between road users	outdoor heading aiding	visual odometry
Building Infrastructure	udated indoor maps database	smart positioning floor	ultrasound anchor architecture	vibration strength based localisation	Local amplifiers	WiFi/ Bluetooth infrastructure	magnetic environment mapping	CCTV image tracking

Fig. 7.13 The sensor technologies and notes on their main usage in pedestrian and vehicle navigation

demands prioritisation between multiple sensor subsystems. The sensors often offer different modes for changing situations. For example, dynamic range could be changed on the fly. The mode of the sensor should be possible to be readjusted during the execution of the system. The selection of the fusion hierarchy and its control depends heavily on the application and the goals that are set for the adaptive system.

The application defines the approach for sensor management and integration. For seamless navigation applications the context detection and its inference should be the focus point. The study and categorisation into commonly accepted contexts would ease further research. Correct navigation filter settings would be set whether mobile device is used in a shop, restaurant, street, countryside or while driving in a car. Best sensor configuration selection for different contexts and best integration practices for these contexts should be evaluated. Computation load is perhaps the most decisive factor in the whole mobile navigation scheme. Studies on algorithm efficiency are in place. Further comparison and combination of the methods mentioned in this document will result in finding novel efficient methods for navigation applications.

Acknowledgements This work was financially supported by EU FP7 Marie Curie Initial Training Network MULTI-POS (Multi-technology Positioning Professionals) under grant nr. 316528.

References

1. Y. Bai et al., Helping the blind to find the floor of destination in multistory buildings using a barometer, in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, July 2013, pp. 4738–4741. doi:10.1109/EMBC.2013.6610606
2. S.J.M. Bamberg et al., Gait analysis using a shoe-integrated wireless sensor system. IEEE Trans. Inform. Technol. Biomed. **12**(4), 413–423 (2008). ISSN:1089-7771. doi:10.1109/TITB.2007.899493
3. S. Boll, A. Asif, W. Heuten, Feel your route: a tactile display for car navigation. IEEE Pervasive Comput. **10**(3), 35–42 (2011). ISSN:1536-1268. doi:10.1109/MPRV.2011.39
4. C.R. Carlson, J.C. Gerdes, J.D Powell, Practical position and yaw rate estimation with GPS and differential wheelspeeds, in *Proceedings of AVEC Japan* (2002)
5. O. Daniel et al., Blind sub-Nyquist GNSS signal detection, in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Mar 2016, pp. 6575–6579. doi:10.1109/ICASSP.2016.7472944
6. Z. Deng et al., Situation and development tendency of indoor positioning. China Commun. **10**(3), 42–55 (2013). ISSN:1673-5447. doi:10.1109/CC.2013.6488829
7. P.D. Groves, *Principles of GNSS, Inertial, and Multisensor Integrated Navigation Systems* (Artech House, Boston, 2013)
8. P.D. Groves, The PNT boom, future trends in integrated navigation, in *Inside GNSS Magazine* (2013)
9. P.D. Groves et al., The four key challenges of advanced multisensor navigation and positioning, in *2014 IEEE/ION Position, Location and Navigation Symposium - PLANS 2014*, May 2014, pp. 773–792. doi:10.1109/PLANS.2014.6851443
10. R. Harle, A survey of indoor inertial positioning systems for pedestrians. IEEE Commun. Surv. Tutorials **15**(3), 1281–1293 (2013). ISSN:1553-877X. doi:10.1109/SURV.2012.121912.00075

11. A.J. Haug, *Bayesian Estimation and Tracking - A Practical Guide* (Wiley, Hoboken, 2012)
12. M. Hazas, A. Hopper, Broadband ultrasonic location systems for improved indoor positioning. IEEE Trans. Mobile Comput. **5**(5), 536–547 (2006). ISSN:1536-1233. doi:10.1109/TMC.2006
13. F. Hflinger et al., Acoustic indoor-localization system for smart phones, in *2014 11th International Multi-Conference on Systems, Signals Devices (SSD)*, Feb 2014, pp. 1–4. doi:10.1109/SSD.2014.6808774
14. G. Jee, J.-H. Song, A study on GPS/DR car navigation system using vehicle movement information, in *Coordinates*, Dec 2010
15. S. Jeon et al., Indoor WPS/PDR performance enhancement using map matching algorithm with mobile phone, in *2014 IEEE/ION Position, Location and Navigation Symposium - PLANS 2014*, May 2014, pp. 385–392. doi:10.1109/PLANS.2014.6851396
16. S. Khalifa, M. Hassan, A. Seneviratne, Adaptive pedestrian activity classification for indoor dead reckoning systems, in *2013 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Oct 2013, pp. 1–7. doi:10.1109/IPIN.2013.6817868
17. T. Kivimaki et al., A review on device-free passive indoor positioning methods. Int. J. Smart Home **8**(1), 71–94 (2014). doi:10.14257/ijsh.2014.8.1.09
18. C. Kray, G. Kortuem, Adaptive positioning for ambient systems. Kunstliche Intell. **21**(4), 56–61 (2007)
19. M. Laverne et al., Experimental validation of foot to foot range measurements in pedestrian tracking, in *Proceedings of ION GNSS* (2011)
20. B. Li et al., How feasible is the use of magnetic field alone for indoor positioning? in *2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Nov 2012, pp. 1–9. doi:10.1109/IPIN.2012.6418880
21. X. Liu et al., Multisensor joint tracking and identification using particle filter and Dempster-Shafer fusion, in *2012 15th International Conference on Information Fusion (FUSION)*, July 2012, pp. 902–909
22. E.S. Lohan, J. Talvitie, G.S. Granados, Data fusion approaches for WiFi fingerprinting, in *2016 International Conference on Localization and GNSS (ICL-GNSS)*, June 2016, pp. 1–6. doi:10.1109/ICL-GNSS.2016.7533847
23. R.C. Luo, C.C. Chang, C.C. Lai, Multisensor fusion and integration: theories, applications, and its perspectives. IEEE Sensors J. **11**(12), 3122–3138 (2011). ISSN:1530-437X. doi:10.1109/JSEN.2011.2166383
24. R. Mautz, Indoor positioning technologies, Ph.D. thesis. ETH Zurich, 2012
25. X. Meng et al., Assessment of UWB for ubiquitous positioning and navigation, in *Ubiquitous Positioning, Indoor Navigation, and Location Based Service (UPINLBS)*, Oct 2012, pp. 1–6. doi:10.1109/UPINLBS.2012.6409783
26. A. Mulloni et al., Indoor positioning and navigation with camera phones. IEEE Pervasive Comput. **8**(2), 22–31 (2009). ISSN:1536-1268. doi:10.1109/MPRV.2009.30
27. T.L. Nguyen, Y. Zhang, M. Griss, ProbIN: probabilistic inertial navigation, in *The 7th IEEE International Conference on Mobile Ad-hoc and Sensor Systems (IEEE MASS 2010)*, Nov 2010, pp. 650–657. doi:10.1109/MASS.2010.5663779
28. Open Geospatial Consortium. IndoorGML Indoor Location Standards Working Group, www.opengeospatial.org. Accessed on 08 Jan 2016
29. J. Pinchin, C. Hide, T. Moore, A particle filter approach to indoor navigation using a foot mounted inertial navigation system and heuristic heading information, in *2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Nov 2012, pp. 1–10. doi:10.1109/IPIN.2012.6418916
30. A.G. Quinchia et al., Analysis and modelling of MEMS inertial measurement unit, in *2012 International Conference on Localization and GNSS*, Jun 2012, pp. 1–7. doi:10.1109/ICL-GNSS.2012.6253129.
31. K. Rein, J. Biermann, Your high-level information is my low-level data—a new look at terminology for multi-level fusion, in *2013 16th International Conference on Information Fusion (FUSION)*, July 2013, pp. 412–417

32. M. Reinstein, M. Hoffmann, Dead reckoning in a dynamic quadruped robot: inertial navigation system aided by a legged odometer, in *2011 IEEE International Conference on Robotics and Automation (ICRA)*, May 2011, pp. 617–624. doi:10.1109/ICRA.2011.5979609
33. I. Rishabh, D. Kimber, J. Adcock, Indoor localization using controlled ambient sounds, in *2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Nov 2012, pp. 1–10. doi:10.1109/IPIN.2012.6418905
34. A. Rivero-Rodriguez, H. Leppkoski, R. Pich, Semantic labeling of places based on phone usage features using supervised learning, in *Ubiquitous Positioning Indoor Navigation and Location Based Service (UPINLBS)*, Nov 2014, pp. 97–102. doi:10.1109/UPINLBS.2014.7033715
35. L. Snidaro et al., Context in fusion: some considerations in a JDL perspective, in *2013 16th International Conference on Information Fusion (FUSION)*, July 2013, pp. 115–120
36. J. Torres-Solis, T.H. Falk, T. Chau, A review of indoor localization technologies: towards navigational assistance for topographical disorientation, in *Ambient Intelligence*, ed. by F.J.V. Molina (INTECH Open Access Publisher, 2010)
37. C. Yang, A. Soloviev, Covariance analysis of spatial and temporal effects of collaborative navigation, in *2014 IEEE/ION Position, Location and Navigation Symposium - PLANS 2014*, May 2014, pp. 989–998. doi:10.1109/PLANS.2014.6851464
38. F. Zampella et al., A constraint approach for UWB and PDR fusion, in *2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Nov 2012, pp. 1–9. doi:10.1109/IPIN.2012.6418929
39. F. Zampella et al., Unscented Kalman filter and magnetic angular rate update (MARU) for an improved pedestrian dead-reckoning, in *2012 IEEE/ION Position Location and Navigation Symposium (PLANS)*, Apr 2012, pp. 129–139. doi:10.1109/PLANS.2012.6236874

Chapter 8
Mapping the Radio World to Find Us

Pedro Figueiredo e Silva, Nunzia Giorgia Ferrara, Ondrej Daniel, Jari Nurmi, and Elena-Simona Lohan

8.1 Introduction

Finding our location on Earth has been a concern for humans throughout ages. While several thousands of years ago we were using stars to guide ourselves throughout the oceans, these days we rely on artificial satellites such as GPS, Galileo, GLONASS and others in order to quickly obtain an accurate position. A lot of the improvements done in this field have resulted from the requirements put forth by the FCC, in its e911 system [10].

However, while GNSS works well in outdoor environments, especially when there is a clear sky view to the satellites, its performance starts to degrade closer to densely populated areas. This is simply due to the increase of multipath components, buildings blocking the sky and possible interference sources [13, 14, 24]. As you roam inside buildings, these effects are emphasised even more due to the presence of walls and different building materials [8, 21, 29]. Still, this is one of the environments where people spend most of their daily time.

In part, this is one of the reasons why FCC is revising and updating its location requirements once again. Over the next years, the regulator expects e911 to achieve a vertical accuracy of 3 m in indoor scenarios [10]. This means, the regulator wants emergency service to pinpoint people's location up to the floor they are on (Fig. 8.1).

While this is a difficult task to tackle, there are solutions that can allow metre and cm level precision, such as using phase measurements for ranging purposes with WiFi [32] and UWB. However, these solutions require setting up a new infrastructure, which might be feasible in certain buildings, but enabling this in a country or even worldwide scale is simply too costly. One should also not forget

P. Figueiredo e Silva (✉) • N.G. Ferrara • O. Daniel • J. Nurmi • E.-S. Lohan
Tampere University of Technology, Korkeakoulunkatu 10, 33720 Tampere, Finland
e-mail: pedro.silva@tut.fi; jari.nurmi@tut.fi; elena-simona.lohan@tut.fi

J. Nurmi et al. (eds.), *Multi-Technology Positioning*,
DOI 10.1007/978-3-319-50427-8_8

Fig. 8.1 Future emergency systems aim to provide first responders with an accurate floor location

the cost of maintenance for such system. Having an equivalent to a GNSS system indoors is simply impractical to achieve.

Yet, the ubiquity of mobile and WiFi networks has put a device nearby or even inside every building. Hence, the question, why not use these systems as indoor positioning systems?

8.1.1 Opportunities

This question has been driving the research community who continues to seek techniques exploiting the multitude of radio signals we are embedded in constantly. Due to their opportunistic nature and the fact most are meant for other means than location purposes, signals from these systems are often referred to as signals of opportunity [5, 6, 23, 25, 28]. Sources of these signals are, for example, DTV, 3G, 4G and WiFi.

There are countless ways to approach this problem and retrieve a derivative of these signals for positioning purposes. Fingerprinting is a popular technique used in the field, which adds little complexity on the user side and is network

Fig. 8.2 Distinct features (*yellow house*) help humans navigate their environment

agnostic. Fingerprinting builds on the fact that each point in space has distinctive characteristic. For humans, our vision provides feedback on the surrounding world allowing us to position ourselves inside a building (Fig. 8.2). In fact, this works since the buildings have distinctive features that allow us to keep track of where we are or where we want to go.

The same things hold for fingerprinting where matching current observations with past observations allows to draw conclusions regarding where the device and therefore the user is located. However, the main goal is to do it so using a cheap and easy approach, which does not require setting up infrastructure and spending the user's battery throughout the process.

Therefore, its flexibility is one of its main advantages in the competing ecosystem of indoor positioning applications.

8.1.2 Overview

Over the chapter, focus will be put on using fingerprinting with RSS from WiFi signals. First off, a brief discussion on path-loss models is provided in order to understand the underlying physical world where this approach is building upon. Afterwards, the reader is walked through the actual technique. Finally, several

shortcomings are presented out in order to allow the reader to understand the lingering technical challenges and hidden disadvantages and costs of the system.

8.2 From Power to Distance

Whenever an electromagnetic wave propagates through space, it experiences a reduction in its power density, which occurs due to interaction with the surrounding environment [4, 17, 31].

8.2.1 Non-free Space Loss

FSL relates to the expected loss an electromagnetic wave experiences while transversing a diffraction-less and dispersive-less environment, where no obstacles are found on its way. Thinking of the wave as a uniform 3D ball, as it travels through space, its surface scales as the square of the radius with the power decaying as the square of the distance.

In the presence of an atmosphere the power is expected to fade more quickly than in vacuum, due to the interference caused by the presence of molecules and surrounding materials, which diffract and reflect electromagnetic waves.

Propagation prediction for indoor radio systems differs from that for outdoor systems. The geometry of the buildings, its building materials, furniture, confined and open spaces have an impact on the propagation of the signals which makes it difficult to model accurately. There are two common models used in the community, one provided by ITU [19] and the empirical log-distance model from [4].

8.2.2 Propagation Impairments

The goal of a path-loss model is to describe the loss occurred in the channel from the transmitter to the receiver. There are many factors that contribute to the loss from one terminal to the other and are known as propagation impairments. In an indoor environment some of these impairments are due to the reflection and diffraction around objects within the room, transmission loss through walls, floors and other obstacles, motion of persons and objects in rooms, among others [19].

8.2.3 ITU-R Model

The ITU-R model [19] is directly obtained from the Friis free space loss propagation equation,

$$P_r(d,f) = P_t - 10\log_{10}\left(\frac{4\pi fd}{c}\right)^2 \tag{8.1}$$

where $P_r(d,f)$ is the received power at distance d (m) for the operating frequency f (Hz), assuming a transmission power of P_t in dBm and a propagation in free space at the constant speed of light, c in m/s. To accommodate the presence of obstacles and other propagation phenomena, the equation should be written as,

$$P_r(d,f) = P_t + C - 20\eta\log_{10}\left(\frac{4\pi f}{c}\right) - 20\eta\log_{10}(d) + v \tag{8.2}$$

where the parameter η models additional losses in the path of the signal and, for that reason, it shall be equal to or larger than one. C is a constant that takes into account other system losses. The additional variable, $v \sim N(0,\sigma^2)$, is a normally distributed random variable which models the slow fading phenomenon.

For most systems, the transmission power can be estimated or then it is known by the user. In some technologies, this information might even be present in the message payload, such as in BLE. Hence, assuming P_t known and to be the same for every emitter, the unknown parameter vector $\mathbf{X} = [C,\ \eta]^T$ is estimated by solving the constrained Least Squares problem,

$$\mathbf{HX} = \mathbf{b} \text{ with } C < 0 \text{ and } \eta \geq 1 \tag{8.3}$$

where

$$\mathbf{H} = \begin{pmatrix} 1 & P_t - 20\log_{10}\left(\frac{4\pi fd}{c}\right) \\ 1 & P_t - 20\log_{10}\left(\frac{4\pi fd}{c}\right) \\ \vdots & \vdots \\ 1 & P_t - 20\log_{10}\left(\frac{4\pi fd}{c}\right) \end{pmatrix} \tag{8.4}$$

and

$$\mathbf{b} = [P_{r,1}, P_{r,2}, \ldots, P_{r,n}] \tag{8.5}$$

the measurement vector corresponding to the received power from each device.

8.2.4 Log-Distance Model

The log-distance model originates from empirical observations and is written as,

$$P_r(d) = P_r(d_0) - 10\eta \log_{10}\left(\frac{d}{d_0}\right) + v, \tag{8.6}$$

where $P_r(\cdot) < 0$ is the received signal power in logarithmic scale dependent on distance d in dBm, d_0 is a reference distance, $\eta > 0$ is the path-loss exponent and $v \sim N(0, \sigma^2)$ is a normally distributed random variable which models the slow fading phenomenon. Both η and w are dependent on the propagation environment and are considered to be different for each path between receiver and transmitter.

For each device, the unknown vector is $\mathbf{X} = \eta$ and by solving the constrained Least Squares,

$$\mathbf{HX} = \mathbf{b} \text{ with } \eta > 0 \tag{8.7}$$

where

$$\mathbf{H} = \begin{pmatrix} P_r(d_0) - 10\log_{10}\left(\frac{d_1}{d_0}\right) \\ P_r(d_0) - 10\log_{10}\left(\frac{d_2}{d_0}\right) \\ \vdots \\ P_r(d_0) - 10\log_{10}\left(\frac{d_n}{d_0}\right) \end{pmatrix} \tag{8.8}$$

and

$$\mathbf{b} = [P_{r,1}, P_{r,2}, \ldots, P_{r,n}] \tag{8.9}$$

as before, the measurement vector for each device.

8.2.5 Typical Values

There are other models available in the literature that take into account other environment effects, such as multipath, interference and different material propagation [4]. However, ITU-R and log-distance method are the simplest and used more often, especially the former. In a log-distance model, the reference distance propagation acts as an accumulator for other losses in the system, making it a suitable model for comparisons.

Table 8.1 Typical path-loss exponent values [4, 31]

Environment	η (log-distance)
Free space	2
Rural	2.7–3.5
City	3–5
Building	4–6
Factory	2–3

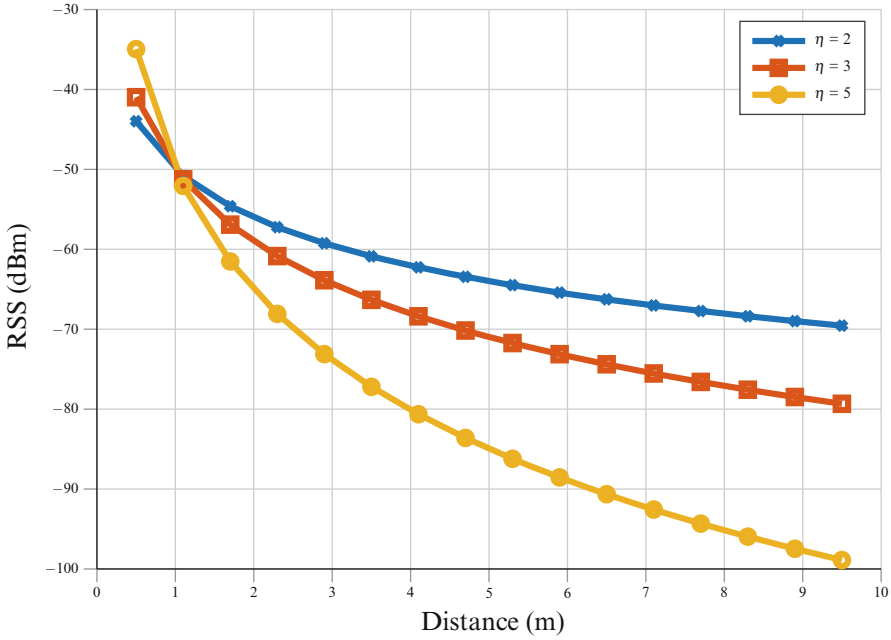

Fig. 8.3 Path-loss illustration with different exponents using log-distance model

When dealing with these models a common necessity is to understand how much the path-loss exponent varies and changes across environments. For that reason, Table 8.1 presents a couple of typical values considered for η in both models.

Figure 8.3 illustrates the effect of η on the RSS with regard to distance. For a typical WiFi terminal, its sensitivity is usually about −100 dBm, meaning that is the lowest signal detection threshold.

These models are generic to any electromagnetic wave, but in this case we are interested in WiFi signals, which typically operate in the 2.4 and 5 GHz band. A higher frequency results in a bigger free space loss, meaning a smaller coverage.

Overall, while path-loss models provide an idea on the expected RSS, they are still very dependent on the environment and sudden changes in it. For that reason, applications that relay on RSS often rely on a filter to take into account

past information. For example, a reasonable approach to take with fingerprinting approaches is to take the mean value of several measurements or several similar paths (e.g. MIMO routers).

8.3 Fingerprinting

Fingerprinting is a technique introduced in [2]. This technique has become the standard for low cost indoor positioning services, since it builds on existing infrastructure. Therefore, the end user can obtain an estimation of its location by simply having its WiFi card on and allowing a third party to forward the data to their servers.

Besides the low cost in infrastructure and low complexity on the user side, this technique is also quite flexible since one can apply its fundamentals to many systems or technologies [3, 9, 25, 26].

It relies on the principle of mapping and matching a particular location to a set of distinctive features, for example, radio signals, sounds, light and smell. For that reason, the technique is divided into two phases: a learning phase and an online phase [7, 12].

Over this chapter, the discussion focuses on using the RSS of WiFi networks as the sole input for fingerprinting [18] (Fig. 8.4).

8.3.1 Learning Phase

The main goal of this phase is to establish the correspondence between each physical location at the location of interest, to a set of physical features. As mentioned, in this case the feature of interest will be the RSS of nearby WiFi networks.

For that reason, in a classic approach, a service provider needs to go through each of the locations in the environment and store the RSS to a database.

8.3.1.1 Building the learning database

A first step in building a successful environment for fingerprinting acquisition is to know the actual real world coordinates of the area of interest. Since GNSS is unavailable, one needs to rely on other technologies, for example, IMU, to keep track of where the measurements are being taken inside the building [30].

This alternative is convenient since most mobile devices these days contain several accelerometers and a few of them even have gyroscopes, magnetometers and barometers. Hence, with these sensors one can extrapolate the direction and speed of the device collecting the data. Other simpler alternatives include having a fixed step velocity or manually compute the new coordinate. The choice taken at this phase will take a toll on the accuracy, however, since the whole system is built

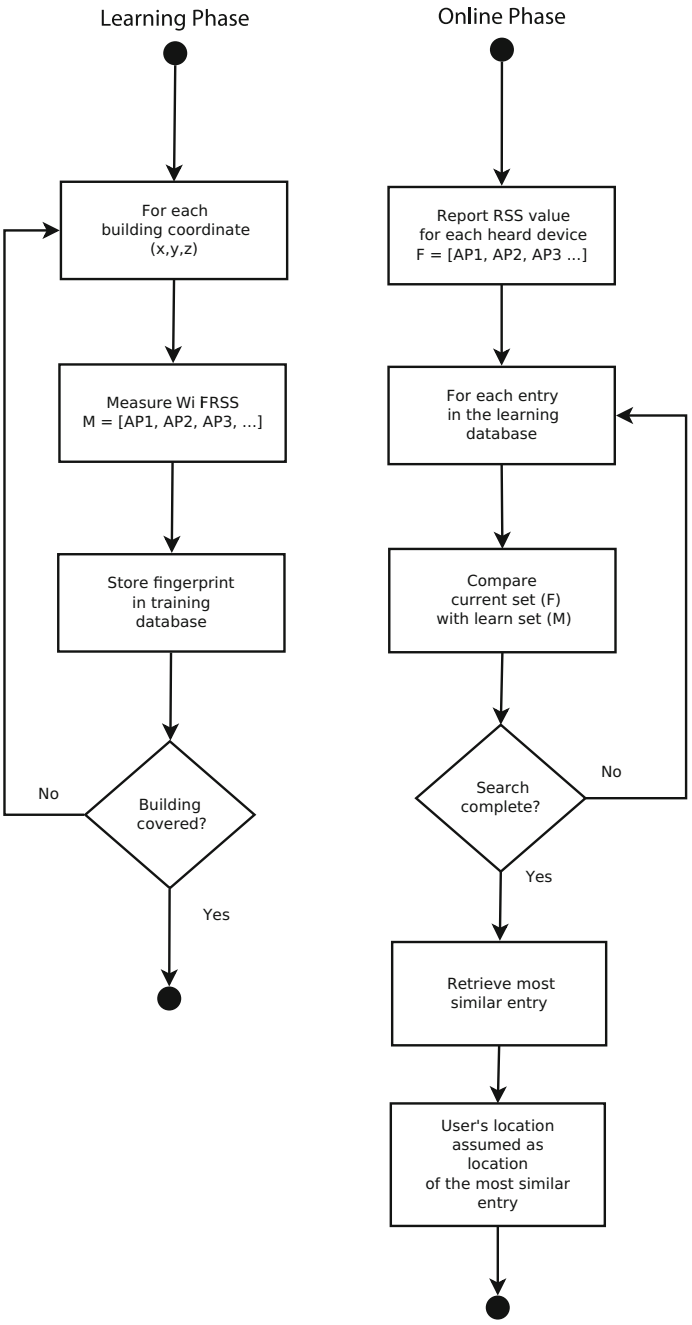

Fig. 8.4 Overview of fingerprinting phases

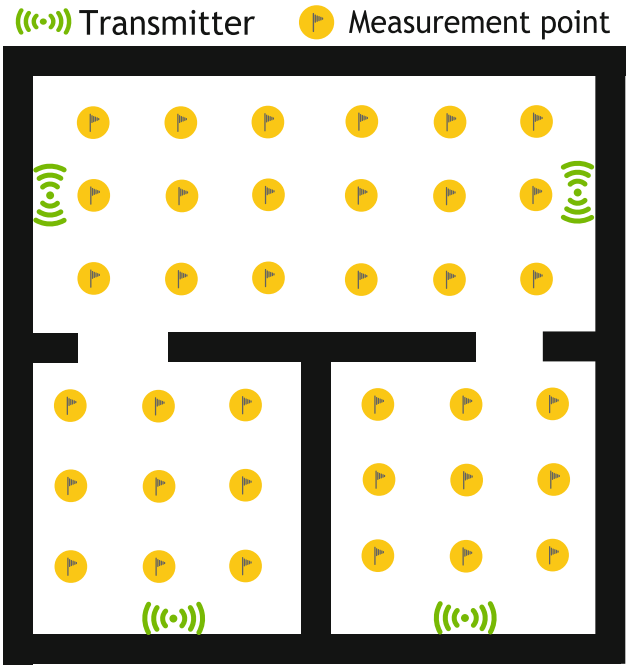

Fig. 8.5 Measurement campaign for the training phase

on already inaccurate measurements, practical approaches tend to be favoured in this phase. Nevertheless, a practical aspect of this phase is to have a well-known reference point which can be translated from WGS-84 into a local reference system.

Once there is an established process on how to map the WiFi measurements to a local coordinate, the operator navigates the environments through several points in the environment, storing the values in a local or remote database. This is illustrated in Fig. 8.5 where the encircled flags represent measurement points and the WiFi access points are represented by the dots and waves.

In each of the measurement points, P_i, depicted in Fig. 8.5, assume the RSS measurements are stored in a vector M defined as,

$$M = [(\text{Addr}, \text{RSS})_1, (\text{Addr}, \text{RSS})_2, \ldots, (\text{Addr}, \text{RSS})_N], \tag{8.10}$$

where Addr_n is the MAC address of the n-th device.

The learning database is complete after visiting and performing the RSS measurements at each of the points. Hence, assume the learning database, T, described as,

$$T = [(P, M)_1, (P, M)_2, \ldots, (P, M)_I], \tag{8.11}$$

where P_i is the i-th point's coordinate and I the total number of measurement points.

8.3.2 Online Phase

In this phase, a service user will request from the provider a location corresponding to a set of features acquired by the user's device. Over at the service provider, the set of incoming features F is mapped into the training database T. This mapping happens by comparing each entry in the database and assuming the user's location as the entry's location with the most similar set of measurements.

There are several approaches to find the most similar or likely entry in the learning database. However, all of them will require testing the incoming set of features F to all the I entries in the database. Thus, while cycling through the I entries of the database, the likelihood for the user device to be at each location will be given by,

$$\mathscr{L}(p,j) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\big(F(k) - M_j(p,k)\big)^2}{2\sigma^2} \right), \tag{8.12}$$

where σ is the variance due to shadowing and p a set of coordinates. $M_j(p,k)$ is the j-th fingerprint belonging to device k and $F(k)$ is the reported measurement for device k.

After going through the full search space, the final likelihood is computed by summing each individual logarithmic likelihood of each device, defined as the cost function, C,

$$C(p) = \sum_{i=1}^{N} \log(\mathscr{L}(p,k)). \tag{8.13}$$

The estimated location of the user is then the arg max of the cost function,

$$\hat{p} = \arg_{\max}\{C(x,y)\}. \tag{8.14}$$

8.3.3 A Simple Example

Assume that a user enters inside a building and its device reports the following measurement matrix F:

$$F = \begin{bmatrix} \text{MAC}_1 & \text{MAC}_2 & \text{MAC}_3 & \text{MAC}_4 & \text{MAC}_5 \\ -70 & -78 & -80 & -90 & -95 \end{bmatrix} \tag{8.15}$$

where the first row is the MAC address of 4 WiFi access points in the building, where the index represents the access point in Fig. 8.6.

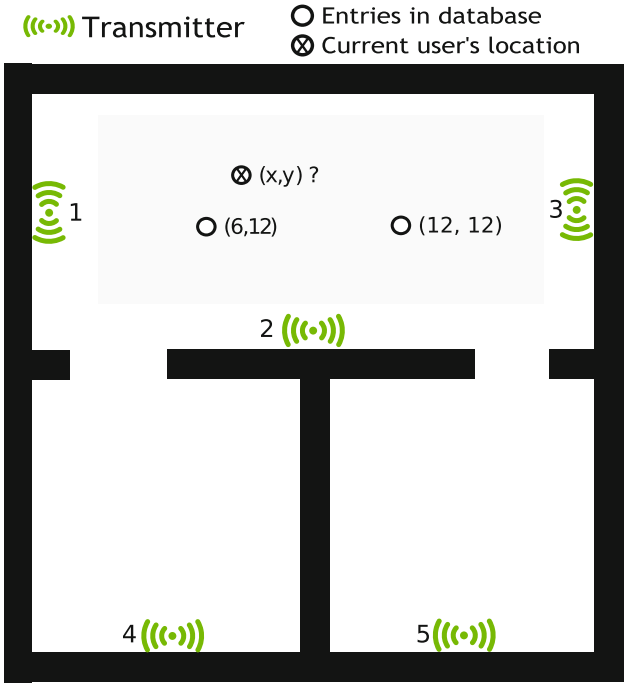

Fig. 8.6 User requesting for the system to determining its location. The *shaded area* denotes an area with a higher change of having similar measurement vectors, due to the lack of obstacles

Determining the user's location is now a matter of searching in the database for the most similar entry. Assume that the training database is similar to the one in Fig. 8.5 and structure the same way as in (8.11). One can go through all its entries and compare which access points were heard at each entry.

Let the training vectors associated with coordinates $p = (6, 12)$ and $p = (12, 12)$, as T_m and T_n, respectively, contain the following information:

$$T_m = \begin{bmatrix} 6 & 12 & & & \\ \text{MAC}_1 & \text{MAC}_2 & \text{MAC}_3 & \text{MAC}_4 & \text{MAC}_5 \\ -68 & -71 & -74 & -97 & -102 \end{bmatrix}, \tag{8.16}$$

$$T_n = \begin{bmatrix} 12 & 12 & & & \\ \text{MAC}_1 & \text{MAC}_2 & \text{MAC}_3 & \text{MAC}_4 & \text{MAC}_5 \\ -84 & -72 & -60 & -107 & -88 \end{bmatrix}. \tag{8.17}$$

Comparing each of these entries to F will result in the following vectors of differences:

$$F - T_m = (-2, -7, -6, 7, 7), \tag{8.18}$$

and

$$F - T_n = (14, 6, 20, 17, -7). \tag{8.19}$$

Applying (8.13) for the cost function calculation leads to $C(p = (6, 12)) = -13$ and $C(p = (12, 12)) = -20$, meaning the location of the user is estimated by (8.14) as $\hat{p} = (6, 12)$.

However, in a real scenario, there might several entries with very similar values, especially on open environments with little obstacles (see shaded area in Fig. 8.6). For that reason, it is common to estimate the user's location as the average of the best candidates found through (8.13).

8.3.4 Shortcomings

Despite its simplicity, fingerprinting has several disadvantages and challenges [1, 15, 18, 20, 22]. Reliance on RSS makes this technique difficult to keep up, regarding consistency of the training databases. Service providers must continuously or periodically re-learn the environment, otherwise the training database will no longer be consistent. This implies additional costs in the maintenance of the system. However, relying on crowdsourced data allows for offsetting this cost in the learning process, which means relying on the data provided by the users for building the training database.

8.3.4.1 Consistency

Regarding consistency, the biggest issue in fingerprinting is to guarantee that changes in the environment do not impact significantly the performance of the system. Since the whole system relies on RSS measurements, a simple shuffling of furniture can disrupt the signal propagation significantly. While the system should be able to cope with some of these changes, including people, permanent interference or changes in the environment will reduce the overall accuracy of the environment [35].

8.3.4.2 Storage

Data storage requirements are quite high and pose a significant cost to the system. As an example, assume that a minimum length fingerprint requires 48 bits for the AP MAC address, another 48 bits for the device's MAC address or user identifier and 8 bits for the RSS measurement. In addition to this, let there be a mean of 4 access points per fingerprinting location. Therefore, while the user identifier can only be sent once, each access point has a different MAC address that needs to be reported,

which means that the mean total bytes per fingerprint would be of $4*(6+1)+6 = 34$ bytes. Assuming a fingerprinting at every metre over an area of 100 m^2 then the total number of bytes per floor will be $100 * 100 * 34 = 340{,}000$ bytes or 340 kB. Hence, if the service is to be extended globally this number will increase exponentially and turn into several terabytes of data. As an example, there are approximately 3,400,000 buildings in London according to [27].

A good way to tackle this is to learn the environment parameters and through statistical interpolation, obtain a synthetic grid of the environment. This results in less data being transmitted to the user and the possibility of offloading the processing to the user side, if needed [11, 33].

8.3.4.3 Privacy

Privacy advocates are fast to point out how this information allows the service providers to easily profile and track people. Most common mapping applications available in Android and iOS will report the measurements back to the service provider, which relies on crowdsourcing data to keep its training database relevant [34].

Nevertheless, the privacy issue relies on the fact that the user is sending identifiable data to the server. While the data is anonymised with an application ID, the service provider has the capability of understanding whom this ID belongs to. Therefore, the end user can be the target of direct advertisement with respect to its location.

8.3.4.4 Hardware

Another limiting factor for fingerprinting is the multitude of devices, which makes it challenging for the service provider to integrate different quality of measurements. In reality, most chip manufacturers report an integer number that is already a derivative of the RSS. Even more, how this indicator is built is most of the times left out of the receiver specification, making it hard to understand how to properly scale the measurement between different devices [1, 16].

8.4 Conclusions

Despite all the challenges and difficulties in correctly modelling path loss, fingerprinting is popular because it works in most environments. That is especially interesting as new technology is emerging in the market, such as BLE. BLE is attempting to shift or call the attention of customers towards a shop, shelf product or a temporary promotion, in a way, a modern approach to the yellow and red promotion stickers supermarkets employ in their shelves. However, as and if these

became ubiquitous, their added presence in the environment should add more distinct features to the environment, increasing the likelihood of a more accurate location through fingerprinting.

However, there are still many hurdles to go through in order to achieve the future requirements of the FCC. This will certainly move the indoor positioning techniques and systems further as new systems and signals appear in order to tackle the performance requirements.

Until then, WiFi is a good complement to cell and GNSS based positioning in indoor scenarios.

Acknowledgements This work was financially supported by EU FP7 Marie Curie Initial Training Network MULTI-POS (Multi-technology Positioning Professionals) under grant nr. 316528.

References

1. N. Alsindi et al., An empirical evaluation of a probabilistic RF signature for WLAN location fingerprinting. IEEE Trans. Wirel. Commun. **13**(6), 3257–3268 (2014). ISSN:15361276. doi:10.1109/TWC.2014.041714.131113
2. P. Bahl, V.N. Padmanabhan, RADAR: an in-building RF-based user location and tracking system, in *Proceedings IEEE INFOCOM 2000. Conference on Computer Communications. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies (Cat. No.00CH37064)* (2000). https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/infocom2000.pdf
3. A.D. Cheok, L. Yue, A novel light-sensor-based information transmission system for indoor positioning and navigation. IEEE Trans. Instrum. Meas. **60**(1), 290–299 (2011). ISSN:0018-9456. doi:10.1109/TIM.2010.2047304
4. D. Cichon, T. Kurner, Propagation prediction models. COST 231 Final Report, pp. 116–208 (1995)
5. A. Coluccia, F. Ricciato, G. Ricci, Positioning based on signals of opportunity. IEEE Commun. Lett. **18**(2), 356–359 (2014). doi:10.1109/LCOMM.2013.123013.132297
6. A. Dammann, S. Sand, R. Raulefs, Signals of opportunity in mobile radio positioning, in *Signal Processing Conference*, Eusipco (IEEE, Bucharest, 2012), pp. 549–553
7. B. Dawes, K.-W. Chin, A comparison of deterministic and probabilistic methods for indoor localization. J. Syst. Softw. **84**(3), 442–451 (2011). ISSN:01641212. doi:10.1016/j.jss.2010.11.888
8. Z. Deng, Y. Yu, X. Yuan, Situation and development tendency of indoor positioning. China Commun. **10**, 42–55 (2013). doi:10.1109/CC.2013.6488829
9. R. Faragher, R. Harle, Location fingerprinting with bluetooth low energy beacons. IEEE J. Sel. Areas Commun. **33**, 2418–2428 (2015). ISSN:0733-8716. doi:10.1109/JSAC.2015.2430281
10. Federal Communications Commission. Fourth Report and Order (2015)
11. J.-A. Francisco, R.P. Martin, A method of characterizing radio signal space for wireless device localization. Tsinghua Sci. Technol. **20**(4), 385–408 (2015). ISSN:1007-0214. doi:10.1109/TST.2015.7173454
12. C. Gentile et al., Geolocation techniques, principles and applications (2013), pp. 59–97. ISBN:9781461418351. doi:10.1007/978-1-4614-1836-8
13. D. Gingras, An overview of positioning and data fusion techniques applied to land vehicle navigation systems (2009). http://www.gel.usherbrooke.ca/LIV/index_htm_files/denis%20gingras%20chapter.pdf. http://www.igi-global.com/chapter/overview-positioning-data-fusion-techniques/5489

14. P.C. Gomez, Bayesian signal processing techniques for GNSS receivers: from multipath mitigation to positioning. Ph.D. Universitat Politcnica de Catalunya, 2009
15. F. Gustafsson, F. Gunnarsson, Mobile positioning using wireless networks: possibilities and fundamental limitations based on available wireless network measurements. IEEE Signal Process. Mag. **22**(4), 41–53 (2005). ISSN:1053-5888. doi:10.1109/MSP.2005.1458284
16. S. Han et al., Cosine similarity based fingerprinting algorithm in WLAN indoor positioning against device diversity, in *2015 IEEE International Conference Communications (ICC)*, 61401119 (2015), pp. 2710–2714
17. H. Hashemi, The indoor radio propagation channel. Proc. IEEE **81**(7), 943–968 (1993). ISSN:00189219. doi:10.1109/5.231342
18. V. Honkavirta et al., A comparative survey of WLAN location fingerprinting methods, in *2009 6th Workshop Positioning, Navigation Communication* (IEEE, New York, 2009), pp. 243–251. ISBN:978-1-4244-3292-9. doi:10.1109/WPNC.2009.4907834
19. ITU-R, Recommendation ITU-R P.1238-7 (2012)
20. C. Laoudias, C.G. Panayiotou, P. Kemppi, On the RBF-based positioning using WLAN signal strength fingerprints, in *Small* (2010), pp. 93–98
21. H. Liu et al., Survey of wireless indoor positioning techniques and systems. IEEE Trans. Syst. Man Cybern. C (Appl. Rev.) **37**(6), 1067–1080 (2007). ISSN:1094-6977. doi:10.1109/TSMCC.2007.905750
22. L. Mailaender, On the CRLB scaling law for received signal strength (RSS) geolocation, in *2011 45th Annual Conference on Information Science and Systems, CISS 2011* (2011), pp. 2–7. doi:10.1109/CISS.2011.5766210
23. R. Mautz, Indoor positioning technologies. Ph.D. thesis, ETH Zurich, 2012
24. C. Mensing, S. Sand, A. Dammann, GNSS positioning in critical scenarios: hybrid data fusion with communications signals, in *2009 IEEE International Conference Communication Work*, 2 June 2009, pp. 1–6. doi:10.1109/ICCW.2009.5207983
25. V. Moghtadaiee, A.G. Dempster, S. Lim, Indoor localization using FM radio signals: a fingerprinting approach, in *2011 International Conference Indoor Position. Indoor Navigation* Sept 2011, pp. 1–7. doi:10.1109/IPIN.2011.6071932
26. V. Pasku et al., A positioning system based on low frequency magnetic fields. IEEE Trans. Ind. Electron. **63**, 2457–2468 (2016). ISSN: 0278-0046. doi:10.1109/TIE.2015.2499251
27. G. Piggott, Number of properties in London (2014). https://www.cityoflondon.gov.uk/business/economic-research-and-information/statistics/Pages/default.aspx
28. M. Robinson, R. Ghrist, Topological localization via signals of opportunity. IEEE Trans. Signal Process. **60**(5), 2362–2373 (2012). ISSN:1053-587X. doi:10.1109/TSP.2012.2187518
29. G. Seco-Granados et al., Challenges in indoor global navigation satellite systems, in *IEEE Signal*, Feb 2012, pp. 108–131
30. J. Seitz et al., Sensor data fusion for pedestrian navigation using WLAN and INS, in *Proceedings of the Symposium Gyro Technology 2007* (Karlsruhe 2007), pp. 1–10. J. Seitz, L. Patino-Studencki, B. Schindler, S. Haimerl, J. Gutierrez, S. Meyer, J. Thielecke, Sensor data fusion for pedestrian navigation using WLAN and INS, in *Symposium Gyro Technology* (2007). https://cris.fau.de/converis/publicweb/Publication/1023736
31. M. Shafi, S. Ogose, T. Hattori, PathLoss measurements for wireless mobile systems, in *Wireless Communications in the 21st Century* (IEEE, New York, 2009), pp. 185–194. doi:10.1109/9780470547076.ch10
32. subpos.org. *SubPos* (2016)
33. J. Talvitie, M. Renfors, E.S. Lohan, Novel indoor positioning mechanism via spectral compression. IEEE Commun. Lett. **20**(2), 352–355 (2016). ISSN:1089-7798. doi:10.1109/LCOMM.2015.2504097
34. X. Wu et al., Privacy preserving RSS map generation for a crowdsensing network. IEEE Wirel. Commun. **22**(4), 42–48 (2015). ISSN:1536-1284. doi:10.1109/MWC.2015.7224726
35. Y. Yuan et al., Estimating crowd density in an RF-based dynamic environment. IEEE Sensors J. **13**(10), 3837–3845 (2013). ISSN:1530-437X. doi:10.1109/JSEN.2013.2259692

Chapter 9
Survey on 5G Positioning

Arash Shahmansoori, Gonzalo Seco-Granados, and Henk Wymeersch

9.1 Introduction

The rapid increase of mobile data volume, the use of smartphones, and the global bandwidth shortage are the main challenges for current wireless networks. At any given location, the maximum available bandwidth for all cellular technologies is 780 MHz with the carrier frequencies ranging from 700 MHz to 2.6 GHz [43]. A tenfold increase of the data rate requires an almost unavoidable increase of the available bandwidth. Given that the goal is to have available a bandwidth on the order of GHz for high data rate communication with low latency and higher localization accuracy, millimeter wave (mm-wave) frequencies are considered as one of the best candidates. Moreover, increasing the bandwidth provides a better time resolution, thereby ensuring the accurate estimation of the time-of-arrival (TOA) that is used for localization. Figure 9.1 shows that the mm-wave spectrum ranging from 30 to 300 GHz provides more spectrum in bands not previously used in cellular. Particularly, for carrier frequencies $f_c < 6$ GHz the spectrum has a maximum bandwidth $B = 0.555$ GHz, in the centimeter wave (cm-wave) frequencies it is possible to achieve a bandwidth $B = 1.3$ GHz with $f_c = 28$ GHz. In the mm-wave frequencies, we achieve a unlicensed bandwidth $B = 7$ GHz at $f_c = 60$ GHz. Spatial processing techniques relying on massive Multiple Input, Multiple Output (MIMO) transceivers can also be applied in mm-wave frequencies [46]. Moreover, the spectral allocations in mm-wave frequencies are closer to each other than pieces of spectrum used by the cellular operators nowadays, which are

A. Shahmansoori (✉) • G. Seco-Granados
Universitat Autonoma de Barcelona, Barcelona, Barcelona, Spain
e-mail: arash.shahmansoori@uab.cat; gonzalo.seco@uab.cat

H. Wymeersch
Chalmers University of Technology, Gothenburg, Sweden
e-mail: henkw@chalmers.se

J. Nurmi et al. (eds.), *Multi-Technology Positioning*,
DOI 10.1007/978-3-319-50427-8_9

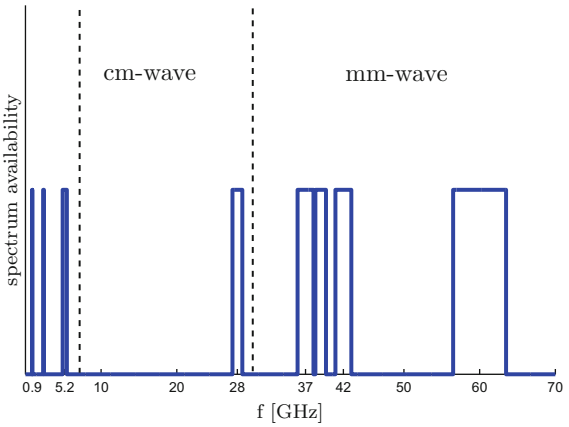

Fig. 9.1 Mm-wave spectrum for 5G

scattered between 700 MHz and 2.6 GHz. This makes mm-wave frequencies more homogenous. Despite the aforementioned advantages, using mm-wave frequencies presents some challenges including path-loss and atmospheric attenuation.

However, it has been shown that attenuation due to rain and atmospheric absorption has a negligible impact on the mm-wave at 28–38 GHz for small distances (i.e., less than 1 km). Due to attenuation at mm-wave frequencies, directional antennas[1] can be used at the transmitters and receivers to overcome the path-loss effects. Using a large number of antennas provides narrow beams towards the user that makes the mm-wave link highly directional. Moreover, the large bandwidth at the mm-wave frequencies provides TOA estimates of high accuracy. Higher directionality and higher TOA estimation accuracy lead to better localization accuracy [66].

Furthermore, the location of the user is extremely important for the transmitter in highly directional communication. Knowing the location of the user the transmitter can steer the beam directly or to a reflected path. For the case that the line-of-sight (LOS) is blocked, steering the beam to the reflected path with the strongest signal power can be helpful for user localization. The data transmission is increased based on the statistical channel knowledge for user location. This leads to a synergy between localization and communication. In today's technologies such as Global Positioning System (GPS), accurate location information cannot be provided for indoors and in urban canyons. Other technologies such as Ultra Wide Band (UWB) can provide indoor localization with the cost of high hardware complexity [53]. Also, WiFi can provide indoor localization at low cost but not so high accuracy as GPS outdoors and UWB indoors [16].

[1] Directional antenna is an antenna designed to radiate or receive greater power in specific directions allowing for reduced interference from unwanted sources.

The use of 5G technologies to obtain position and orientation was previously explored in [14, 47, 60] for mm-wave and in [17, 25, 49] for massive MIMO. Beam training protocols through direction of arrival[2](DOA) were considered in [47]. A hypothesis testing user localization approach was presented in [14] using the concept of channel sparsity that is due to few and clustered paths. These methods limit virtual angle spacings $1/N_{\mathrm{Tx}}$ and $1/N_{\mathrm{Rx}}$ due to the limited number of antenna elements in the transmitter N_{Tx} and receiver N_{Rx}. Localization based on received signal strength (RSS) was considered in [60]. This approach provides meter-level positioning accuracy. A method to estimate the position of the user devices using an extended Kalman filter combined with travel time of the signal from the transmitter to the receiver (TOA) and DOA estimations in the uplink was proposed in [28, 62]. This method assumes LOS propagation, thanks to the high density of the access nodes and it estimates the clock offsets between user nodes. A method based on DOA and RSS estimation for non-cooperative transmitter localization was considered in [63]. This method uses an antenna structure that can selectively receive energy from different sectors (sectorized antennas) to obtain sector-powers as sufficient statistics for DOA and RSS estimation. However, the method assumes that different samples in a sectorized antenna are received sequentially in time, what can slow down the localization. Using massive MIMO systems, the work in [25] considered AOA/angle-of-departure (AOD)[3] estimation for localization, and [17] considered the localization in an LOS scenario by joint TOA, AOA, and AOD.

In this chapter, it is shown that mm-wave and massive MIMO, both candidates for 5G networks, are also enabling technologies for localization. First, a brief overview of 5G systems and the main challenges including path-loss effects are provided. Different path-loss models are presented and the main differences between the path-loss effects in the mm-wave frequencies and UWB systems are explained. For the sake of comparison, UWB systems are used due to providing higher localization accuracy for indoor applications compared to WiFi. Since the estimation of mm-wave channels is of critical importance for user localization, the physical channel model for mm-wave systems together with the limited scattering property is presented. This property leads to the sparsity of the mm-wave channels, which differs from UWB channels since the later are rich in scattering. It is demonstrated that the TOA, AOA, and AOD can be estimated using the sparsity of the mm-wave channels. Hybrid beamformers are explained as the most promising solution for accurate beam steering in mm-wave; they can be used to generate narrow beams used for the user localization by beam training protocols. Finally, different localization techniques based on the TOA, AOA, and AOD and their combination are presented as the promising solutions in the mm-wave frequencies.

[2]DOA or angle-of-arrival (AOA) is defined as the angle between the received beam with respect to a reference line in the receive antenna array.

[3]AOD is defined as the angle between the transmitted beam with respect to a reference line in the transmit antenna array.

This chapter is organized as follows. Section 9.2 briefly explains the relation between 5G and cognitive radio for localization. Section 9.3 represents an overview of 5G systems. Section 9.4 proposes a physical channel model in the mm-wave frequencies and the sparsity in delay and angle subspace. Section 9.5 provides an overview of the hybrid beamformers and beam training protocols for AOA and AOD estimation. Section 9.6 presents different localization techniques based on combined delay and angle information. Section 9.7 provides the simulation results. Section 9.8 concludes the chapter.

9.2 The Relation Between 5G and Cognitive Radio for Localization

Both 5G and cognitive radio are considered as future technologies. The future 5G networks require low cost, high spectral efficiency with the large number of connected devices, and low latency. One of the promising techniques to meet these requirements is cognitive radio. Specifically, cognitive radio increases spectral efficiency using opportunistic and shared spectrum access. In particular, the location of a primary user enables several key capabilities in cognitive radio such as spatio-temporal spectrum sensing, intelligent location-aware power control and routing, as well as aiding security and spectrum policy enforcement. When GPS information is unavailable due to either cost considerations or other limitations, different localization techniques introduced in this chapter for 5G networks including DOA and RSS can be used for primary user localization in cognitive radio [61].

9.3 On 5G Systems

In this section, we briefly describe 5G systems, their properties, and benefits. First, mm-wave systems are explained in terms of their carrier frequencies, bandwidth, and data rate. Second, the benefits and challenges of the massive MIMO systems are described. Third, the concept of device-centric architecture in 5G systems is addressed. Finally, the concepts of device-to-device (D2D) communication, location-aware communications, and ultra dense networks are described.

9.3.1 Mm-Wave

The mm-wave band provides 5G systems with an amount of bandwidth on the order of GHz. Some of the implications of using the mm-wave spectrum include:

- The possibility to use of cognitive radio techniques to share the spectrum with satellite or radar systems.

- The capability to generate very narrow beams with smaller directional and adaptive antenna arrays, thanks to small wavelength.

Moreover, mm-wave can provide high peak, average and outage rates on the order of gigabit per second (Gbps), as required in different 5G scenarios, e.g., autonomous driving.

9.3.2 Massive MIMO

Massive MIMO systems can operate either at mm-wave frequencies or lower ones [56]. Massive MIMO systems are considered as systems with large number of antenna elements[4] in the transmitter $N_{\text{Tx}} \gg 1$ and with P single-antenna or multi-antenna terminals. To suppress interference and achieve the sum capacity of the multi-user channel, it is required to have P channel vectors mutually orthogonal (favorable propagation). For the case of mutually non-orthogonal channel vectors, advanced signal processing methods (e.g., dirty paper coding [8]) are used. Favorable propagation can be achieved with sufficiently large number of antenna elements N_{Tx} (e.g., $N_{\text{Tx}} = 100$) for a given number of single-antenna terminals P (e.g., $P = 12$) in Non Line-of-Sight (NLOS) environments with rich scattering or LOS environments with dropping a few worst terminals that cause non-orthogonal channel vectors. Massive MIMO enables simple spatial multiplexing/de-multiplexing procedures. However, channel estimation is a challenging step in the massive MIMO systems. The channel coherency is limited by the propagation environment, user mobility, and the carrier frequency that limits the number of orthogonal pilots. Moreover, reuse of pilots leads to pilot contamination that needs to be mitigated [10]. Figure 9.2 shows a massive MIMO system in the uplink and downlink for LOS propagation with the Base Station (BS) equipped with N_{Tx} antennas that serves P single-antenna terminals.

9.3.3 Device-Centric Architecture

Device-centric architectures provide a promising approach to meet the increasing demand for throughput that is required by applications in today's mobile devices, such as video streaming that requires at least 0.5 Mbps data rate. The uplink and downlink as well as control and data channels need to be reconsidered. In particular, the cell-centric architecture should evolve into a device-centric meaning that a given device should be able to exchange multiple information flows through different sets of heterogeneous nodes [5]. Figure 9.3 shows the cell-centric and device-centric networks where in the cell-centric network each user is communicating with the BS

[4] At a typical cellular frequency of 2 GHz; the wavelength is 15 cm and up to 400 dual-polarized antennas can thus be deployed in a 1.5 m × 1.5 m array.

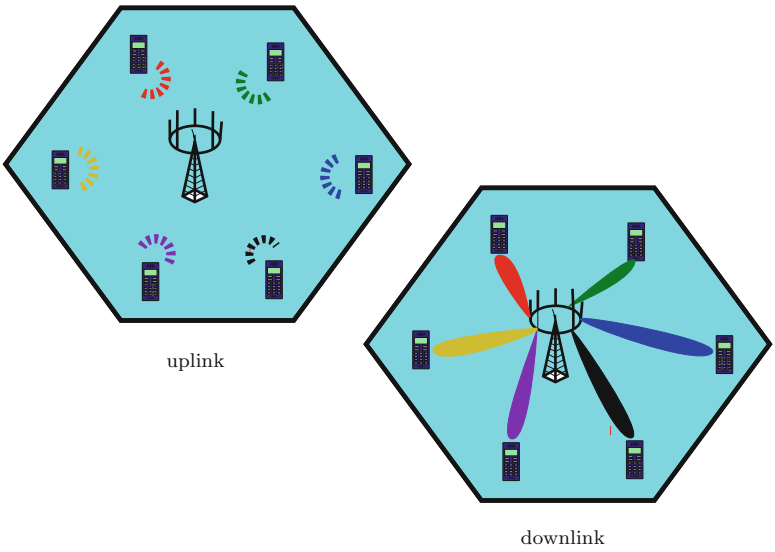

Fig. 9.2 Illustration of the massive MIMO system in the uplink and downlink for LOS propagation

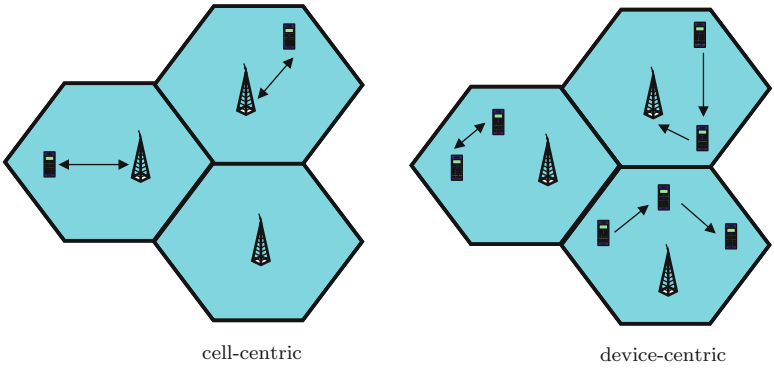

Fig. 9.3 Illustration of the cell-centric (*left*) and the device-centric network (*right*)

of the same cell directly while in the device-centric network each user can cooperate with the other users directly or act as a relay to other users for the communication with the BS or other users. More details on the different types of device-to-device communication are described later in this section.

Among the device centric and massive MIMO for 5G, mm-wave is a good candidate for localization due to higher bandwidth and smaller size of antenna arrays due to smaller wavelength. This enables highly directional links that are the key for the estimation of AOA/AOD for localization purposes. Consequently, the main focus is on the mm-wave localization for the rest of this chapter.

9.3.3.1 Device-Centric and Cell-Centric Positioning

In cell-centric localization, each agent communicates with multiple anchors.[5] This requires high density of the anchors and long-range transmissions. In device-centric localization, the agents can obtain the information from both anchors and agents. Consequently, high anchor density or long-range transmissions are no longer required [64]. If an agent cannot obtain its position based on distance estimates with respect to the anchors, device-centric localization is used to cooperatively obtain the position. This increases localization accuracy and coverage.[6]

9.3.4 D2D Communication

D2D communication can potentially reduce latency and power consumption and increase peak data rates. In the device-level 5G cellular network, each device communicate directly to another device or through the support of other devices. The BS either partially or fully controls resource, destination, and relaying devices or not have any control. Four types of device-level communications are briefly described [58].

9.3.4.1 Device Relaying Communication with Base Station Controlled Link

For a device located at the edge of a cell, the signal strength is poor and it is required to communicate with the BS by relaying the information through the other devices.

9.3.4.2 Direct Device-to-Device Communication with Base Station Controlled Link

In this architecture, the two devices are directly communicating with the links information provided by the BS.

9.3.4.3 Device Relaying Communication with Device Controlled Link

Both communication and links information are provided by the other devices and the BS is not involved in communication and link information.

[5] At least three anchors are required for 3D localization.

[6] The fraction of nodes with accurate location estimate is called coverage.

9.3.4.4 Direct Device-to-Device Communication with Device Controlled Link

Two devices are directly communicating and the link information is controlled by the devices.

Two of the most important challenges in D2D communication are security and interference management [9]. Security is important due to the fact that the routing of information is through the other users. Interference becomes important especially for the case of device relaying communication with device controller and direct device-to-device communication where the centralized methods cannot be employed.

9.3.5 Location-Aware Communications

5G networks can benefit from location information. In particular for the case of D2D resource allocation, D2D links share the same cellular resources that potentially interfere with each other. For instance, for the case of reusing the uplink resources, D2D transmissions interfere with cellular transmissions at the base station. To limit the interference, either maximum transmission power should be limited or D2D should not be allowed in the regions close to the base station. Consequently, position information of the user is of critical importance for resource allocation to ensure sufficiently large physical separation between the D2D and base station. To this end, distance and virtual sectoring based resource allocation techniques are proposed in [29]. Distance based resource allocation uses a pre-selected distance constraint to control interference between D2D and cellular nodes. Virtual sectoring based resource allocation relies on AOA measurements. In this approach, a D2D pair will reuse the radio resource that belongs to the vertically opposite sector based on the specified number of virtual sectors in the cell.

Another method to minimize the interference to the primary users in cognitive radio is the spatial spectrum sensing that can be adapted in 5G. In this approach, Gaussian processes (GPs) are used for predicting location-dependent channel qualities and provide statistical description of channel quality measurement in any location and any time. More specifically, the power from primary users can be estimated through secondary users resulting power density maps that allows resource allocation in the frequency bands that are not crowded [12, 36, 48, 57].

For vehicle-to-vehicle (V2V) networks, large-scale characteristics of the wireless channel (i.e., path-loss) may be captured using channel or position/motion measurements [13]. It has been shown that the feedback of position information to accomplish link adaptation is favorable compared to the overhead for the feedback of path-loss information. Particularly, for the case that path-loss changes rapidly.

9.3.6 Ultra Dense Networks

The throughput of a user in 5G networks is increased by the network densification. Densification in 5G networks is achieved in the spatial and frequency domain through the deployment of small cells and using additional spectrum (e.g., millimeter wave bands spanning from 30 to 300 GHz) [4]. Due to the use of small cells or cell splitting for spatial densification, path-loss is reduced while both desired and interfering signals are increased. Consequently, to translate densification into enhanced user experience backhaul densification is required together with space and frequency densification. Cloud radio access network (Cloud-RAN) architecture with coordinated multipoint processing (CoMP) where transmit/receive processing is centralized at a single processor transforms the systems into a near interference free system. Massive MIMO and mm-wave communication serve as the other candidates to improve capacity for wireless backhaul.

9.3.6.1 Mobility Management in Ultra Dense Networks

A moving node in a network or a group of such nodes form a moving network that can communicate with the other fixed or mobile nodes. This enhances the coverage for potentially large populations of jointly moving communication devices [18]. Tracking and predicting the device locations in the radio network is beneficial from various points of views. Location-aware communications may be considered as one of the advantages of predicting the user locations in the wireless network. The combination of radio environment maps and predicted user node locations are used within the network for proactive radio resource management (RRM). This leads to power consumption and load balancing at the moment and near future together with proactively allocating orthogonal radio resources in time and frequency [22]. Predicted user locations can be used for different applications including location data for self-driving cars, autonomous vehicles, and robots.

9.4 Mm-Wave Channels

In this section, we briefly describe mm-wave channels and the methods to estimate the channel parameters including AOA/AOD and TOA using the sparsity of the mm-wave channels. First, different path-loss models for mm-wave channels are described. Second, a double directional channel model is presented. Third, some estimation techniques are proposed. Finally, the sparsity of the mm-wave channel is applied for parameter estimation.

9.4.1 Path-Loss

First, a frequency dependent path-loss model is presented that is mainly used for communication purposes. Then, we present a path-loss model based on geometry statistics, which is more suitable for localization purposes.

9.4.1.1 Frequency Dependent Path-Loss

An important challenge in mm-wave frequencies is the high path-loss. The effect of path-loss in mm-wave systems is much higher than in wideband and ultra-wideband systems. This can be seen using the free space path-loss (FSPL) formula, defined as the path-loss for two isotropic antennas at the distance d, and is given by

$$\mathrm{FSPL}(d,f) = \left(\frac{\lambda}{4\pi d}\right)^2, \tag{9.1}$$

with $\lambda = c/f$ is the wavelength corresponding to a given frequency f (e.g., of the order of 60 GHz for the mm-wave frequencies). Since the FSPL is proportional to the squared frequency, the attenuation is much more severe than the UWB systems. In practice, the path-loss is expressed in dB relative to a reference power or transmitted power, and relative to the carrier frequency. Therefore, the path-loss in 28 and 38 GHz in urban microcells can be written as [33, 38, 60]

$$\mathrm{PL}(d,f) = \alpha + 10\bar{\beta}\log_{10}(d) + 20\gamma\log_{10}\left(\frac{f}{f_c}\right), \tag{9.2}$$

in which $f \in [-B/2, B/2]$ and B denotes the bandwidth (e.g., on the order of 6 GHz for the mm-wave frequencies), α is the path-loss at the carrier frequency f_c and at distance of 1 m, $\bar{\beta}$ is the average path-loss exponent, and 2γ is the exponent of the frequency dependency. As it is clear from (9.2), increasing the fractional bandwidth $\mathrm{FB} = f/f_c$ makes the effect of frequency dependent term more pronounced, see [33, 38, 60] for the typical values of path-loss for different parameters in (9.2).

9.4.1.2 Geometry Based Statistical Path-Loss

This approach is based on a cluster based channel model for the mm-wave frequencies. Each cluster is defined as the set of parameters including TOA, AOA/AOD, and complex channel gains with close values. We call the parameters in each cluster as intra-cluster parameters and the parameters from the different clusters as inter-cluster parameters. When a path arrives to the receiver, it has already gone through k_r reflections and k_d diffractions (object-intersections) [30, 31]. The probabilities

of encountering k_r reflections and k_d diffractions at distance d are described by the geometrical distribution of the environment and follow the Poisson distributions

$$p_r(k_r|d) = \frac{\exp(-\lambda_r d)(\lambda_r d)^{k_r}}{k_r!},$$

and

$$p_d(k_d|d) = \frac{\exp(-\lambda_d d)(\lambda_d d)^{k_d}}{k_d!},$$

where $1/\lambda_r$ and $1/\lambda_d$ are calibration parameters. They can be found similarly as done in [24], that is, as the mean distance a ray can travel before it intersects with an object that leads to reflection or diffraction. Using the above definitions, the shadow fading of the n-th cluster can be calculated as

$$\sigma^2_{\mathrm{SF}}(d_n) = \sum_{k_r} p_r(k_r|d_n)\sigma^2_{\mathrm{SF},r}(k_r) + \sum_{k_d} p_d(k_d|d_n)\sigma^2_{\mathrm{SF},d}(k_d), \tag{9.3}$$

where $\sigma^2_{\mathrm{SF},r}(k_r)$ and $\sigma^2_{\mathrm{SF},d}(k_d)$ are reflection and diffraction losses caused by k_r and k_d scattering events, respectively, $d_n = c\tau_n$ where $c = 3 \times 10^8$ m/s is the speed of light in free space and τ_n is the TOA of the first sub-path component within the n-th cluster, and $\tau_0 = d_{\mathrm{los}}/c$ where d_{los} is the LOS distance between the transmitter and receiver. Consequently, the path-loss of the n-th cluster with traveling distance d_n is

$$\rho_n = \sigma^2_{\mathrm{SF}}(d_n)\xi^2(d_n)\left(\frac{\lambda_c}{4\pi d_n}\right)^2, \tag{9.4}$$

where $\xi^2(d_n)$ is the atmospheric attenuation, and the last term denotes the free space path-loss. We use the path-loss model in (9.4) for localization purposes instead of (9.2) due to the geometrical relationships between the clusters, transmitter, and receiver contained in the formula. Following the inter-cluster path-loss ρ_n the intra-cluster path-loss can be obtained, see [11, 20, 54] for more details and typical values of the path-loss using geometry based statistical model.

Figure 9.4 shows the geometry based statistical model for two clusters with one reflection for the first cluster and two reflections for the second cluster, and total path lengths of $d_1 = d_{11} + d_{12}$ and $d_2 = d_{21} + d_{22} + d_{23}$, respectively.

9.4.2 Mm-Wave MIMO Channel Model

In a mm-wave MIMO system, channel parameters including AOA/AOD, channel gains, and TOA (i.e., the parameters that describe multipath components (MPCs)) are used for the localization purposes. A common approach for modeling the mm-wave MIMO channels is to group a set of rays with some close parameters in a

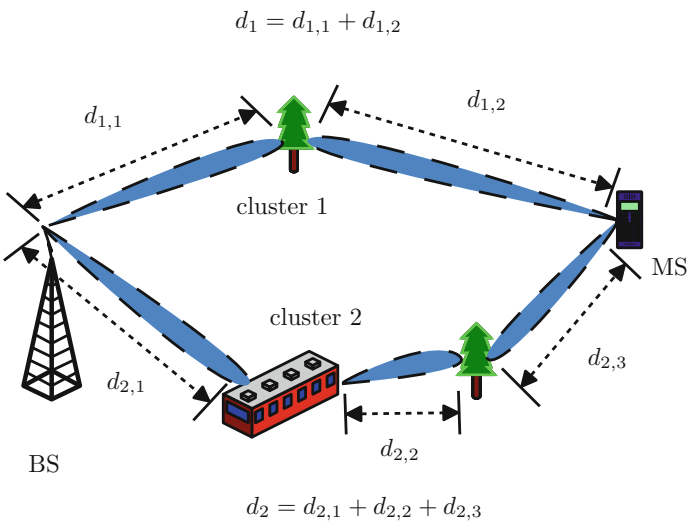

Fig. 9.4 Geometry based statistical path-loss model for mm-wave MIMO channel with two clusters and first and second order reflections

cluster. Consequently, the channel response between the receiver and the transmitter can be written as the sum of K specular MPCs and the LOS as [2, 19, 44]

$$\mathbf{H}(t,f) = \sum_{k=0}^{K} \rho_k \mathbf{B}_{\mathrm{Rx}}(f,\theta_{\mathrm{Rx},k}) \mathbf{X}_k \mathbf{B}_{\mathrm{Tx}}^{\mathrm{T}}(f,\theta_{\mathrm{Tx},k}) e^{-j2\pi f \tau_k} e^{j2\pi t \nu_k}, \tag{9.5}$$

where ρ_k was given in (9.4), $\mathbf{B}_{\mathrm{Tx}}(f,\theta_{\mathrm{Tx},k}) \in \mathbb{C}^{N_{\mathrm{Tx}}\times 2}$ and $\mathbf{B}_{\mathrm{Rx}}(f,\theta_{\mathrm{Rx},k}) \in \mathbb{C}^{N_{\mathrm{Rx}}\times 2}$ denote the complex beam pattern of the transmit array and the receive array with horizontal and vertical polarization, respectively, $\mathbf{X}_k \in \mathbb{C}^{2\times 2}$ contains the four polarimetric transmission coefficients for the k-th MPC, τ_k is the k-th TOA, and ν_k denotes the Doppler frequency for the k-th MPC. For the case of conventional (non-polarized) uniform linear arrays (ULAs), (9.5) turns into [19, 51, 59]

$$\mathbf{H}(t,f) = \sum_{k=0}^{K} \rho_k h_k \mathbf{a}_{\mathrm{Rx}}(f,\theta_{\mathrm{Rx},k}) \mathbf{a}_{\mathrm{Tx}}^{\mathrm{T}}(f,\theta_{\mathrm{Tx},k}) e^{-j2\pi f \tau_k} e^{j2\pi t \nu_k}, \tag{9.6}$$

where $\mathbf{a}_{\mathrm{Rx}}(f,\theta_{\mathrm{Rx},k}) \in \mathbb{C}^{N_{\mathrm{Tx}}\times 1}$ and $\mathbf{a}_{\mathrm{Tx}}(f,\theta_{\mathrm{Tx},k}) \in \mathbb{C}^{N_{\mathrm{Rx}}\times 1}$ denote the array steering vectors for the ULAs, and h_k is the complex channel gain for the k-th cluster. The effect of polarization is not considered in the channel modeling. This is due to the fact that polarization has no significant effect in NLOS, while in the LOS it can double the spectral efficiency. In general, more research on the effect of polarization in the mm-wave channels is required.

9.4.3 Parameter Estimation

Some of the typical algorithms used in channel estimation in 5G are:

1. Space-alternating generalized expectation maximization (SAGE).
2. Joint and iterative maximum likelihood estimation in [45], named RIMAX.

In the parameter estimation of the MPC, it is usually assumed the impulse response in (9.5) or (9.6) to consist of specular scattering.

9.4.3.1 SAGE

Particularly, SAGE algorithm (that is an algorithm based on expectation maximization and successively cancels interference) uses this assumption [15]. The SAGE algorithm jointly estimates MPC parameters, i.e., AOA/AOD, channel gains, Doppler shifts, and TOAs.

9.4.3.2 RIMAX

In addition to specular scattering, considering diffuse scattering improves the parameter estimation. RIMAX is an estimation method that considers the diffuse scattering in addition to specular scattering to improve the estimation of parameters. Moreover, an extended Kalman filter can be used for tracking the parameters in a sequential way for the case of time-varying channels.

The SAGE method can be considered as the preferred algorithm to estimate the MPC parameters due to the fact that in the mm-wave frequencies most of the power can be attributed to specular components. In the SAGE algorithm, the channel response in (9.5) or (9.6) consists of the superposition of $K + 1$ plane waves where K is the number of MPCs with specular scattering property, and the index 0 denotes the LOS path, which is omitted for the obstructed-line-of-sight (OLOS) scenario.

9.4.4 Sparsity

The propagation environment has a different effect for the mm-wave channels due to smaller wavelength. Due to the reduced fresnel zone diffraction is lower while penetration losses can be much larger. Few and clustered paths lead to channel sparsity unlike UWB channels [6, 35]. Consequently, the channel is "sparse" in the angular and time domains. The sparsity of the mm-wave channels significantly simplifies the estimation of channel parameters including TOA, Doppler spread, and AOA/AOD, which are difficult to estimate based on the physical channel model in (9.6) due to the non-linear dependence of $\{\theta_{\mathrm{Tx},k}, \theta_{\mathrm{Rx},k}, \tau_k, \nu_k\}$ on $\mathbf{H}(t,f)$. To this

end, we use a virtual representation for the physical channel model in (9.6) for the estimation of channel parameters by sampling in time, frequency, and space with the aid of a 4D Fourier transform [3, 27, 50]. For indoor applications, we will encounter a very slowly varying channel due to the fact that we are dealing with small values for the velocity and considering high data rate radios operating at 60 GHz. Consequently, time variation or Doppler spread is not a serious problem and can be neglected for indoor localization. Hence, we can eliminate the Doppler spread in (9.6) and write the virtual representation as

$$\mathbf{H}(f) = \sum_{l=0}^{L-1} \mathbf{A}_{\mathrm{Rx}} \mathbf{H}_{\mathrm{v}}^{\mathrm{T}}(l) \mathbf{A}_{\mathrm{Tx}}^{\mathrm{H}} e^{-j2\pi \frac{l}{B} f}, \tag{9.7}$$

where $L = \lceil B\tau_{\max} \rceil + 1$ denotes the maximum number of resolvable delays and $\tau_{\max}$ is the delay spread, $\mathbf{A}_{\mathrm{Rx}}$ and $\mathbf{A}_{\mathrm{Tx}}$ are $N_{\mathrm{Rx}} \times N_{\mathrm{Rx}}$ and $N_{\mathrm{Tx}} \times N_{\mathrm{Tx}}$ unitary matrices comprising $\mathbf{a}_{\mathrm{Rx}}(m/N_{\mathrm{Rx}})$ and $\mathbf{a}_{\mathrm{Tx}}(m/N_{\mathrm{Tx}})$ as their columns, and the k-th element of $\mathbf{a}_{\mathrm{Rx}}(m/N_{\mathrm{Rx}})$ is equal to $1/\sqrt{N}_{\mathrm{Rx}} \exp(-j2\pi(k-1)m/N_{\mathrm{Rx}})$, and similarly for $\mathbf{a}_{\mathrm{Tx}}(m/N_{\mathrm{Tx}})$. Besides, $\mathbf{H}_{\mathrm{v}}(l) = [\mathbf{h}_{\mathrm{v},1}(l), \ldots, \mathbf{h}_{\mathrm{v},N_{\mathrm{Rx}}}(l)]$ is an $N_{\mathrm{Tx}} \times N_{\mathrm{Rx}}$ matrix with the ith column $\mathbf{h}_{\mathrm{v},i}(l)$ consists of the matrix of virtual channel coefficients $\{H_v(i,m,l)\}$ obtained as

$$H_v(i,m,l) = \sum_{k \in S_{\mathrm{Rx},i} \cap S_{\mathrm{Tx},m} \cap S_{\tau,l}} \frac{1}{B} \int_{\frac{-B}{2}}^{\frac{B}{2}} H_{b,i,m,k}(f) e^{j2\pi \frac{l}{B} f} \, \mathrm{d}f, \tag{9.8}$$

where

$$H_{b,i,m,k}(f) = \rho_k h_k D_{N_{\mathrm{Rx}}}(\widetilde{\theta}_{\mathrm{Rx},k}(f) - i\Delta\widetilde{\theta}_{\mathrm{Rx}}) D_{N_{\mathrm{Tx}}}(\widetilde{\theta}_{\mathrm{Tx},k}(f) - m\Delta\widetilde{\theta}_{\mathrm{Tx}}) e^{-j2\pi f \tau_k}, \tag{9.9}$$

in which $D_N(\theta) = \sin(\pi N\theta)/\sin(\pi\theta)$ is the Dirichlet sinc function, $\widetilde{\theta}_{\mathrm{Tx},k}(f) = (d/\lambda)\sin(\theta_{\mathrm{Tx},k})$ with $\lambda = c/(f + f_c)$ and $\widetilde{\theta}_{\mathrm{Rx},k}(f)$ is defined similarly by replacing the subscript Tx by Rx, $\Delta\widetilde{\theta}_{\mathrm{Tx}} = 1/N_{\mathrm{Tx}}$ and $\Delta\widetilde{\theta}_{\mathrm{Rx}} = 1/N_{\mathrm{Rx}}$ are the orthogonal beam spacings for the transmit and receive ULAs, and $S_{\mathrm{Rx},i}$, $S_{\mathrm{Tx},m}$, and $S_{\tau,l}$ are the subsets for partitioning the $K+1$ paths defined as

$$S_{\mathrm{Rx},i} = \{k : \widetilde{\theta}_{\mathrm{Rx},k}(f) \in (i/N_{\mathrm{Rx}} - \Delta\widetilde{\theta}_{\mathrm{Rx}}/2, i/N_{\mathrm{Rx}} + \Delta\widetilde{\theta}_{\mathrm{Rx}}/2]\}, \tag{9.10}$$

$$S_{\mathrm{Tx},m} = \{k : \widetilde{\theta}_{\mathrm{Tx},k}(f) \in (m/N_{\mathrm{Tx}} - \Delta\widetilde{\theta}_{\mathrm{Tx}}/2, m/N_{\mathrm{Tx}} + \Delta\widetilde{\theta}_{\mathrm{Tx}}/2]\}, \tag{9.11}$$

$$S_{\tau,l} = \{k : \tau_k \in (l/B - 1/2B, l/B + 1/2B]\}. \tag{9.12}$$

Note that for sufficiently small fractional bandwidth (e.g., smaller than 0.02) $\widetilde{\theta}_{\mathrm{Rx},k}(f)$ and $\widetilde{\theta}_{\mathrm{Tx},k}(f)$ are constant within the frequency band of interest and belong to the above intervals. However, when the fractional bandwidth is larger, the resolution in the estimation of AOA/AOD becomes worse as more intervals are required to

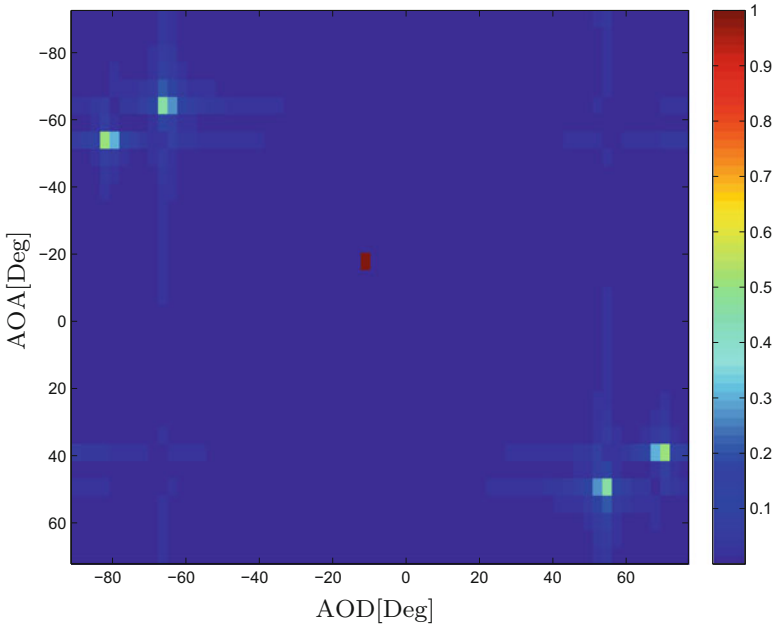

Fig. 9.5 Normalized magnitude of the virtual representation $\{H_v(i,m,l)\}$ in AOA/AOD subspace

include $\widetilde{\theta}_{\mathrm{Rx},k}(f)$ and $\widetilde{\theta}_{\mathrm{Tx},k}(f)$ within the bandwidth of interest. The estimated values of the AOA/AOD and TOA are obtained by finding the indices corresponding to the non-zero entries of $\{H_v(i,m,l)\}$ as

$$\{i^{\max}, m^{\max}, l^{\max}\} = \underset{i,m,l}{\operatorname{argmax}} \{H_v(i,m,l)\}. \tag{9.13}$$

Figures 9.5 and 9.6 show the normalized magnitude of the virtual representation $\{H_v(i,m,l)\}$ of the mm-wave channel for four different clusters and the LOS in the AOA/AOD and AOA/TOA planes. The values of the AOA/AOD and TOA are dominant only in the direction of clusters and the LOS. Using the sparsity of the mm-wave channels, a training based method can be used for the parameter estimation. Training based methods consist in sensing and reconstruction. Sensing corresponds to the design of training signals that are used by the transmitter to probe the channel, while reconstruction is to recover the channel in the receiver. The training based methods used in other rich scattering scenarios cannot be applied for the mm-wave channels due to the large number of antennas and bandwidth, what justifies the need to exploit the sparsity.

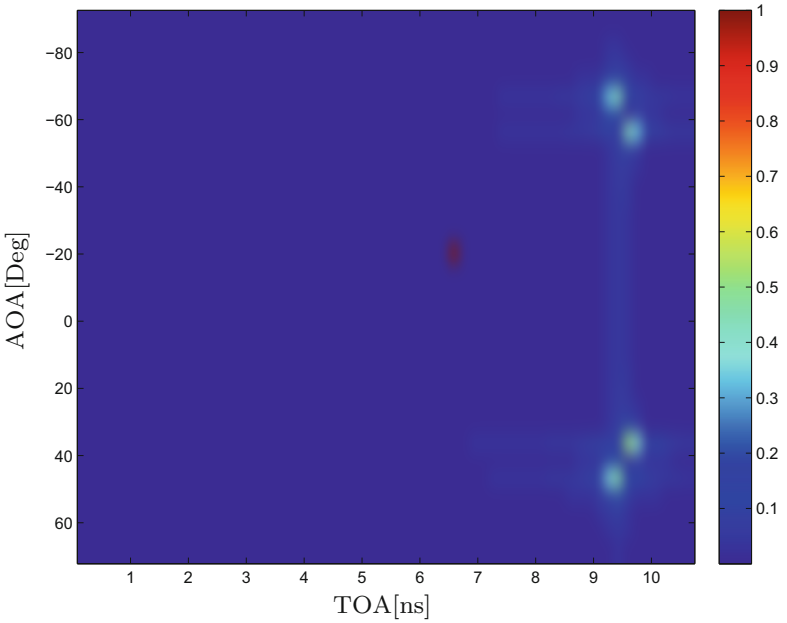

Fig. 9.6 Normalized magnitude of the virtual representation $\{H_v(i, m, l)\}$ in AOA/TOA subspace

9.5 Multi-Beam Transmission

To overcome the severe effect of path-loss frequencies[7] in mm-wave, one can increase the number of antenna elements to achieve beamforming gain. There exist some challenges in using a large number of antennas (from a few tens to hundreds of antennas) in the transmitter and receiver. One of the main challenges in using large number of antenna elements is to design beamformers that can generate narrow beams. In practice, analog beamformers using phase shifters suffer from the quantization error and fail to point the beam with sufficient accuracy [23, 41]. Moreover, digital beamformers in their conventional form require digital-to-analog-converter (DAC) for each antenna element in the transmitter and ADC for each antenna element in the receiver. Considering the large number of antenna elements in the transmitter and receiver, and the fact that DACs and ADCs consume a lot of power at mm-wave, one needs to use a more efficient way for beamforming. Moreover, multi-stream transmission using hybrid beamformers is required for both communication and localization purposes [1, 39, 68]. Particularly, it is critical to have more than one beam towards each user in order to make localization possible, as it will be explained in more detail later on.

[7] For a given distance, the FSPL at 60 GHz is 28 dB larger than 2.4 GHz.

In this section, first we review the hybrid beamformers as an important way for multi-beam transmission to obtain AOA/AOD that are used for localization purposes using the sparsity of the mm-wave MIMO channel in the beamspace. Second, a beam training protocol to find the strongest link between transmitter and receiver and consequently estimation of AOA/AOD as a key step for the localization is investigated.

9.5.1 Hybrid Beamformers

Hybrid beamformers are used to avoid the complexity in the implementation of the typical digital beamformers that require DAC for each antenna of the transmitter and ADC for each antenna of the receiver, i.e., N_{Tx} DACs in the transmitter and N_{Rx} ADCs in the receiver. Instead, hybrid beamformers use $M_t < N_{\mathrm{Tx}}$ and $M_r < N_{\mathrm{Rx}}$ DACs and ADCs in the transmitter and receiver, respectively, where M_t and M_r denote the number of transmit and received beams that are much smaller than the number of antenna elements. Moreover, they provide multi-beam transmission like digital beamformers but with less complexity. Especially, in the estimation of AOD and AOA in LOS conditions one needs to send more than one beam at each transmission as will be explained in the next section. Hybrid beamformers are comprised of a baseband digital pre-coder, DACs, radio frequency (RF) chains, and an RF analog pre-coder in the transmitter; and analog RF combiner, RF chains, ADCs, and a baseband digital combiner in the receiver. More details about hybrid beamformers in the lower frequencies can be found in [55, 65].

Figure 9.7 shows the MIMO architecture at mm-wave using a hybrid beamformer in which N_s data streams are fed to the baseband digital pre-coder, $M_t > N_s$ outputs of the baseband pre-coder are converted to analog and used to generate M_t beams

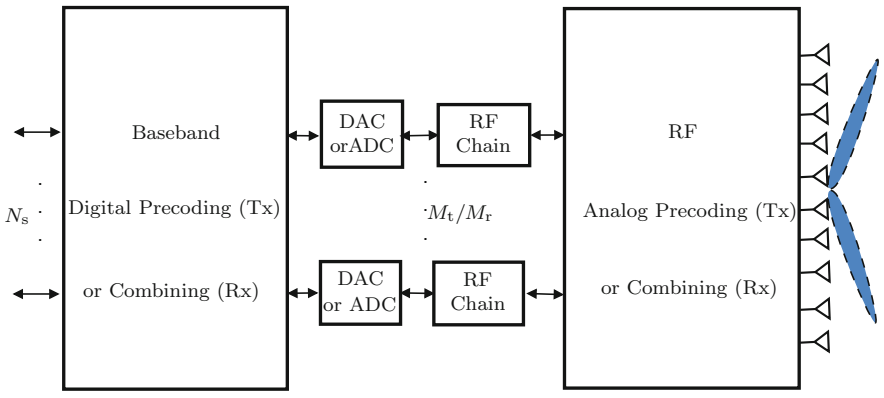

Fig. 9.7 MIMO architecture at mm-wave based on hybrid analog-digital precoding and combining

through RF chains that are connected to the antenna arrays by the RF analog pre-coder. On the receiver side, the received signals are fed to the RF analog combiner to capture the M_r beams, then the resulting signals are converted to digital and fed to the baseband digital combiner to reconstruct the N_s transmitted data streams.

Hybrid analog pre-coder/combiner can be implemented in two different ways using phase shifters and switches [21, 34, 40]. Although the lack of precision of analog shifter can be compensated in the digital pre-coder/combiner, using switches instead of analog phase shifters exploits the sparse nature of the mm-wave channel by implementing a compressed spatial sampling on the received signal and further reduces the complexity of the hybrid architecture using phase shifters.

9.5.2 Beam Training Protocols

The beam training protocol is a very important step in the AOA/AOD estimation and will be briefly explained in this section. The beam training protocol included in IEEE 802.11ad includes three major steps [26]:

- Sector Level Sweep (SLS): This stage is based on a coarse combination between the sector (at the transmitter side) and antenna (at the receiver side). The transmitter sends signals for each of its sectors, with a number of sectors up to 64 per antenna. After completing the sweep by the transmitter, the MS selects the best sector and sends feedback to the transmitter. At the end of this stage a coarse estimation of the AOD is obtained.
- Beam Refinement Protocol (BRP): In this stage, the coarse estimation of the AOD will be refined by sending the orthogonal beams within the optimal sector found from the previous stage. The receiver sends feedback to the transmitter regarding the success of the new beam. At the end of this stage a refined estimation of the AOD is obtained.
- Beam tracking: This stage includes a periodic refinement over a small number of antenna configurations.

Beam training protocols can be generalized for hybrid precoding rather than only for analog beamformers. The main advantage is the capability to steer the beam with more accuracy than using only phase shifters, thanks to the compensation of the error in analog part using the digital pre-coder. This approach starts with the coarse search for the best AOA/AOD and channel gains (SLS step) and refines the estimated values (BRP step) in the final stages using a novel multi-resolution beamforming codebook.

Figure 9.8 illustrates the beam training protocol as an important strategy to find the best link between the BS and the mobile station (MS). Particularly, when one link is not strong enough or is blocked and cannot be used to estimate the channel parameters (i.e., AOA/AOD, delay, and channel gain), using first the beam training protocol, we can obtain the sector that provides the LOS conditions, and then we use

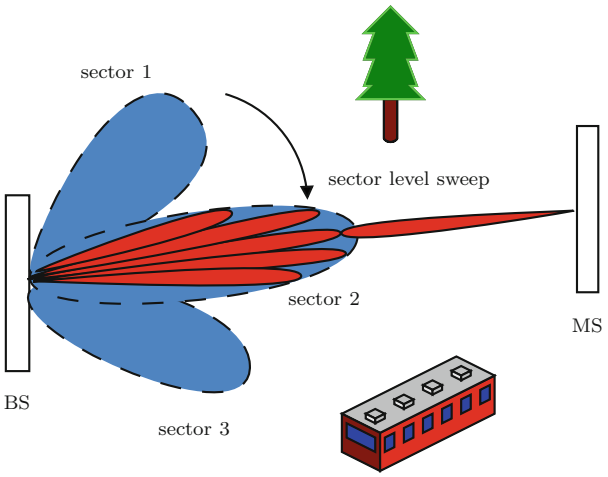

Fig. 9.8 Finding the optimal sector and beam for the localization of the MS by the SLS and BRP

the LOS link[8] to localize the MS using the AOA/AOD and delay estimates, as will be discussed below. In what follows, we provide an overview on the localization of the MS using AOA/AOD, and TOA in a mm-wave MIMO system.

9.6 Localization Based on Delay, AOA, and AOD

In this section, we describe some localization techniques that can be used for 5G systems. First we briefly describe common localization approaches using range measurements, range-difference measurements, triangulation, and fingerprinting. Then, the localization techniques based on the combination of range and angle information are described for 5G systems. We consider 2D localization in this section for simplicity.[9] Nevertheless, the methods can be easily extended to 3D localization.

9.6.1 General Localization Techniques

9.6.1.1 Localization Using Range Measurements

Localization using range measurements is a technique for localization based on the received signal strength (RSS) or the TOA from the BSs to the MS [42]. If the BSs

[8] Although, it is possible to use the information from the NLOS link for localization as will be discussed in the next section.

[9] The reader is referred to [17, 52] for more details.

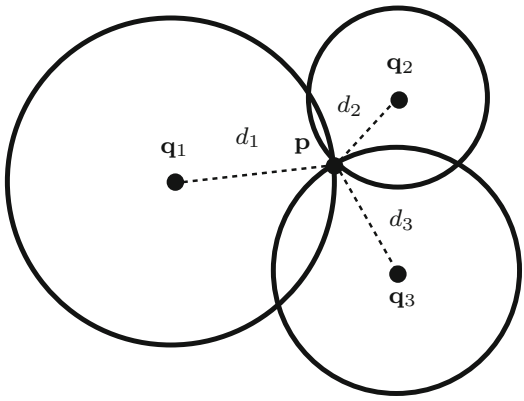

Fig. 9.9 Range measurement approach for the 2D localization of the MS with at least 3 BSs

are located at the positions $\mathbf{q}_i = [q_{i,x}, q_{i,y}]^{\mathrm{T}}$, and the MS is located at the unknown position of $\mathbf{p} = [p_x, p_y]^{\mathrm{T}}$ with the distance $d_i = c\tau_i$ from the BS, we obtain the following geometrical relation:

$$(q_{i,x} - p_x)^2 + (q_{i,y} - p_y)^2 = d_i^2. \tag{9.14}$$

Putting together expressions like (9.14) corresponding to M anchors, we obtain a systems of equation that can be solved for the position of the MS, $\mathbf{p}$. Figure 9.9 depicts the range measurement approach using at least 3 BSs for 2D localization.

9.6.1.2 Localization Using Range-Difference Measurements

Range-difference measurement is a localization technique based on time difference of arrival (TDOA) of the signals transmitted by the different BSs [67]. The time/distance difference between BSs $i \neq j$ can be computed as

$$d_i - d_j = \Delta d_{i,j}, \tag{9.15}$$

with d_i obtained from (9.14). The above equation is the mathematical representation of a hyperbola. For different pairs of BSs, (9.15) forms a system of equations whose solution is the location of the MS, $\mathbf{p}$. An example for TDOA technique are long-term evolution (LTE) networks. Unlike TOA based localization, in the localization using TDOA there is no need for synchronization between transmitters and receivers. Figure 9.10 shows 2D localization based on TDOA with at least 3 BSs.

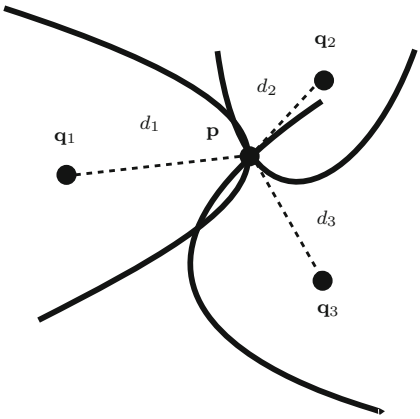

Fig. 9.10 Range-difference measurement approach for the 2-D localization of the MS with at least 3 BSs

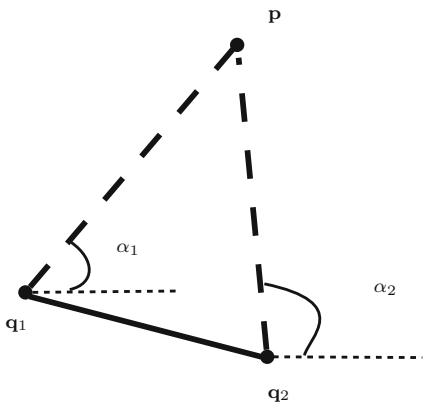

Fig. 9.11 Triangulation approach for the 2-D localization of the MS

9.6.1.3 Triangulation

Triangulation is a localization technique based on angle measurements [32]. This technique usually requires the use of antenna arrays in the transmitter and receiver to measure the AOA and AOD. Estimation algorithms such as multiple signal classification (MUSIC) and estimation of signal parameters by rotational invariance techniques (ESPRIT) can be applied to estimate the AOA and AOD [25, 37]. Localization of the MS is possible if the AODs from two BSs are known (in the downlink) or the AOAs at the two BSs are known (in the uplink). But, if the MS only knows the AOA from two BS transmissions and the rotation of the MS is unknown, then positioning is not possible, and more AOA measurements are needed. Moreover, this method fails for the localization of the MS if the MS is aligned with the BSs as no triangle can be formed in this case. Figure 9.11 demonstrates the triangulation approach for the 2D localization of the MS with two BSs.

9.6.1.4 Fingerprinting

Fingerprinting approach is based on the fact that radio waves emitted from the BSs leave a unique radio fingerprint at a given location that can be used for localization [7]. This requires a training phase to collect the fingerprints at known locations that later on can be used for the localization of the MS based on probabilistic or deterministic positioning techniques, e.g., maximum likelihood estimator or k-nearest-neighbor (kNN).

9.6.2 Mm-Wave Localization Techniques

From the above discussion, we interpret that all the aforementioned methods either use the information from angles, delays, or RSS. However, one may envision that both angles and delays can be used for the localization at the same time. Particularly, large number of antenna elements in the transmitter and receiver in the 5G systems provides steerable narrow beams that can be used for localization with AOA/AOD and TOA.

Figure 9.12 shows the LOS link for the localization of the MS using joint angle and delay measurements. The TOA provides a circle with the radius of d_0 from the MS centered in $\mathbf{q}$, AOD and AOA provide lines that eventually lead to the localization of the MS as shown in Fig. 9.13. This can be simply expressed as

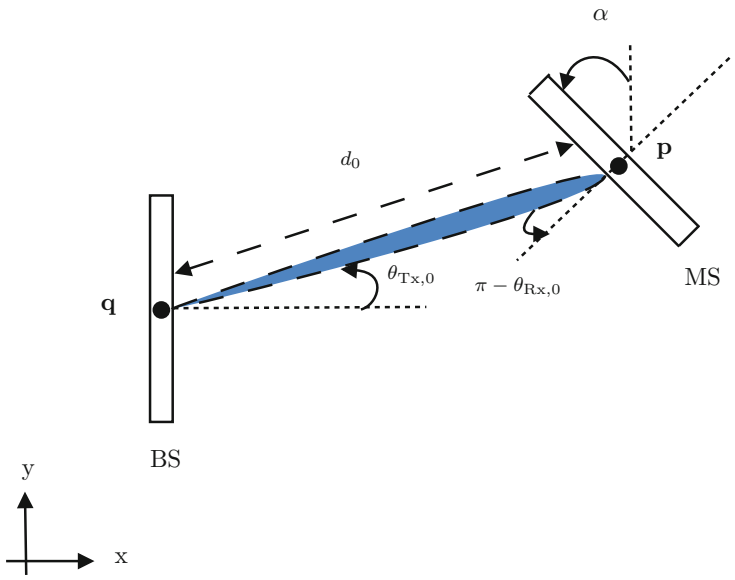

Fig. 9.12 LOS link for the localization based on joint AOA/AOD and TOA estimation

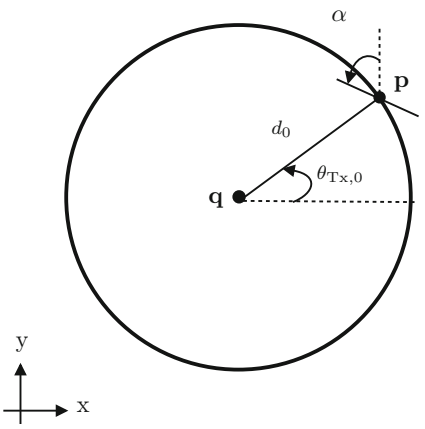

Fig. 9.13 Demonstration of the localization in the LOS with TOA and AOA/AOD

$$(q_x - p_x)^2 + (q_y - p_y)^2 = d_0^2, \tag{9.16}$$

and

$$\tan(\theta_{\mathrm{Tx},0}) = \frac{p_y - q_y}{p_x - q_x}. \tag{9.17}$$

Solving (9.16) and (9.17) leads to

$$\mathbf{p} = \mathbf{q} + d_0 \mathbf{u}(\theta_{\mathrm{Tx},0}),$$

where $\mathbf{u}(\theta_{\mathrm{Tx},0}) = [\cos(\theta_{\mathrm{Tx},0}), \sin(\theta_{\mathrm{Tx},0})]^{\mathrm{T}}$. Moreover, the orientation is obtained as $\alpha = \pi + \theta_{\mathrm{Tx},0} - \theta_{\mathrm{Rx},0}$. Figure 9.14 shows the LOS link in the presence of clusters. In this case, the presence of clusters reduces the localization accuracy depending on the location of the clusters towards the LOS link as will be shown in the simulation results. Moreover, the orientation is *only* estimated through the LOS link and the clusters do not provide any information on the orientation of the MS.

Figure 9.15 demonstrates the use of NLOS links[10] for the localization of the MS using joint angle and delay measurements and a given orientation α_0. In this case the location of the MS can be obtained using the following equations:

$$\|\mathbf{p} - \mathbf{s}_1\| + \|\mathbf{q} - \mathbf{s}_1\| = d_1, \tag{9.18}$$

$$\|\mathbf{p} - \mathbf{s}_2\| + \|\mathbf{q} - \mathbf{s}_2\| = d_2, \tag{9.19}$$

$$\mathbf{s}_1 = \mathbf{q} + d_{1,1}\mathbf{u}(\theta_{\mathrm{Tx},1}), \tag{9.20}$$

$$\mathbf{s}_2 = \mathbf{q} + d_{2,1}\mathbf{u}(\theta_{\mathrm{Tx},2}), \tag{9.21}$$

[10]This can also be considered as the blocked LOS as in the mm-wave frequencies blockage happens quite often especially for indoor localization.

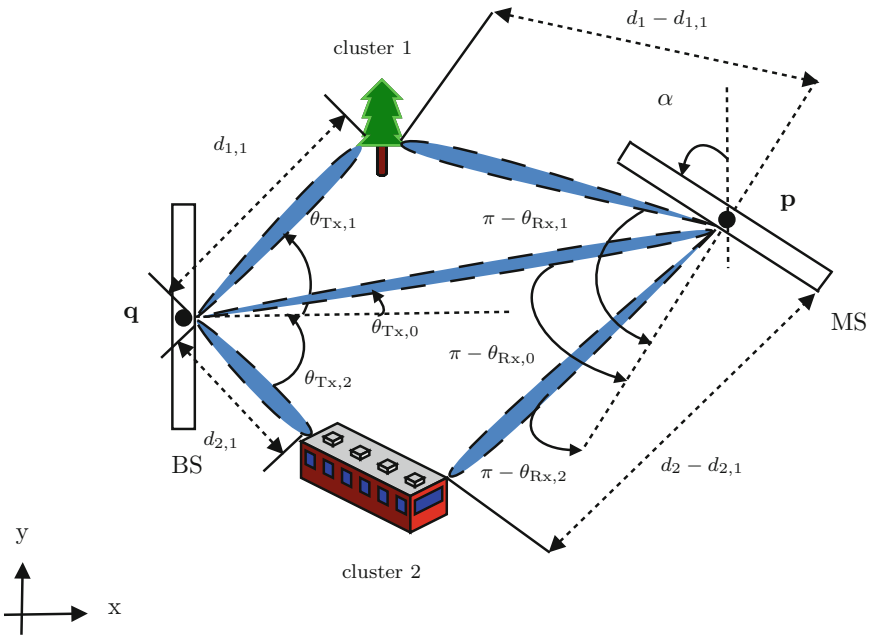

Fig. 9.14 LOS in the presence of clusters for the localization based on joint AOA/AOD and TOA estimation

$$\tan(\pi - (\theta_{\mathrm{Rx},1} + \alpha_0)) = \frac{s_{1,y} - p_y}{p_x - s_{1,x}}, \tag{9.22}$$

$$\tan(\pi - (\theta_{\mathrm{Rx},2} + \alpha_0)) = \frac{-s_{2,y} + p_y}{p_x - s_{2,x}}, \tag{9.23}$$

where $\mathbf{q}$ is known, $\{\theta_{\mathrm{Tx},k}, \theta_{\mathrm{Rx},k}, d_k\}$ denotes the set of estimated parameters that are assumed to be known, $\mathbf{p}$ is the unknown location of the MS, and $\{\mathbf{s}_k, d_{k,1}\}$ denotes the set of unknown parameters including the location of the k-th cluster $\mathbf{s}_k$ and the distance between the k-th cluster and the BS $d_{k,1}$. Considering 2-D localization, there are 8 unknown parameters that can be obtained by the above set of equations. The TOA from two clusters provides the intersection from two circles as shown in Fig. 9.16, while the AOAs provide the lines for the localization of the MS.

9.7 Simulation Results

In this section, we analyze the performance of the mm-wave localization techniques by means of numerical simulations. Performance is measured in terms of the position error bound (PEB, expressed in meters) and the rotation error bound

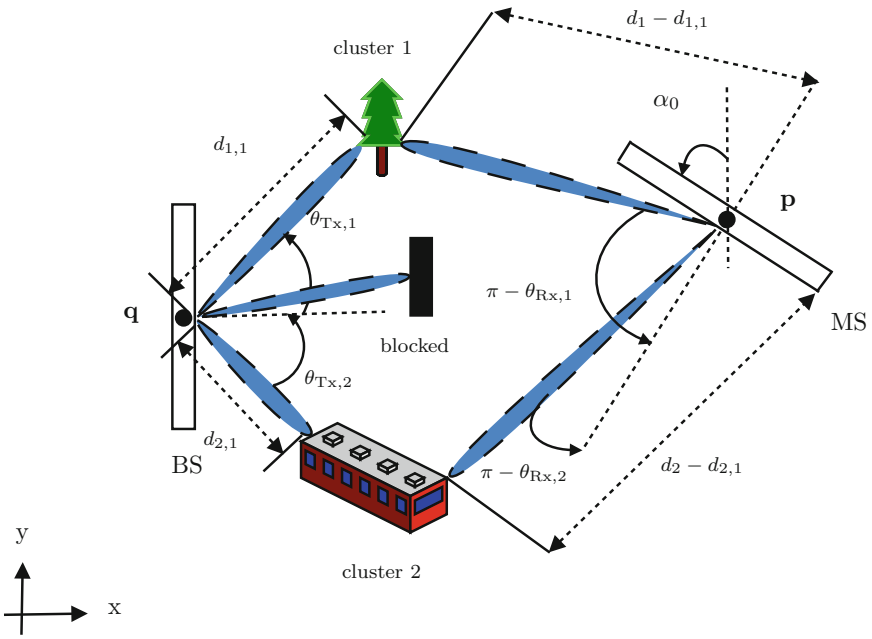

Fig. 9.15 NLOS link for the localization based on joint AOA/AOD and TOA estimation

(REB, expressed in radians), where PEB $\triangleq \sqrt{\mathrm{tr}\left\{\left[\mathbf{J}_{\boldsymbol{\eta}}^{-1}\right]_{1:2,1:2}\right\}}$ and REB $\triangleq \sqrt{\mathrm{tr}\left\{\left[\mathbf{J}_{\boldsymbol{\eta}}^{-1}\right]_{3,3}\right\}}$ with $\mathbf{J}_{\boldsymbol{\eta}}$ being the Fisher information matrix (FIM) of the unknown parameter $\boldsymbol{\eta} \triangleq [\mathbf{p}^{\mathrm{T}}, \alpha, \boldsymbol{\kappa}^{\mathrm{T}}]^{\mathrm{T}}$. For the case of LOS, the parameter $\boldsymbol{\kappa}$ that is related to the clusters (i.e., for the l-th cluster $\boldsymbol{\kappa}_l = [\tau_l, \boldsymbol{\theta}_l^{\mathrm{T}}, \mathbf{h}_l^{\mathrm{T}}]^{\mathrm{T}}$ where $\boldsymbol{\theta}_l = [\theta_{\mathrm{Tx},l}, \theta_{\mathrm{Rx},l}]$ and $\mathbf{h}_l = [\mathrm{Re}\{h_l\}, \mathrm{Im}\{h_l\}]^{\mathrm{T}}$) is set to zero. For the case of NLOS, we assume that $\alpha = \alpha_0$ and consequently it disappears from the unknown parameter $\boldsymbol{\eta}$. We compute the PEB for different locations of the MS for the BS located at a fixed position. The comparison between PEB for the case of LOS and in the presence of the clusters is provided. Finally, the performance of the NLOS link with two clusters is shown for a given orientation for different locations of the MS and the BS located at a fixed position.

9.7.1 Simulation Setup

We set $f_c = 60\,\mathrm{GHz}$, $B = 600\,\mathrm{MHz}$, and $N_0 = 2\mathrm{W/GHz}$. The inter-element spacing is assumed to be $d = \lambda_c/2$. The number of transmit and receive antennas for the non-polarized ULAs are set to $N_{\mathrm{Tx}} = 64$ and $N_{\mathrm{Rx}} = 8$.

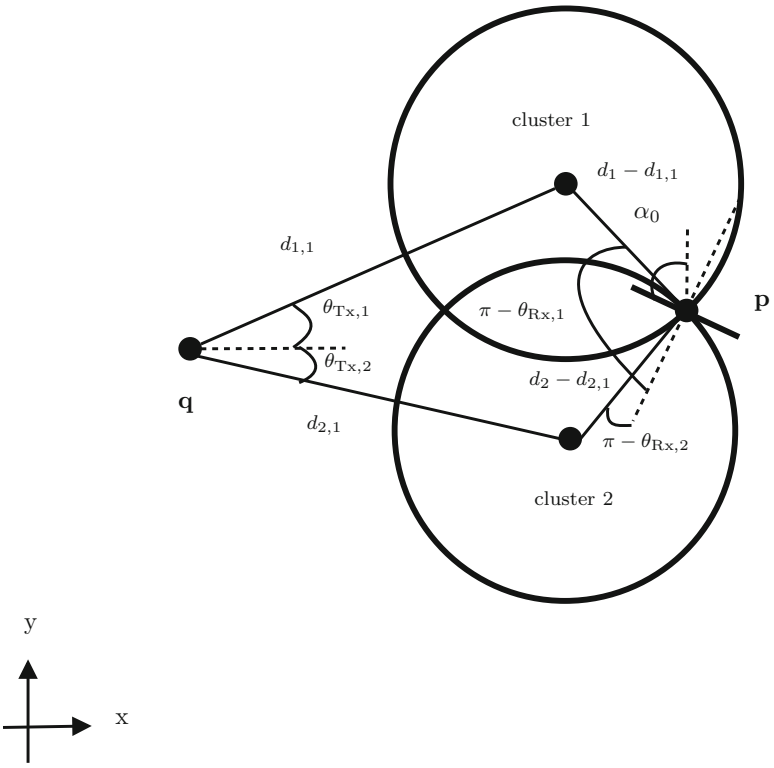

Fig. 9.16 Demonstration of the localization in the NLOS with TOA and AOA/AOD

9.7.2 Results and Discussion

Figures 9.17 and 9.18 show the AOD, AOA, and TOA in the LOS and the resulting values of the PEB and REB for different locations of the MS. It is observed that in the directions of the beams, the AOA/AOD and TOA together with the PEB and REB have the lowest values, while in the other locations the values are much higher. More specifically, for a distance of $\|\mathbf{q} - \mathbf{p}\| = 0.5\,\mathrm{m}$, the PEB in the directions of the beams is approximately 5 cm while the highest PEB is approximately 3 m. The REB values range from 0.01 rad in the direction $\pi/3$, over 0.02 rad in the direction $-\pi/3$, up to 0.2 rad outside any of the beams. We observe the impact of the extra beam in the direction of $\pi/3$ that provides increased SNR, leading to better TOA information (in the Fisher sense) and good information regarding AOA/AOD that leads to reducing the PEB and REB. On the other hand, the single beam that is transmitted in the direction of $-\pi/3$ provides good TOA and AOA information but leads to poor AOD information (except for MS locations close to the BS). Hence, the PEB and REB in the direction of $-\pi/3$ are higher.

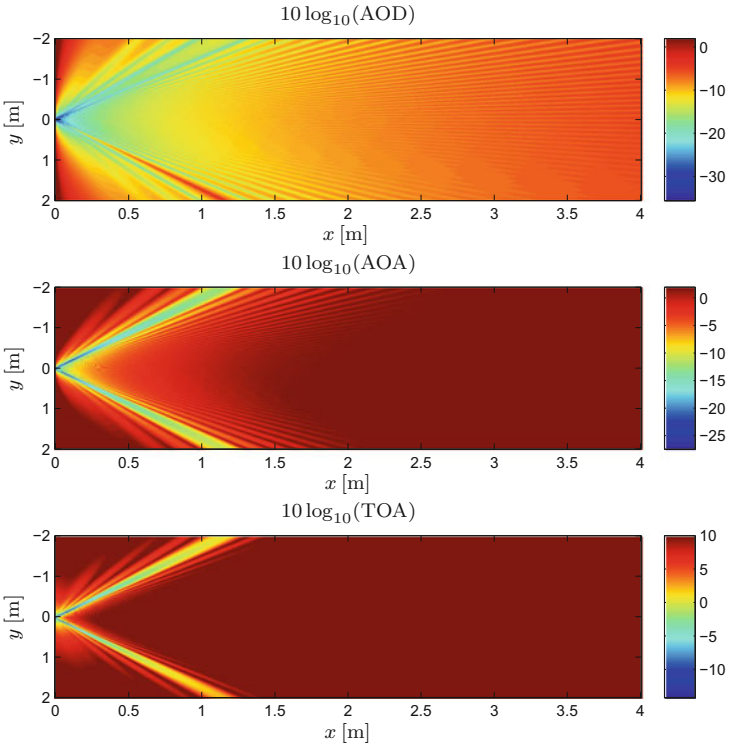

Fig. 9.17 Performances of the AOD, AOA, and TOA in the LOS for different locations of the MS and the BS located at $\mathbf{q}$ [m] $= [0, 0]^{\mathrm{T}}$, for a scenario with three beams in the directions $[\pm\pi/3, \pi/3 + 0.01]$

In the presence of LOS, the effect of NLOS paths on position estimation is shown in Fig. 9.19. It can be observed that adding the clusters sequentially reduces the position information (in the Fisher sense) that leads to higher values of the PEB. Moreover, sufficiently good localization accuracy can be obtained even at low SNR.

Figure 9.20 compares the PEB for the LOS in presence of clusters and NLOS conditions. It can be observed that the PEB for the NLOS is much higher than the LOS condition (around 35 dB at snr $\approx$ 1 dB). Moreover, adding the clusters make the performance worse due to reducing the position information in the Fisher sense.

9.8 Conclusions

Among 5G candidate technologies, mm-wave provides promising solutions for localization, thanks to the large bandwidth and highly directional links made possible by the small wavelength at mm-wave frequencies. The effect of path-loss

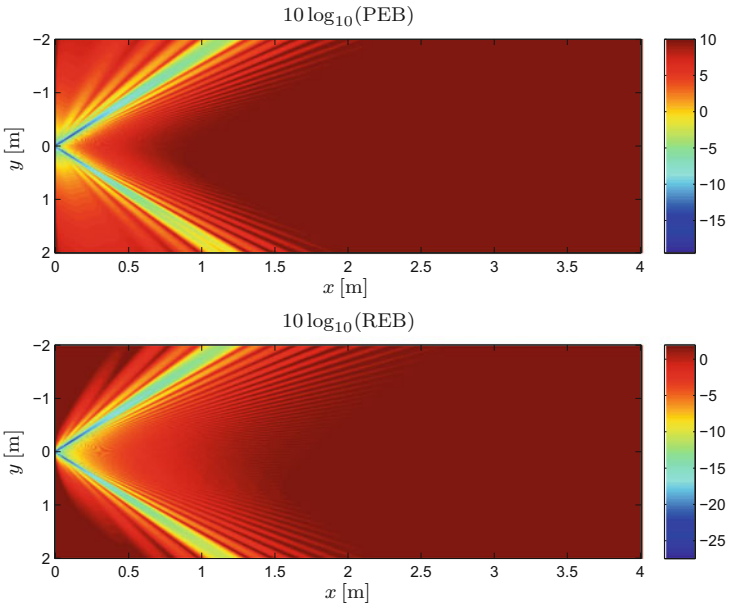

Fig. 9.18 Performances of the PEB (*top*) and REB (*bottom*) in the LOS for different locations of the MS and the BS located at $\mathbf{q}$ [m] $= [0, 0]^{\mathrm{T}}$, for a scenario with three beams in the directions $[\pm\pi/3, \pi/3 + 0.01]$

in mm-wave frequencies can be compensated by using the antenna arrays in the transmitter and receiver. Hybrid beamformers and beam training protocols provide powerful tools for AOA/AOD estimation that can be used for localization of the MS. Specifically, the NLOS links provide valuable information for the localization of the MS as beam tracking protocols lead to finding the strongest NLOS link and estimation of the AOA/AOD and TOA. The localization of the MS is based on the geometry of the environment and a geometrical statistical path-loss model. In general, the sparsity of the mm-wave wave channel is the key for estimation of the channel parameters especially in the NLOS conditions. Finally, the accuracy in terms of PEB was proposed by exploiting the delay and angle information for LOS and NLOS conditions in the simulation results.

Acknowledgements This work was financially supported by EU FP7 Marie Curie Initial Training Network MULTI-POS (Multi-technology Positioning Professionals) under grant nr. 316528.

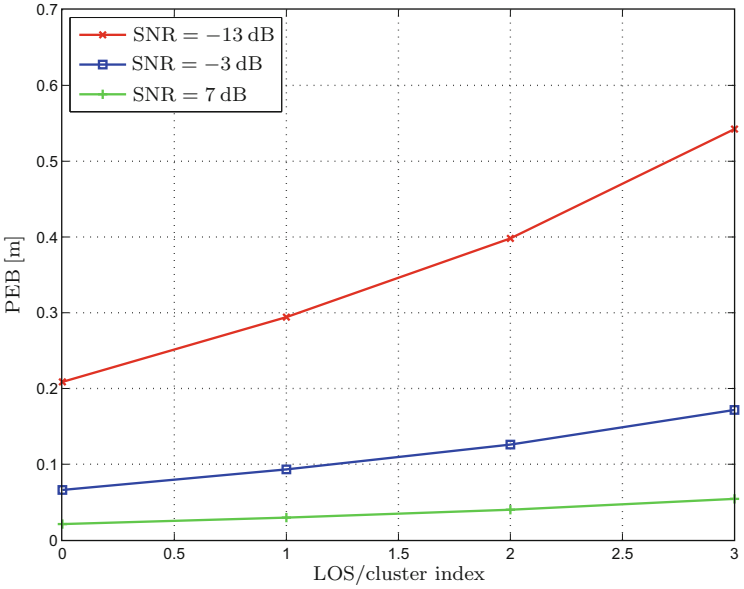

Fig. 9.19 The effect of adding NLOS links on the PEB for the case that the beams are transmitted in the LOS and in the directions of the clusters. The clusters are added sequentially by moving away from the LOS link as $\mathbf{s}_k$ [m] $= [1.5, 1 + (k-1) \times 0.1]^{\mathrm{T}}$ for $k = 1, \ldots, 3$, $\mathbf{q}$ [m] $= [0, 0]^{\mathrm{T}}$, and $\mathbf{p}$ [m] $= [0, 4]^{\mathrm{T}}$. The index $k = 0$ denotes the LOS condition

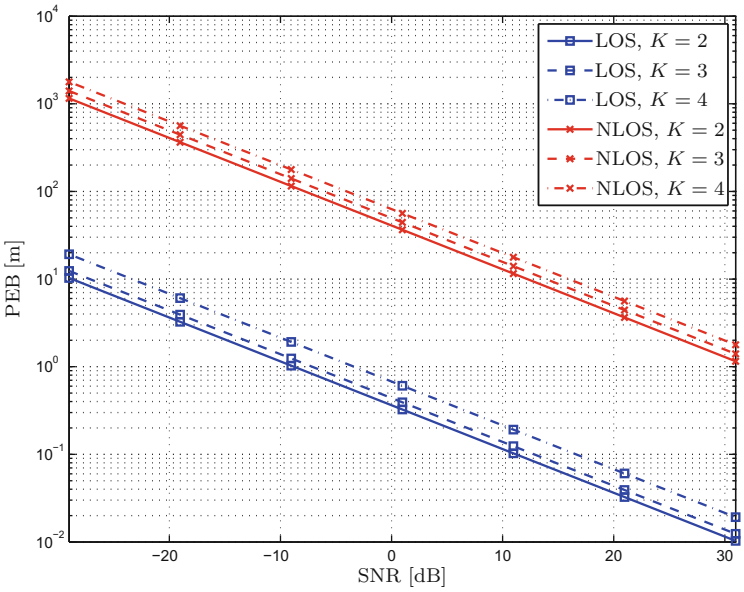

Fig. 9.20 Performances of the PEB for the LOS in presence of clusters and NLOS conditions with different number of clusters $\mathbf{s}_k$ [m] $= [1.5, 1 + (k-1) \times 0.1]^{\mathrm{T}}$ for $k = 2, \ldots, 4$, $\mathbf{q}$ [m] $= [0, 0]^{\mathrm{T}}$, and $\mathbf{p}$ [m] $= [0, 4]^{\mathrm{T}}$

References

1. A. Alkhateeb et al., Channel estimation and hybrid precoding for millimeter wave cellular systems. IEEE J. Sel. Top. Sign. Process. **8**(5), 831–846 (2014)
2. P. Almers et al., Survey of channel and radio propagation models for wireless MIMO systems. EURASIP J. Wirel. Commun. Netw. **2007**(1), 56–56 (2007)
3. P. Bello, Characterization of randomly time-variant linear channels. IEEE Trans. Commun. **11**(4), 360–393 (1963)
4. N. Bhushan et al., Network densification: the dominant theme for wireless evolution into 5G. IEEE Commun. Mag. **52**(2), 82–89 (2014)
5. F. Boccardi et al., Five disruptive technology directions for 5G. IEEE Commun. Mag. **52**(2), 74–80 (2014)
6. J. Brady, N. Behdad, A. Sayeed, Beamspace MIMO for millimeter-wave communications: system architecture, modeling, analysis, and measurements. IEEE Trans. Antennas Propag. **61**(7), 3814–3827 (2013)
7. M. Bshara et al., Fingerprinting localization in wireless networks based on received-signal-strength measurements: a case study on WiMAX networks. IEEE Trans. Veh. Technol. **59**(1), 283–294 (2010)
8. G. Caire, S. Shamai, On the achievable throughput of a multiantenna Gaussian broadcast channel. IEEE Trans. Inf. Theory **49**(7), 1691–1706 (2003)
9. I. Cha et al., Trust in M2M communication. IEEE Veh. Technol. Mag. **4**(3), 69–75 (2009)
10. Z. Chen, C. Yang, Pilot decontamination in massive MIMO systems: exploiting channel sparsity with pilot assignment, in *IEEE Global Conference on Signal and Information Processing (GlobalSIP)* (2014)
11. C.C. Chong et al., A new statistical wideband spatio-temporal channel model for 5-GHz band WLAN systems. IEEE J. Sel. Areas Commun. **21**(2), 139–150 (2003)
12. A. Dammann et al., WHERE2 location aided communications, in *European Wireless Conference* (2013)
13. R.C. Daniels, R.W. Heath Jr., Link adaptation with position/motion information in vehicle-to-vehicle networks. IEEE Trans. Wirel. Commun. **11**(2), 505–509 (2012)
14. H. Deng, A. Sayeed, Mm-wave MIMO channel modeling and user localization using sparse beamspace signatures, in *International Workshop on Signal Processing Advances in Wireless Communications*, pp. 130–134 (2014)
15. B.H. Fleury et al., Channel parameter estimation in mobile radio environments using the SAGE algorithm. IEEE J. Sel. Areas Commun. **17**(3), 434–450 (1999)
16. S. Folea et al., Indoor localization based on Wi-Fi parameters influence, in *International Conference on Telecommunications and Signal Processing (TSP)*, pp. 190–194 (2013)
17. A. Guerra, F. Guidi, D. Dardari, Position and orientation error bound for wideband massive antenna arrays, in *ICC Workshop on Advances in Network Localization and Navigation* (2015)
18. A. Gupta, R.K. Jha, A survey of 5G network: architecture and emerging technologies. IEEE Access **3**, 1206–1232 (2015)
19. C. Gustafson, 60 GHz wireless propagation channels: characterization, modeling and evaluation. Ph.D. thesis, Lund University, 2014
20. G. Gustafson et al., On mm-wave multipath clustering and channel modeling. IEEE Trans. Antennas Propag. **62**(3), 1445–1455 (2014)
21. A. Hajimiri et al., Integrated phased array systems in silicon. Proc. IEEE **93**(9), 1637–1655 (2005)
22. A. Hakkarainen et al., High-efficiency device localization in 5G ultra-dense networks: prospects and enabling technologies, in *IEEE Vehicular Technology Conference (VTC)* (2015)
23. S. Han et al., Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G. IEEE Commun. Mag. **53**(1), 186–194 (2015)
24. M. Hassan-Ali, K. Pahlavan, A new statistical model for site-specific indoor radio propagation prediction based on geometric optics and geometric probability. IEEE Trans. Wirel. Commun. **1**(1), 112–124 (2002)

25. A. Hu et al., An ESPRIT-based approach for 2-D localization of incoherently distributed sources in massive MIMO systems. IEEE J. Sel. Top. Sign. Process. **8**(5), 996–1011 (2014)
26. ISO/IEC/IEEE International Standard for Information technology Telecommunications and information exchange between systems–Local and metropolitan area networks Specific requirements-part 11: wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 3: enhancements for very high throughput in the 60 GHz Band (adoption of IEEE Std 802.11ad-2012). ISO/IEC/IEEE 8802-11:2012/Amd.3:2014(E) vol. 59, pp. 1–634 (2014)
27. R.S. Kennedy, *Fading Dispersive Communication Channels* (Wiley-Interscience, New York, 1969)
28. M. Koivisto et al., Joint device positioning and clock synchronization in 5G ultra-dense networks (2016, submitted for publication)
29. N.P. Kuruvatti et al., Robustness of location based D2D resource allocation against positioning errors, in *IEEE Vehicular Technology Conference* (2015)
30. Q.C. Li, G. Wu, T.S. Rappaport, Channel model for millimeter wave communications based on geometry statistics, in *IEEE Globecom Workshop*, pp. 427–432 (2014)
31. Q.C. Li et al., Validation of a geometry-based statistical mmwave channel model using ray-tracing simulation, in *IEEE Vehicular Technology Conference*, pp. 1–5 (2015)
32. H. Liu et al., Survey of wireless indoor positioning techniques and systems. IEEE Trans. Syst. Man Cybern. C (Appl. Rev.) **37**(6), 1067–1080 (2007)
33. G.R. MacCartney et al., Path loss models for 5G millimeter wave propagation channels in urban microcells, in *IEEE Global Telecommunications (GLOBECOM) Conference* (2013)
34. R. Mendez-Rial et al., Channel estimation and hybrid combining for mmWave: phase shifters or switches? in *Information Theory and Applications Workshop (ITA), 2015*, pp. 90–97, Feb (2015). doi:10.1109/ITA.2015.7308971
35. J. Mo et al., Channel estimation in millimeter wave MIMO systems with one-bit quantization, in *IEEE Asilomar Conference on Signals, Systems and Computers* (2014)
36. I. Nevat, G.W. Peters, I.B. Collings, Location-aware cooperative spectrum sensing via Gaussian processes, in *Australian Communications Theory Workshop* (2012)
37. K. Papakonstantinou, D. Slock, ESPRIT-based estimation of location and motion dependent parameters, in *EEE Vehicular Technology Conference (VTC)* (2009)
38. S. Piersanti, L.A. Annoni, D. Cassioli, Millimeter waves channel measurements and path loss models, in *IEEE Wireless Communications Symposium (ICC)* (2012)
39. E. Pisek et al., High throughput millimeterwave MIMO beamforming system for short range communication, in *IEEE Consumer Communications and Networking Conference (CCNC)* (2014)
40. F. Pivit, V. Venkateswaran, Joint RF-feeder network and digital beamformer design for cellular base-station antennas, in *Antennas and Propagation Society International Symposium (APSURSI)*, pp. 1274–1275 (2013)
41. A. Poon, M. Taghivand, Supporting and enabling circuits for antenna arrays in wireless communications. Proc. IEEE **100**(7), 2207–2218 (2012)
42. H. Qasem, L. Reindl, Precise wireless indoor localization with trilateration based on microwave backscatter, in *IEEE Annual Wireless and Microwave Technology Conference* (2006)
43. T. Rappaport et al., Millimeter wave mobile communications for 5G cellular: it will work! IEEE Access **1**, 335–349 (2013)
44. A. Richter, Estimation of radio channel parameters: models and algorithms. Ph.D. thesis, The Ilmenau University of Technology, 2005
45. J. Richter, A. Salmi, V. Koivunen, An algorithm for estimation and tracking of distributed diffuse scattering in mobile radio channels, in *IEEE 7th Workshop on Signal Processing Advances in Wireless Communications* (2006)
46. R. Rusek et al., Scaling up MIMO: opportunities and challenges with very large arrays. IEEE Signal Process. Mag. **30**(1), 40–60 (2013)

47. P Sanchis et al., A novel simultaneous tracking and direction of arrival estimation algorithm for beam-switched base station antennas in millimeter-wave wireless broadband access networks, in *IEEE Antennas and Propagation Society International Symposium* (2002)
48. S. Sand et al., Position aware adaptive communication systems, in *Asilomar Conference on Signals, Systems and Computers* (2009)
49. V. Savic, E.G. Larsson, Fingerprinting-based positioning in distributed massive MIMO systems, in *IEEE Vehicular Technology Conference* (2015)
50. A.M. Sayeed, A virtual representation for time- and frequency-selective correlated MIMO channels, in *IEEE International Conference on Acoustics, Speech and Signal Processing* (2003)
51. A.M. Sayeed, T. Sivanadyan, Wireless communication and sensing in multipath environments using multi-antenna transceivers, in *Handbook on Array Processing, Sensor Networks*, ed. by K.J.R. Liu, S. Haykin (IEEE-Wiley, New York, 2010)
52. A. Shahmansoori et al., 5G position and orientation estimation through millimeter wave MIMO, in *IEEE Global Telecommunications (GLOBECOM) Conference* (2015)
53. Y. Shen, M.Z. Win, Fundamental limits of wideband localization Part I: a general framework. IEEE Trans. Inf. Theory **56**(10), 4956–4980 (2010)
54. P.F.M. Smulders, Statistical characterization of 60-GHz indoor radio channels. IEEE Trans. Antennas Propag. **57**(10), 2820–2829 (2009)
55. P. Sudarshan et al., Channel statistics-based RF pre-processing with antenna selection. IEEE Trans. Wireless Commun. **5**(12), 3501–3511 (2006)
56. A.L. Swindlehurst et al., Millimeter-wave massive MIMO: the next wireless revolution? IEEE Commun. Mag. **52**(9), 56–62 (2013)
57. R.D. Taranto et al., Location-aware communications for 5G networks: how location information can improve scalability, latency, and robustness of 5G. IEEE Signal Process. Mag. **31**(6), 102–112 (2014)
58. M.N. Tehrani, M. Uysal, H. Yanikomeroglu, Device-to-device communication in 5G cellular networks: challenges, solutions, and future directions. IEEE Commun. Mag. **52**(5), 86–92 (2014)
59. D. Tse, P. Viswanath, *Fundamentals of Wireless Communication* (Cambridge University Press, Cambridge, 2007)
60. M. Vari, D. Cassioli, mmWaves RSSI indoor network localization, in *IEEE Workshop on Advances in Network Localization and Navigation (ICC)* (2014)
61. J. Werner et al., Primary user localization in cognitive radio networks using sectorized antennas, in *Annual Conference on Wireless On-demand Network Systems and Services* (2013)
62. J. Werner et al., Joint user node positioning and clock offset estimation in 5G ultra-dense networks, in *IEEE Global Telecommunications (GLOBECOM) Conference* (2015)
63. J. Werner et al., Performance and Cramer-Rao bounds for DoA/RSS estimation and transmitter localization using sectorized antennas. IEEE Trans. Veh. Technol. **65**, 3255–3270 (2015)
64. H. Wymeersch, J. Lien, M.Z. Win, Cooperative localization in wireless networks. Proc. IEEE **97**(2), 427–450 (2009)
65. X. Zhang, A.F. Molisch, S. Kung, Variable-phase-shift-based RF-baseband codesign for MIMO antenna selection. IEEE Trans. Signal Process. **53**(11), 4091–4103 (2005)
66. Q. Zhao, J. Li, Rain attenuation in millimeter wave ranges, in *Proceedings of IEEE International Symposium Antennas, Propagation and EM Theory*, pp. 1–4 (2006)
67. Y. Zhou, J. Li, L. Lamont, Multilateration localization in the presence of anchor location uncertainties, in *IEEE Global Communications Conference (GLOBECOM)* (2012)
68. D. Zhu, J. Choi, R.W. Heath Jr., Auxiliary beam pair design in mmWave cellular systems with hybrid precoding and limited feedback, in *IEEE International Conference on Acoustics, Speech, and Signal Processing* (2016)

Chapter 10
Formation Control of Multi-Agent Systems with Location Uncertainty

Markus Fröhle, Themistoklis Charalambous, Henk Wymeersch, Siwei Zhang, and Armin Dammann

10.1 Introduction

Increasingly sophisticated algorithms implemented on autonomous agents/robots with massive computational capabilities have enabled solving complex tasks in uncertain environments, such as mapping of disaster areas and search-and-rescue operations. For agents to explore and interact with the environment, it is important that they have a coherent view of this environment and their positions within it (see, for example, [19, 25] and references therein). Such situational awareness is typically achieved through simultaneous localisation and mapping (SLAM) [32] and localisation using heterogeneous sensor fusion [36]. While research on situational awareness has traditionally focused on improving the localisation accuracy, the focus has now shifted to localisation methods that have knowledge of computational complexity limitations of the agents, energy and communication constraints as well as the agent's higher-level task [8, 10]; for instance, a higher-level task may be the navigation of the robot in the environment from its current position to its final position. This chapter considers the problem of how position uncertainty in

M. Fröhle (✉) • H. Wymeersch
Department of Signals and Systems, Chalmers University of Technology, Gothenburg 41296, Sweden
e-mail: frohle@chalmers.se

T. Charalambous
Department of Electrical Engineering and Automation, Aalto University, 02150 Espoo, Finland
e-mail: themistoklis.charalambous@aalto.fi

S. Zhang • A. Dammann
German Aerospace Center (DLR), Institute of Communications and Navigation Communications Systems, Oberpfaffenhofen-Wessling 82234, Germany

J. Nurmi et al. (eds.), *Multi-Technology Positioning*,
DOI 10.1007/978-3-319-50427-8_10

multi-agent systems affects control and communication, and how positioning can be improved taking into account the underlying limitations of such systems.

To support positioning, a robot is typically equipped with a plurality of sensors providing position information. Since using any of these sensors consumes energy, the selection *of which sensor(s) to use when* is important. This *sensor selection* problem refers to the agent choosing a subset out of all its available sensors, in order to optimise an objective (e.g., the trace or determinant of the state estimation error covariance [16, 26], an expected utility defined by Bayes risk [4], or a measure of information such as conditional entropy [35]). The objective can be optimised (a) over a single time-step [3, 4, 16, 26] or (b) over a prediction window [14, 28, 33, 35]. In the first class of approaches, the problems are often assumed to be temporally separable, leading to efficient solutions [3, 4, 16, 26]. In the more general second class of approaches, the problems are temporally inherently inseparable, leading to exponential complexity in the prediction horizon. This complexity is reduced through appropriate techniques [14, 28, 33, 35]. These techniques generally focus on minimising a function of the state error covariance, which is not always relevant for battery-constrained devices. For such devices, minimising energy consumption is more meaningful, while ensuring a certain positioning quality.

While sensor selection can be performed on a per-agent basis, agents can also *cooperate* to solve tasks by combining agents with heterogeneous sensors, such as formation control [27]. Accomplishing such tasks relies on accurate and fresh location information. Existing approaches [6, 17, 18, 24, 38] assume perfect position information, which may not be available in reality. The limited range of the anchor signals can be countered by cooperation among agents, in order to determine the position of all agents. Cooperative positioning based on belief propagation and message passing algorithms was presented in [36]. This was extended by [40], accounting for the overhead and cost related to accessing the radio channel by a time-division multiple access protocol using an orthogonal-frequency division multiplexing (OFDM) signal. The Cramér-Rao bound (CRB) for cooperative positioning accuracy introduced in [29] provides the fundamental limit of this approach. Therefore, formation control should account for localisation errors [39], which can be improved through active information seeking [21, 22].

Finally, when performing a task, agents need to know their own absolute location, both for control and to maintain connectivity with each other over the wireless channel. Connectivity can be improved through *prediction of channel gains* [7]. The channel comprises three components: deterministic path-loss, shadowing, and small-scale fading [11]. The latter two components are generally modeled as random variables. For typical wave-lengths, small-scale fading decorrelates over a few centimetres, whereas shadowing decorrelates over 50–100 m outdoors [11] and 1–5 m indoors [2, 15], based on standard shadowing correlation models [13, 34]. For a multi-agent system, [1] modeled shadowing through a spatial loss field. Channel prediction was studied in [8], which proposed a Gaussian Process (GP) framework, and [20], which considered the impact of the channel parameters on the prediction variance. An assumption in [1, 8, 20] was the availability of perfect location information. This was partially addressed in [37], which extended [20] to determine

the impact of localisation errors on channel prediction. Location uncertainty at the receiver but not the transmitter side was explicitly accounted for in [23].

This chapter covers recent progress in the above-mentioned areas: sensor selection, cooperation, and channel prediction. In particular, the rest of the chapter is organised as follows. We first describe several fundamental concepts in the areas of communication, control, and localisation in Sect. 10.2. We then consider the interaction between control and localisation in Sect. 10.3. In particular, Sect. 10.3.1 considers the sensor selection problem, where the aim is to find a path to a goal with minimum sensing cost, while maintaining a certain localisation quality. Section 10.3.2 focuses on a cooperative scenario, where again a goal must be achieved, but agents cooperate to maintain a certain localisation quality, rather than relying on a variety of sensors. In Sect. 10.4, the problem of channel prediction under location uncertainty will be tackled. These problems highlight the need to maintain good position quality in Multi-Agent Systems (MAS).

10.2 Localisation for Communication and Control

This section briefly highlights the use of localisation for communication and control of MAS. We first describe a standard communication model in Sect. 10.2.1, followed by a generic optimal control problem in Sect. 10.2.2, and a metric to assess localisation algorithms in Sect. 10.2.3.

10.2.1 Inter-Agent Communication

In MAS, the communication between agents occurs over the wireless medium. The channel power between a transmitter (TX) and a receiver (RX), with locations $\mathbf{p}_{\mathrm{TX}} \in \mathbb{R}^D$ and $\mathbf{p}_{\mathrm{RX}} \in \mathbb{R}^D$, respectively, where D is the dimensionality of the space, can be modelled as a spatially correlated random process. Assuming communication averages out small-scale fading, either in time (over a time window) or frequency (average power over a large frequency band), the received signal power in dBm can be expressed as [11]

$$P_{\mathrm{RX}}(\mathbf{p}_{\mathrm{TX}}, \mathbf{p}_{\mathrm{RX}}) = P_{\mathrm{TX}} + L_0 - 10\,\eta\,\log_{10}\frac{\|\mathbf{p}_{\mathrm{TX}} - \mathbf{p}_{\mathrm{RX}}\|}{d_0} + \Psi(\mathbf{p}_{\mathrm{TX}}, \mathbf{p}_{\mathrm{RX}}), \tag{10.1}$$

where the transmitted power is P_{TX}, L_0 comprises antenna and propagation gains, d_0 is a reference distance (herein we assume that d_0 is 1 m) and η is the path-loss exponent. Shadowing (in the dB domain) is modelled as a zero-mean Gaussian random process with a given spatial correlation [13, 31] and exhibiting channel reciprocity (i.e., $\Psi(\mathbf{p}_{\mathrm{TX}}, \mathbf{p}_{\mathrm{RX}}) = \Psi(\mathbf{p}_{\mathrm{RX}}, \mathbf{p}_{\mathrm{TX}})$). Provided there are no significant changes in the environment, the received power is thus largely predictable and

dependent on the absolute positions of the TX and RX. Hence, in order to predict the received signal power, the locations of the TX and RX should be known.

10.2.2 Multi-Agent Control

Control of MAS usually involves the design of optimal control strategies that should be distributed, relying on local information only. These optimal control strategies emerge from having the interconnected agents to optimise an objective function $f : \mathscr{X} \times \mathscr{U} \mapsto \mathbb{R}$, i.e.,

$$\begin{aligned} &\underset{\mathbf{u}\in\mathscr{U}}{\text{minimise}} \;\; f(\mathbf{x}, \mathbf{u}) \\ &\text{subject to } \dot{\mathbf{x}} = h(\mathbf{x}_0, \mathbf{u}, t), \end{aligned} \tag{10.2}$$

where $\mathbf{x} \in \mathscr{X}$ comprises the state of the agents (e.g., position and velocity of multiple agents over a certain time window), $\mathbf{x}_0$ represents the initial state, $\mathbf{u}$ denotes the control input (e.g., acceleration), belonging to some fixed set $\mathscr{U}$, and $h(\cdot)$ represents the dynamics. From the above formulation, it is clear that solving the problem in a decentralised fashion, where agents find the optimal solution without a central coordinator, information exchange between the interconnected agents regarding their state (including the position) is necessary.

10.2.3 Bounding Uncertainty of Estimators: The Cramér-Rao Bound

The Cramér-Rao Bound (CRB) is a lower bound on the estimation error variance of any unbiased estimator [29]. Due to its computational simplicity, the CRB is an attractive tool for system verification [40] and adaptive algorithm design [5]. The CRB is expressed as

$$\mathbb{E}\big[\|\hat{\mathbf{x}}^{(k)} - \mathbf{x}^{(k)}\|^2\big] \geq \mathrm{CRB}[\mathbf{x}^{(k)}] = \mathrm{tr}\big((\mathbf{F}_{\mathbf{x}}^{(k)})^{-1}\big), \tag{10.3}$$

where $\hat{\mathbf{x}}^{(k)}$ is the estimate of $\mathbf{x}^{(k)}$ and $\mathbf{F}_{\mathbf{x}}^{(k)}$ is the so-called Fisher information matrix (FIM) with respect to $\mathbf{x}$ at time-step k. The FIM is defined as

$$\mathbf{F}_{\mathbf{x}}^{(k)} = -\mathbb{E}_{\mathbf{y}^{(k)}|\mathbf{x}^{(k)}}\left[\nabla_{\mathbf{x}^{(k)}}^{\mathrm{T}} \nabla_{\mathbf{x}^{(k)}} \log p(\mathbf{y}^{(k)}|\mathbf{x}^{(k)})\right], \tag{10.4}$$

where $\mathbf{y}^{(k)}$ is the measurement associated with state $\mathbf{x}^{(k)}$.

10.3 Impact of Location Uncertainty on Mission Goals

In this section, we will focus on the connection between localisation and control of agents towards a goal. Two different problems will be discussed: the first one deals with selecting paths that minimise a long-term cost, while the second one deals with progressively moving towards a goal, while also optimising positioning quality. Both of these problems rely on the concepts previously introduced: optimization for the control actions and FIM for the positioning quality.

10.3.1 Limiting Location Uncertainty

10.3.1.1 Problem Formulation

Exemplified in Fig. 10.1, a mobile agent with an estimated start position $\mathbf{p}_0$ aims at reaching a goal position $\mathbf{p}_{\text{goal}} \in \mathbb{R}^2$ in discrete time steps. The agent is equipped with M sensors providing location information, with the use of sensor m having a cost (e.g., power consumption) $c_m \geq 0$. The agent can use $J \geq 1$ different paths towards the goal position, two of which are shown in Fig. 10.1. At each time and along each possible path, the agent can use at most one sensor. Each path has a certain inherent cost associated with its length (agent movement cost). The goal of the agent is to find the least expensive path to the goal, while providing a certain quality of position information.

We will consider the agent's position as its state, with the initial state $\mathbf{x}^{(0)}$ modeled as a Gaussian random variable $\mathbf{x}^{(0)} \sim \mathcal{N}(\boldsymbol{\mu}^{(0)}, \mathbf{P}^{(0)})$. The goal position $\mathbf{p}_{\text{goal}}$ is known exactly, as is the floor plan of the environment and the quality of each sensor within the environment. We will focus on one of the J paths, which comprises $N_j + 1$ positions, $\mathbf{p}_j^{(0)}, \mathbf{p}_j^{(1)}, \ldots, \mathbf{p}_j^{(N_j)}$, where $\mathbf{p}_j^{(0)} = \boldsymbol{\mu}^{(0)}$ and $\mathbf{p}_j^{(N_j)} = \mathbf{p}_{\text{goal}}$. When the agent moves along the path, its state statistics can be modeled with a process model

$$\mathbf{x}^{(k)} = \mathbf{A}^{(k-1)}\mathbf{x}^{(k-1)} + \mathbf{B}^{(k-1)}\mathbf{u}^{(k-1)} + \mathbf{n}^{(k-1)}, \tag{10.5}$$

where $\mathbf{u}^{(k)}$ is the control signal at time k, $\mathbf{A}^{(k-1)}$ and $\mathbf{B}^{(k-1)}$ are known matrices, and $\mathbf{n}^{(k-1)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{Q})$ is a Gaussian process noise with error covariance matrix $\mathbf{Q}$, and a measurement model of sensor m:

$$\mathbf{y}_m^{(k)} = \mathbf{H}_m^{(k)}\mathbf{x}^{(k)} + \mathbf{v}_m^{(k)}, \tag{10.6}$$

where $\mathbf{H}_m^{(k)}$ is a known matrix and $\mathbf{v}_m^{(k)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \mathbf{R}_m^{(k)})$ is the measurement noise associated with sensor m when used in location $\mathbf{x}^{(k)}$. This linear measurement model can be motivated from a loose coupling view, where the measurement corresponds to a position estimate, as opposed to range estimates. The linear process and measurement models enable us to use the Kalman filter [30].

The agent now performs the following two steps, before proceeding to the goal:

1. for each path, the agent predicts a required control sequence $\mathbf{u}^{(0:N-1)}$, based on $\mathbf{p}^{(k)}$;
2. for a given path and a given sequence of selected sensing systems, the agent can compute (i) the expected accuracy of its predicted position using a Kalman filter and (ii) the associated cost.

The expected accuracy is determined by the FIM, introduced in Sect. 10.2.3. In particular, when the agent had a FIM $\boldsymbol{F}^{(k-1)}$ at time $k-1$ and activates sensor m, the FIM at time k will be

$$\begin{aligned}\mathbf{F}^{(k)} = {} & \mathbf{H}_m^{(k)\mathrm{T}} \left(\mathbf{R}_m^{(k)}\right)^{-1} \mathbf{H}_m^{(k)} + \mathbf{Q}^{-1} - \mathbf{Q}^{-1}\,\mathbf{A}^{(k-1)} \\ & \times \left(\mathbf{F}^{(k-1)} + \left(\mathbf{A}^{(k-1)}\right)^{\mathrm{T}} \mathbf{Q}^{-1}\,\mathbf{A}^{(k-1)}\right)^{-1} \left(\mathbf{A}^{(k-1)}\right)^{\mathrm{T}} \mathbf{Q}^{-1}\,, \end{aligned} \tag{10.7}$$

Based on this information, the agent can determine the expected cost of any of the J paths, and then select the least expensive path.

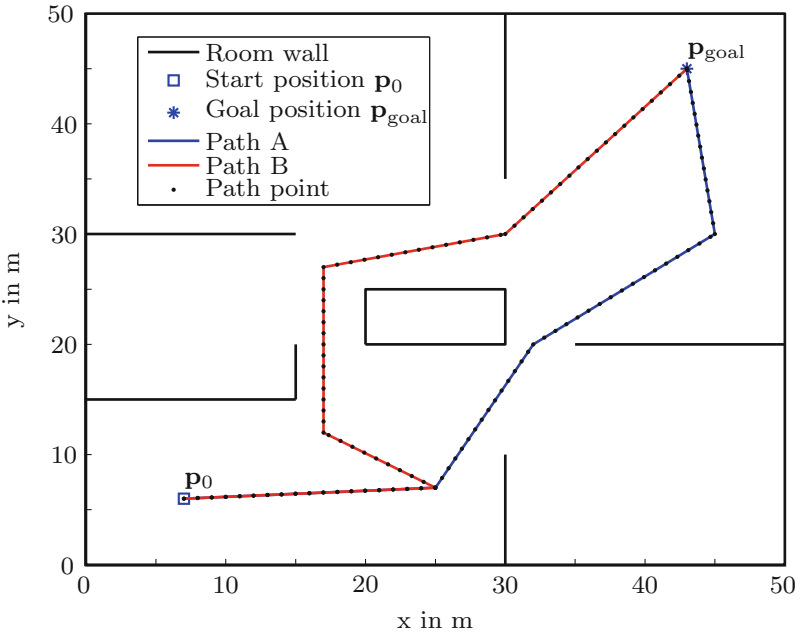

Fig. 10.1 A scenario with two paths A and B from a start position $\mathbf{p}_0 = [7, 6]^{\mathrm{T}}$ to a goal position $\mathbf{p}_{\text{goal}} = [43, 45]^{\mathrm{T}}$. The objective of the agent is to determine which path has the lowest cost, while ensuring a certain positioning quality

10.3.1.2 Optimisation Formulation

We can now cast the sensor selection problem in a standard form as follows:

$$\underset{\mathbf{D}}{\text{minimise}} \quad \mathbf{c}^{\mathrm{T}}\mathbf{D}\mathbf{1} \tag{10.8a}$$

$$\text{subject to} \quad \mathbf{D}^{\mathrm{T}}\mathbf{1} = \mathbf{1} \tag{10.8b}$$

$$\mathbf{D} \in \{0,1\}^{M\times N} \tag{10.8c}$$

$$\mathrm{tr}\left(\left(\mathbf{F}^{(k)}\right)^{-1}\right) \leq \Delta^2,\ k \in \{1,2,\ldots,N\}, \tag{10.8d}$$

where $\mathbf{c} = [c_1, c_2, \ldots, c_M]^{\mathrm{T}}$, $\mathbf{1}$ is a column vector containing all ones, $\mathbf{F}^{(k)}$ is the FIM from Sect. 10.2.3 and $\mathbf{D}$ denotes the optimisation variable:

$$\mathbf{D} = \begin{bmatrix} d_1^{(1)} & \ldots & d_1^{(N)} \\ \vdots & \ddots & \vdots \\ d_M^{(1)} & \ldots & d_M^{(N)} \end{bmatrix}, \tag{10.9}$$

where $d_m^{(k)} = 1$ means that the m-th sensor system is used in the k-th time step; (10.8b) ensures that only one out of M measurement systems is used at any time; (10.8d) ensures the trace of the inverse of the posterior FIM to stay below a specific threshold (Δ^2, in m^2). This makes sure that the expected[1] root mean square position error (RMSE) does not exceed Δ. From (10.7), it is readily verified that $\mathbf{F}^{(k)}$ in (10.8d) can be expressed as a linear function of $\mathbf{D}$.

Problem (10.8) is combinatorial in nature, rendering its solution intractable as time horizon N increases. Rather than solving (10.8) directly, lower and upper bounds on the cost per path can be obtained through semi-definite programming (SDP), which involves relaxing the integer constraint on $\mathbf{D}$, and dynamic programming (DP), respectively. In the DP approach, we discretise the positioning quality and build up a hidden Markov model with the selected measurement systems as input. This allows us to run the Viterbi algorithm and efficiently compute the cost of any combination of measurement system selections. Additional details can be found in [9].

10.3.1.3 Performance Evaluation

We consider a scenario detailed in [9]. In brief, the scenario is in the environment of Fig. 10.1, and uses a geometric path planner to generate 196 paths. We considered

[1] Note that we only determine a path with lowest *expected* cost, under the assumption that the agent measures in the positions $\mathbf{p}_k$.

$M = 4$ sensors: Sensor 1 corresponds to no sensing, with $\mathbf{R}_1^{(k)} = 10^6\,\mathbb{I}$; Sensor 2 corresponds to a GPS-like sensor which is better near the windows, but poor inside the environment; Sensor 3 is an RFID-like sensor with tags placed near the corners; Sensor 4 is an ultra-wideband-like (UWB) sensor with four reference nodes in the corners. Such a system would exhibit high sensor measurement quality in the middle of the environment, but not near the corners. We assign costs as follows: $c_1 = 1$ comprising the cost of movement, $c_2 = 3$, $c_4 = 2$, $c_4 = 4$.

The DP implementation had 10 states, while the SDP implementation used the software package CVX [12]. As benchmarks, we have also implemented a greedy approach, selecting at each time step the cheapest sensor that can ensure (10.8d) is satisfied.

The path cost with the three different methods is evaluated for different process noise levels $\sigma_Q^2 \in \{0.01, 0.1, 0.2\}$, shown in Fig. 10.2, where $\mathbf{Q} = \sigma_Q^2\,\mathbb{I}$. Paths are sorted by length. We note that when σ_Q^2 is low (top and middle plot of Fig. 10.2), there is a large gap between the solution using the greedy method compared to using the DP method. Hence, for low σ_Q^2 there is a clear benefit in using DP considering a long horizon. For large values of σ_Q^2, the sensors must be used more frequently to maintain acceptable positioning quality. Hence, the greedy and DP method exhibit similar performance.

Furthermore, the SDP method leads to costs close to the trivial lower bound (where the cheapest sensor is used all the time irrespective whether (10.8d) is met or not). This is because the SDP method can allow for fractional sensor usage, always exactly meeting the RMSE constraint.

10.3.2 Location-Aware Formation Control

The ideas from the previous section were mainly limited by the high complexity. Here, we will not consider control over a time horizon, but rather a greedy framework. This allows us to address a richer set of problems. As before, we will have a goal, but now consider multiple cooperating agents that help each other in localising as they move to the goal.

10.3.2.1 Problem Formulation

We consider a swarm of M agents that want to move towards a goal with location $\mathbf{q} \in \mathbb{R}^2$, supported by anchors/base stations. We thus aim to find (global) control actions $\mathbf{u}^{(k)}$ at time k moving the agents according to

$$\mathbf{p}^{(k)} = \mathbf{p}^{(k-1)} + \mathbf{u}^{(k)} + \boldsymbol{\epsilon}^{(k)}, \tag{10.10}$$

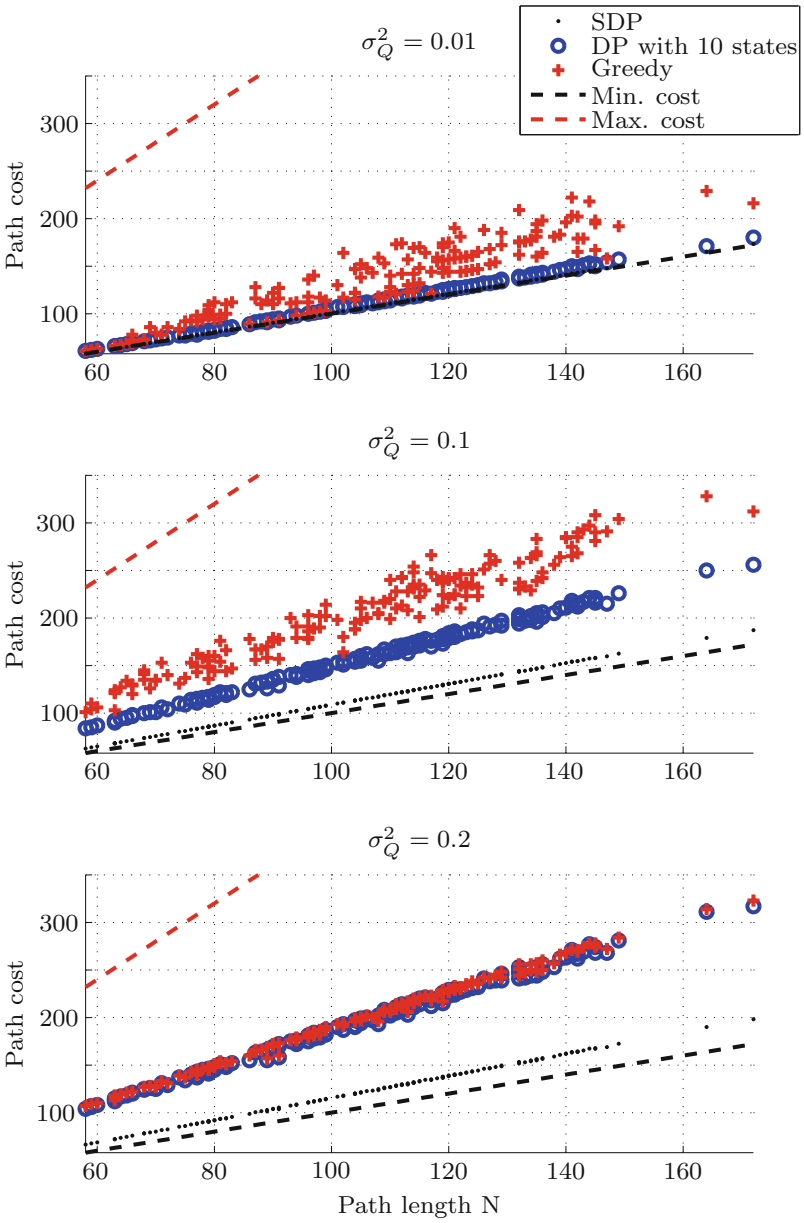

Fig. 10.2 Path costs for the different paths sorted by path length. *Top to bottom* plot corresponds to different values of the process noise variance $\sigma_Q^2 \in \{0.01, 0.1, 0.2\}$. The minimum and the maximum path costs are obtained by using the sensor with lowest and highest cost along the whole path irrespective whether (10.8d) is met or not. The lower bound (SDP) and upper bound (DP) are compared to the greedy solution

where $\boldsymbol{\epsilon}^{(k)} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$ is the global transition noise with diagonal covariance matrix $\mathbf{Q}$. Under perfect location information this can be done by solving the optimisation problem

$$\text{Problem } \mathscr{P}_\alpha: \quad \underset{\mathbf{u}_\alpha^{(k)}}{\text{minimise}} \ \|\mathbf{p}^{(k)} - \mathbf{1}_{M\times 1} \otimes \mathbf{q}\| \qquad (10.11)$$
$$\text{subject to } \mathbf{u}_\alpha^{(k)} \in \mathscr{U}_\alpha.$$

where $\mathscr{U}_\alpha$ is a set representing all valid control signals. We will consider $\mathscr{U}_\alpha = \{\mathbf{u} \in \mathbb{R}^{2M} |\ \|\mathbf{u}\| = \mu_\alpha\}$. A goal approaching command can thus be by moving along the gradient of the objective:

$$\mathbf{u}_\alpha^{(k)} = -\mu_\alpha \frac{\mathbf{p}^{(k-1)} - \mathbf{1}_{M\times 1} \otimes \mathbf{q}}{\|\mathbf{p}^{(k-1)} - \mathbf{1}_{M\times 1} \otimes \mathbf{q}\|}. \qquad (10.12)$$

In case the position is not known, we can replace $\mathbf{p}^{(k-1)}$ by an estimate $\hat{\mathbf{p}}^{(k-1)}$.

The problem with the above approach is that in general $\hat{\mathbf{p}}^{(k)}$ will deviate more and more from $\mathbf{p}^{(k)}$ as time k progresses. To ensure that all agents stay localised, we solve the following problem

$$\text{Problem } \mathscr{P}_\beta: \quad \underset{\mathbf{u}_\beta^{(k)}}{\text{minimise}} \ \mathbb{E}\left[\|\hat{\mathbf{p}}^{(k)} - \mathbf{p}^{(k)}\|\right] \qquad (10.13)$$
$$\text{subject to } \mathbf{u}_\beta^{(k)} \in \mathscr{U}_\beta,$$

where $\mathscr{U}_\beta = \{\mathbf{u} \in \mathbb{R}^{2M} |\ \|\mathbf{u}\| = \mu_\beta\}$ and the expectation should be interpreted as being over $\hat{\mathbf{p}}^{(k)}$.

Our goal is thus to design a swarm formation controller for joint goal-approaching and location information seeking. An example of the described swarm system is illustrated in Fig. 10.3, where a swarm of three agents moves from region A to region B. Three base stations are located at each region. The white solid lines in the figure illustrate the radio connections between nodes for communications and ranging.

10.3.2.2 Optimisation Formulation

We will consider an alternating optimisation approach, where the swarm alternately solves the problems $\mathscr{P}_\alpha$ and $\mathscr{P}_\beta$. Since we already provided an expression to solve $\mathscr{P}_\alpha$, we will focus on $\mathscr{P}_\beta$ and drop the time index. The problem can be expressed in terms of the Fisher Information Matrix (FIM):

$$\underset{\mathbf{u}_\beta \in \mathscr{U}_\beta}{\text{minimise}} \quad \text{tr}\left(\mathbf{F}_\mathbf{p}^{-1}\right). \qquad (10.14)$$

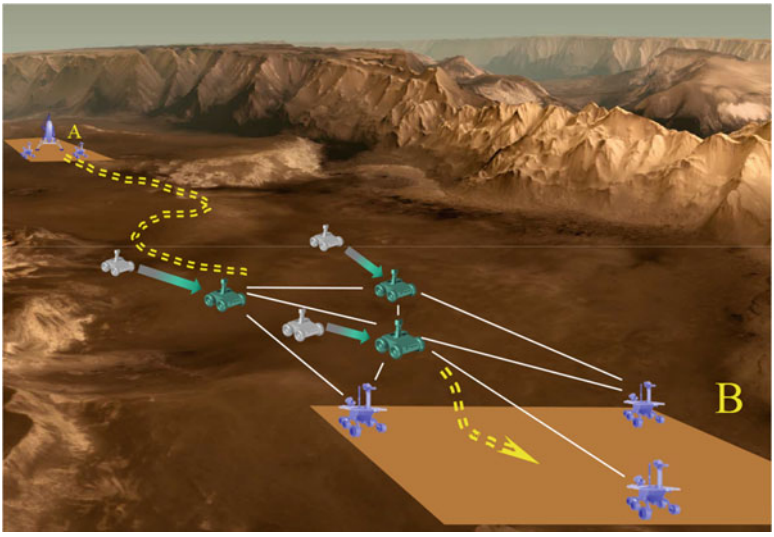

Fig. 10.3 A swarm navigation system on Mars (not in real scale): three agents move from region *A* to region *B*. Three base stations are located at each region. The *white solid lines* illustrate the radio connections between nodes for communications and ranging. Background picture credits: ESA / DLR / FU Berlin (G. Neukum)

The specific form of the FIM depends on the scenario at hand, the type of measurements, whether a priori information is available, etc. Considering a non-Bayesian setting, $\mathbf{F}_{\mathbf{p}}$ depends on the unknown value $\mathbf{p}$, which should be interpreted as the position after implementing the control [i.e., $\mathbf{p}^{(k)}$ in (10.10)]. The gradient $\mathbf{c} \in \mathbb{R}^{2M}$ of the objective function is

$$\mathbf{c} = \left[\mathbf{c}_1^{\mathrm{T}}, \ldots, \mathbf{c}_u^{\mathrm{T}}, \ldots, \mathbf{c}_M^{\mathrm{T}}\right]^{\mathrm{T}} = \nabla_{\mathbf{p}} \mathrm{tr}\left(\mathbf{F}_{\mathbf{p}}^{-1}\right), \tag{10.15}$$

where $\mathbf{c}_u \in \mathbb{R}^2$ is the gradient component of agent u [41]. The steepest descent gradient controller solving $\mathscr{P}_\beta$ with a step size μ_β can be expressed as

$$\mathbf{u}_\beta = -\mu_\beta \frac{\mathbf{c}}{\|\mathbf{c}\|}. \tag{10.16}$$

Since the true positions of the agents are unknown, the gradient $\mathbf{c}$ is evaluated in the position estimates $\hat{\mathbf{p}}$ and then utilised to generate the control command $\mathbf{u}_\beta$.

As a specific instance of this problem, and to exemplify the potential for closed-form expressions, we will consider a common measurement model, as defined in Sect. 10.2.3. We let the measurements correspond to distance estimates with neighbouring agents and anchors. The statistical model for the distance estimation error between nodes u and v with positions $\mathbf{p}_u$ and $\mathbf{p}_u$ is a zero-mean Gaussian random variable with variance $\sigma_{u,v}^2$. The ranging variance $\sigma_{u,v}^2$ is, in general, distance dependent by the radio propagation model [40]. In this work, the analytical ranging

variance model proposed in [41] is applied to capture the main features of radio-based ranging. For short distances, the ranging variance is quadratically proportional to the distance. After a certain distance, the ranging variance rapidly increases to the maximum ranging variance due to the low signal-to-noise ratio (SNR).

10.3.2.3 Performance Evaluation

Simulations are conduced to illustrate the performance of the considered formation control algorithm. As initialisation, 25 agents are uniformly deployed in an area of 10 m × 10 m at region A. Agents need to move to region B 600 m away, based on their position estimates. In each region three base stations are deployed. Each node is equipped with a radio transceiver with coverage range of 90 m. There is a blind region of over 400 m along the path where none of the base stations can be directly connected due to low SNR. The step sizes are set to $\mu_\alpha = 0.15$ m and $\mu_\beta = 0.1$ m. Agents' transition noise variance at each dimension is set to 0.01 m^2. Two snapshots at time step 30,000 and 31,200 are shown in Fig. 10.4. Agents are

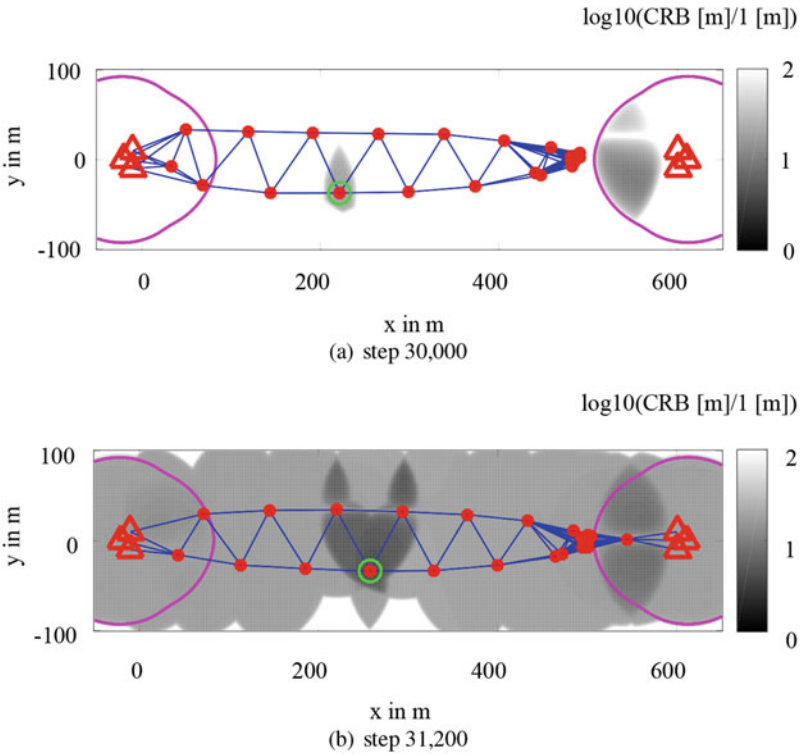

Fig. 10.4 Formation at time steps 30,000 (*upper*, **a**) and 31,200 (*lower*, **b**): agents are shown as *red dots* and base stations as *red triangles*. *Two magenta curves* show the radio coverage borders of base stations. The *grey shaded background* is the averaged swarm localisation CRB when the agent of interest (marked with a *green circle*) moves to that position

shown as red dots and base stations as red triangles. Two magenta curves show the radio coverage border of base stations. The grey shaded background is the averaged swarm localisation CRB, i.e. $\sqrt{\text{tr}(\mathbf{F}_{\mathbf{p}}^{-1})/M}$, when the agent of interest (marked with a green circle) moves to that position. The agent of interest can move inside the connected dark area without jeopardising the global position accuracy. The two snapshots demonstrate a procedure of establishing a virtual bridge in order to propagate high localisation accuracy into the blind region. The swarm automatically stretches to a double-row formation along x dimension to extend the locatable area. At step 30,000, the agents in the middle, e.g., the one marked with green, can only move within a limited area since they are crucial vertices to maintain rigidity of the whole network, whereas the agents clustered in front of the swarm are more flexible to move due to the redundant links. As a consequence, the crucial agents in the middle mainly perform location information seeking and the ones in front are dominantly controlled with goal-approaching. At step 31,200, the front agents have already moved inside the coverage of the base stations at region B and can propagate precise position information acquired from base stations as well. Therefore, the agents behind are released from their role of crucial vertices and can apply goal-approaching again. From the figure it can be seen that the connected dark area in which the agent of interest can move is significantly increased. With this numerical result, it can be concluded that the swarm location information seeking and goal approaching can be achieved with the considered formation control algorithm. The location information seeking framework can be adapted to different applications, e.g. swarm return-to-base application, where both swarm and base positions are to be estimated [42].

10.4 Impact of Location Uncertainty on Channel Gain Prediction

In this section, we step away from the control aspect and rather deal with channel modelling and prediction in the presence of location uncertainty. We will again rely on ideas introduced in Sect. 10.2.

10.4.1 Problem Formulation

A database of N power measurements is available between different TX and RX locations. Each measurement is of the form $y = P_{\text{RX}}(\mathbf{x}) + n$, where $n \sim \mathcal{N}_n(0, \sigma_n^2)$ and the total set of measurements is denoted $\mathbf{y} = [y_1, y_2, \ldots, y_N]^{\text{T}}$. The locations of both, TX and RX, are known statistically, through the probability density functions (pdfs) $p(\mathbf{p}_{\text{TX}})$ and $p(\mathbf{p}_{\text{RX}})$, which are assumed to be described by a finite number of parameters $\mathbf{s}$ (e.g., means and covariances of both agents' locations). Hence, the distributions associated with the N measurements can be

denoted by $\mathbf{S} = [\mathbf{s}_1^{\mathrm{T}}, \mathbf{s}_2^{\mathrm{T}}, \ldots, \mathbf{s}_N^{\mathrm{T}}]^{\mathrm{T}}$. For notational convenience, the positions of a TX–RX pair will be denoted by $\mathbf{x} = [\mathbf{p}_{\mathrm{TX}}^{\mathrm{T}}, \mathbf{p}_{\mathrm{RX}}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{2D}$, allowing expressions such as $P_{\mathrm{RX}}(\mathbf{x})$ and $P_{\mathrm{RX}}(\mathbf{s}) = \int P_{\mathrm{RX}}(\mathbf{x}) p(\mathbf{x}) \mathrm{d}\mathbf{x}$.

The problems considered in this section are (a) learning of channel parameters; (b) prediction of the channel at an unvisited location.

10.4.2 Channel Prediction

The Gaussian process (GP) framework of [23], called uncertain GP (uGP), is here adapted for learning and prediction of the wireless channel considering TX and RX location uncertainty. This uGP framework is contrasted to classical GP (cGP), wherein location uncertainty is ignored. To simplify the exposition, we will ignore the deterministic path loss, and consider only shadowing. The received power at uncertain location pair $\mathbf{x}$ (described by its distribution parameters $\mathbf{s}$) is modelled as a GP, denoted by

$$P_{\mathrm{RX}}(\mathbf{s}) \sim \mathscr{GP}\left(0, k(\mathbf{s}, \mathbf{s}')\right), \tag{10.17}$$

where a suitable choice for the covariance function is

$$k(\mathbf{s}, \mathbf{s}') = \iint c(\mathbf{x}, \mathbf{x}') p(\mathbf{x}) p(\mathbf{x}') \mathrm{d}\mathbf{x}\, \mathrm{d}\mathbf{x}', \tag{10.18}$$

in which $c(\mathbf{x}, \mathbf{x}')$ is a covariance function under precise location information. Note that (10.18) reverts back to a classical GP (cGP) approach, with classical covariance functions [8], when there is no location uncertainty. In order to obtain a closed-form expression for (10.18), $c(\mathbf{x}, \mathbf{x}')$ should be limited to certain specific families, e.g., a squared exponential. In that case, with $p = 2$, one can write

$$c(\mathbf{x}, \mathbf{x}') = \sigma_\Psi^2 \exp\left(-\frac{\|\mathbf{p}_{\mathrm{TX}} - \mathbf{p}'_{\mathrm{TX}}\|^p}{d_c^p} - \frac{\|\mathbf{p}_{\mathrm{RX}} - \mathbf{p}'_{\mathrm{RX}}\|^p}{d_c^p} \right), \tag{10.19}$$

when $\mathbf{x} \neq \mathbf{x}'$ and $c(\mathbf{x}, \mathbf{x}) = \sigma_\Psi^2 + \sigma_{\mathrm{proc}}^2$. The spatial correlation of the received power between two links $\mathbf{x}$ and $\mathbf{x}'$ depends on the TX and RX displacement as illustrated in Fig. 10.5. The problem of learning now involves determining the parameters $\boldsymbol{\theta} = [d_c, \sigma_\Psi, \sigma_{\mathrm{proc}}]^{\mathrm{T}}$ from the database. Once $\boldsymbol{\theta}$ has been determined, the prediction problem can then be solved, by virtue of using a GP framework.

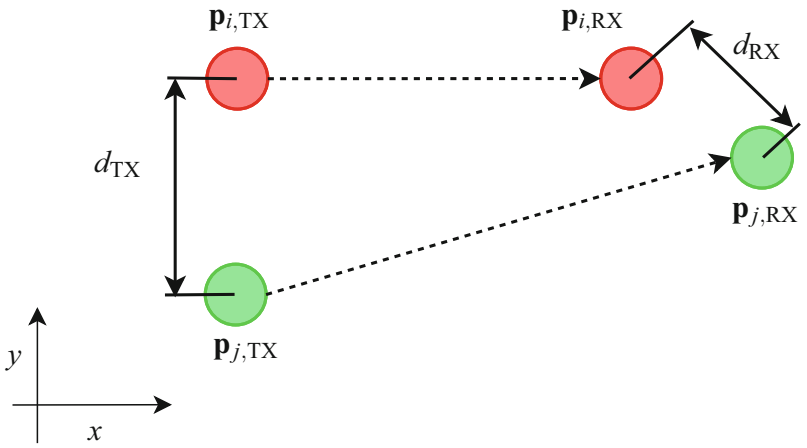

Fig. 10.5 Example of two links i and j in $\mathbb{R}^2$. The transmitter displacement is given by $d_{\mathrm{TX}} = \|\mathbf{p}_{i,\mathrm{TX}} - \mathbf{p}_{j,\mathrm{TX}}\|$ and the receiver displacement is $d_{\mathrm{RX}} = \|\mathbf{p}_{i,\mathrm{RX}} - \mathbf{p}_{j,\mathrm{RX}}\|$

10.4.3 Performance Evaluation

We consider a one-dimensional zero-mean field of 15 m length: $P_{\mathrm{RX}}(\mathbf{x}) : [0, 15] \times [0, 15] \to \mathbb{R}$. The true field is obtained from a 2-dimensional GP with (10.19), with parameters $p = 1$, $d_c = 3$, $\sigma_n = 0.01$, and $\sigma_\Psi = 10$. To ensure reciprocity the field is only generated for $\mathbf{p}_{\mathrm{TX}} \geq \mathbf{p}_{\mathrm{RX}}$ and then copying the values for $\mathbf{p}_{\mathrm{TX}} < \mathbf{p}_{\mathrm{RX}}$. The resolution of the field is 25 cm, corresponding to possible 40 TX and RX locations, which in total corresponds to 1,600 samples. For the training set $N = 250$ samples are randomly selected out of these 1,600 samples. This includes their reciprocal counterpart. They are then perturbed by location uncertainty. In accordance with [23], heterogeneous location errors have been considered with error variance σ_i. The error variance follows an exponential distribution parametrised by the average location error standard deviation λ. Assuming the agents know σ_n, the hyper-parameter vector is then given by $\boldsymbol{\theta} = [d_c, \sigma_\Psi, \sigma_{\mathrm{proc}}]^{\mathrm{T}}$. The covariance function of cGP corresponds to the Gudmundson model [13].

For parameter learning, a Monte Carlo simulation is performed over 30 realisations of the shadowing field.

10.4.3.1 Learning

The influence of location uncertainty of the training samples (parametrised by λ) in the database on the estimation of $\boldsymbol{\theta}$ for cGP (with $p = 1$) and uGP (with $p = 2$) is demonstrated in Fig. 10.6. For cGP, it can be observed that an increase of λ leads to an increase of d_c. Furthermore, the parameter responsible for capturing input

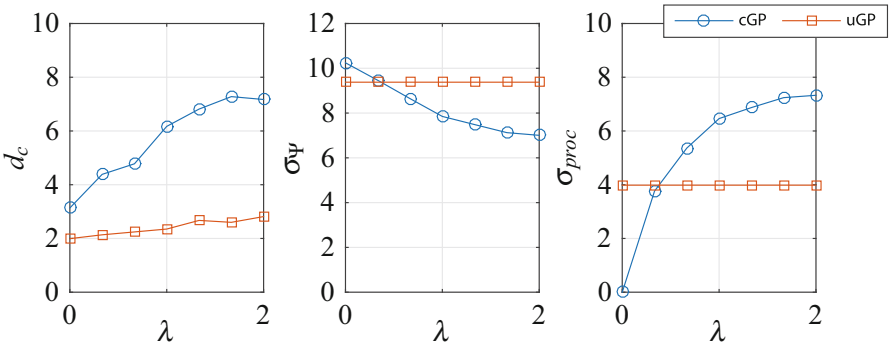

Fig. 10.6 Estimated hyper parameters for cGP and uGP for different levels of location uncertainty of the training samples parametrised by the average location error standard deviation λ in metres. *Left*: decorrelation distance d_c, *middle*: shadow standard deviation σ_Ψ, *right*: process standard deviation σ_{proc}

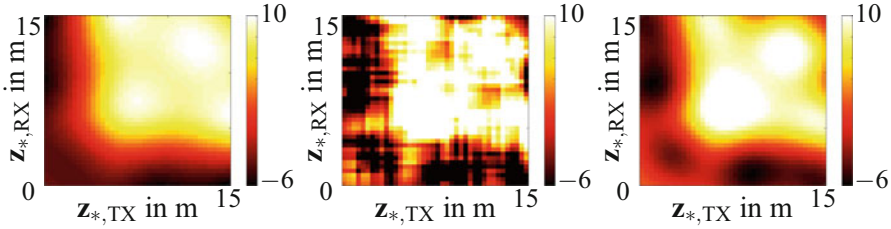

Fig. 10.7 *Left*: expected received power $P_{\text{RX,avg}}$ for one realisation of the spatially correlated shadowing field for different expected positions $\mathbf{z}_{*,\text{TX}}$, and $\mathbf{z}_{*,\text{RX}}$, respectively. The channel is symmetric reciprocal along the diagonal ($\mathbf{z}_{*,\text{TX}} = \mathbf{z}_{*,\text{RX}}$), i.e., the field is symmetric. *Middle*: prediction of $P_{\text{RX,avg}}$ using cGP. *Right*: prediction of $P_{\text{RX,avg}}$ using uGP. Note, cGP and uGP maintain channel reciprocity in their prediction

uncertainty σ_{proc} also increases. To maintain the total signal variance the parameter σ_Ψ needs to decrease. In contrast to this, uGP is able to properly incorporate different levels of location uncertainty with the effect that the estimated d_c remains almost constant independent of the location uncertainty of the training samples. The parameter σ_{proc} captures the mismatch between the covariance functions of the true field [where $p = 1$ in (10.19)] and uGP [where $p = 2$ in (10.19)]. This mismatch is independent of the location uncertainty of the training samples λ. In the right panel of Fig. 10.6, this can be observed by the estimated values of σ_{proc}. This means, since σ_{proc} is constant over λ so has to be σ_Ψ in order to maintain the total variance (see middle panel of Fig. 10.6).

10.4.3.2 Prediction

The prediction performance is shown in Fig. 10.7, assuming the database had no location uncertainty (i.e., $\lambda = 0$), but the received power should be predicted at a

location where a location uncertainty with standard deviation 2 m is present. The expected received power is expressed by $P_{\mathrm{RX,avg}}(\mathbf{s}_*) = \int P_{\mathrm{RX}}(\mathbf{x}_*)p(\mathbf{x}_*)\mathrm{d}\mathbf{x}_*$. Here, $p(\mathbf{x}_*) = \mathcal{N}_{\mathbf{x}_*}(\mathbf{z}_*, 4\mathbf{I})$. The mean value of the random location $\mathbf{x}_*$ is composed of the TX and RX mean location by $\mathbf{z}_* = [\mathbf{z}_{*,\mathrm{TX}}, \mathbf{z}_{*,\mathrm{RX}}]^{\mathrm{T}}$. The panels of Fig. 10.7 contain from left to right: $P_{\mathrm{RX,avg}}$ plot for different mean transmitter and receiver locations ($\mathbf{z}_{*,\mathrm{TX}}$ and $\mathbf{z}_{*,\mathrm{RX}}$), predicted expected received power with cGP, and with uGP. Observe that with both methods, cGP and uGP, channel reciprocity is maintained. Furthermore, the prediction of the true expected field with uGP is much better than with cGP.

10.5 Conclusion

The foreseeable emerging applications suggest that MAS are likely to become more prevalent in the coming years. Specific instances include intelligent transportation systems, networks of unmanned aerial vehicles (e.g., drones), networks of wheel-driven robots, or even a network comprised of heterogeneous agents. All these systems rely heavily on precise location information. This chapter treated how location information can be obtained, maintained, and disseminated over a network of agents. It also revealed a close connection between communication and localisation, and described a channel prediction strategy in the presence of location uncertainty. Furthermore, the coupling between communication and control and their dependencies on localisation have been demonstrated. Significant challenges remain, both to provide accurate location information and to operate MAS under limited, noisy or no information about the location of the agents.

Acknowledgements This work was financially supported by EU FP7 Marie Curie Initial Training Network MULTI-POS (Multi-technology Positioning Professionals) under grant nr. 316528. This work was also partially supported by the German project VaMEx-CoSMiC, which is supported by the Federal Ministry for Economic Affairs and Energy on the basis of a decision by the German Bundestag, grant 50NA1521 administered by DLR Space Administration, the EU project HIGHTS MG-3.5a-2014-636537, the European Research Council under Grant No. 258418 (COOPNET); and the DLR project Dependable Navigation.

References

1. P. Agrawal, N. Patwari, Correlated link shadow fading in multi-hop wireless networks. IEEE Trans. Wirel. Commun. **8**(8), 4024–4036 (2009)
2. A. Böttcher, P. Vary, C. Schneider, R.S. Thomä, De-correlation distance of the large scale parameters in an urban macro cell scenario, in *6th European Conference on Antennas and Propagation (EUCAP)* (2012), pp. 1417–1421
3. A.S. Chhetri, D. Morrell, A. Papandreou-Suppappola, On the use of binary programming for sensor scheduling. IEEE Trans. Signal Process **55**(6), 2826–2839 (2007)

4. D. Cohen, D.L. Jones, S. Narayanan, Expected-utility-based Sensor Selection for State Estimation, in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (IEEE, New York, 2012), pp. 2685–2688
5. K. Das, H. Wymeersch, Censoring for Bayesian cooperative positioning in dense wireless networks. IEEE J. Sel. Areas Commun. **30**(9), 1835–1842 (2012)
6. N.M.M. de Abreu, Old and new results on algebraic connectivity of graphs. Linear Algebra Appl. **423**(1), 53–73 (2007)
7. T. Eyceoz, A. Duel-Hallen, H. Hallen, Deterministic channel modeling and long range prediction of fast fading mobile radio channels. IEEE Commun. Lett. **2**(9), 254–256 (1998)
8. J. Fink, Communication for Teams of Networked Robots. Ph.D. thesis, Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA, August 2011
9. M. Fröhle, A.A. Zaidi, E. Ström, H. Wymeersch, Multi-step sensor selection with position uncertainty constraints, in *IEEE Globecom Workshops* (2014), pp. 1439–1444. doi: 10.1109/GLOCOMW.2014.7063636.
10. E.G. Garcia, L.S. Muppirisetty, E.M. Schiller, H. Wymeersch, On the trade-off between accuracy and delay in cooperative UWB localization: performance bounds and scaling laws. IEEE Trans. Wirel. Commun. **13**(8), 4574–4585 (2014)
11. A. Goldsmith, *Wireless Communications* (Cambridge University Press, Cambridge, 2005)
12. M. Grant, S. Boyd, *CVX: Matlab Software for Disciplined Convex Programming*, March 2014. http://cvxr.com/cvx
13. M. Gudmundson, Correlation model for shadow fading in mobile radio systems. Electron. Lett. **27**(23), 2145–2146 (1991). ISSN: 0013-5194. doi:10.1049/el:19911328
14. M.F. Huber, On multi-step sensor scheduling via convex optimization, in *2nd International Workshop on Cognitive Information Processing (CIP)*, June 2010, pp. 376–381
15. N. Jalden, Analysis and Modelling of Joint Channel Properties from Multi-site, Multi-Antenna Radio Measurements. Ph.D. thesis, KTH, Signal Processing, 2010, pp. xviii, 224
16. S. Joshi, S. Boyd, Sensor selection via convex optimization. IEEE Trans. Signal Process. **57**(2), 451–462 (2009)
17. Y. Kim, M. Mesbahi, On maximizing the second smallest eigenvalue of a state-dependent graph Laplacian. IEEE Trans. Autom. Control **51**(1), 116–120 (2006)
18. Y. Kim, G. Zhu, J. Hu, Optimizing formation rigidity under connectivity constraints, in *IEEE Conference on Decision and Control (CDC)*, 2010, pp. 6590–6595
19. G.J.M. Kruijff et al., Experience in system design for human–robot teaming in urban search and rescue, in *Field and Service Robotics* (Springer, Berlin, Heidelberg, 2014), pp. 111–125
20. M. Malmirchegini, Y. Mostofi, On the spatial predictability of communication channels. IEEE Trans. Wirel. Commun. **11**(3), 964–978 (2012)
21. F. Meyer, H. Wymeersch, M. Fröhle, F. Hlawatsch, Distributed estimation With information-seeking control in agent networks. IEEE J. Sel. Areas Commun. **33**(11), 2439–2456 (2015). doi:10.1109/JSAC.2015.2430519
22. F. Morbidi, G.L. Mariottini, Active target tracking and cooperative localization for teams of aerial vehicles. IEEE Trans. Control Syst. Technol. **21**(5), 1694–1707 (2013)
23. L.S. Muppirisetty, T. Svensson, H. Wymeersch, Spatial wireless channel prediction under location uncertainty. IEEE Trans. Wirel. Commun. **15**(2), 1031–1044 (2016). doi: 10.1109/TWC.2015.2481879
24. R. Olfati-Saber, Flocking for multi-agent dynamic systems: algorithms and theory. IEEE Trans. Autom. Control **51**(3), 401–420 (2006). ISSN: 0018-9286. doi:10.1109/TAC.2005.864190
25. G. Schirner, D. Erdogmus, K. Chowdhury, T. Padir, The future of human-in-the-loop cyber-physical systems. Computer **46**(1), 36–45 (2013)
26. M. Shamaiah, S. Banerjee, H. Vikalo, Greedy sensor selection: leveraging submodularity, in *49th IEEE Conference on Decision and Control (CDC)* (IEEE, New York, 2010), pp. 2572–2577
27. J.S. Shamma. *Cooperative Control of Distributed Multi-agent Systems* (Wiley Online Library, Chichester, 2007)
28. X. Shen, P.K. Varshney, Sensor selection based on generalized information gain for target tracking in large sensor networks. IEEE Trans. Signal Process. **62**(2), 363–375 (2014)

29. Y. Shen, H. Wymeersch, M.Z. Win, Fundamental limits of wideband localization. Part II: cooperative networks. IEEE Trans. Inform. Theory **56**(10), 4981–5000 (2010). ISSN: 0018-9448. doi:10.1109/TIT.2010.2059720
30. D. Simon, *Optimal State Estimation: Kalman, H infinity, and Nonlinear Approaches* (Wiley, 2006)
31. S.S. Szyszkowicz, H. Yanikomeroglu, J.S. Thompson, On the feasibility of wireless shadowing correlation models. IEEE Trans.Veh. Technol. **59**(9), 4222–4236 (2010)
32. S. Thrun, Y. Liu, Multi-robot SLAM with sparse extended information filers, in *Robotics Research* (Springer, Berlin, 2005), pp. 254–266
33. M.P. Vitus, W. Zhang, A. Abate, J. Hu, C.J. Tomlin, On efficient sensor scheduling for linear dynamical systems. Automatica **48**, 2482–2493 (2012)
34. Z. Wang, E.K. Tameh, A.R. Nix, Joint shadowing process in urban peer-to-peer radio channels. IEEE Trans. Veh. Technol. **57**(1), 52–64 (2008). ISSN: 0018-9545. doi:10.1109/TVT.2007.904513
35. J.L. Williams, J.W. Fisher, A.S. Willsky, Approximate dynamic programming for communication-constrained sensor network management. IEEE Trans. Signal Process. **55**(8), 4300–4311 (2007)
36. H. Wymeersch, J. Lien, M.Z. Win, Cooperative localization in wireless networks. Proc. IEEE **97**(2), 427–450 (2009). ISSN: 0018-9219. doi:10.1109/JPROC.2008.2008853
37. Y. Yan, Y. Mostofi, Impact of localization errors on wireless channel prediction in mobile robotic networks, in *IEEE Globecom, Workshop on Wireless Networking for Unmanned Autonomous Vehicles*, December 2013
38. M. Zavlanos, M. Egerstedt, G. Pappas, Graph-theoretic connectivity control of mobile robot networks. Proc. IEEE **99**(9), 1525–1540 (2011)
39. S. Zhang, R. Raulefs, Multi-agent flocking with noisy anchor-free localization, in *11th International Symposium on Wireless Communications Systems (ISWCS)* (2014), pp. 927–933
40. S. Zhang et al., System-level performance analysis for Bayesian cooperative positioning: from global to local, in *International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2013, pp. 1–10
41. S. Zhang, M. Fröhle, H. Wymeersch, A. Dammann, R. Raulefs, Location-aware formation control in swarm navigation, in 2015 *IEEE Globecom Workshops* (2015), pp. 1–6. doi: 10.1109/GLOCOMW.2015.7414165
42. S. Zhang, R. Raulefs, A. Dammann, Localization-driven formation control for swarm return-to-base application, in *European Signal Processing, 2016 IEEE/EURASIP Conference on (EUSIPCO)*, August 2016

Chapter 11
Positioning Technology Applications Related to Environmental Issues

Anna Kolomijeca

11.1 Introduction

The Earth's natural environment consists of climate, weather and natural resources. Unfortunately, human activities negatively influence all these three factors by extraction of natural resources and polluting the environment. Although nature is able to accumulate certain amounts of pollutants and many natural resources are renewable—this is an extremely time-consuming process. With human population growth of around 80 million people per year [72], which increases the need of natural resources and end product consumption, there is not enough time for nature to regenerate itself. Remote sensing of the environment is a very important application of positioning technologies. This aims to monitor/study climate, weather and natural resources in order to help authorities to implement successful environmental policy and management plans.

11.2 Definitions

Climate: specific environmental/weather condition of a region. The climate is characterized by temperature, air pressure, humidity, sunshine, clouds, precipitation and wind.

Natural resources: Renewable (can be naturally replaced) or non-renewable (cannot be naturally replenished in short time) materials that naturally exist on Earth that humans can use to make products. For example, food, medicine, energy and spiritual recreation. Natural resources can be further divided into two parts: biotic

A. Kolomijeca (✉)
Universitat Autonoma de Barcelona, 08193 Bellaterra, Barcelona, Spain
e-mail: anna_kolomijeca@inbox.lv

J. Nurmi et al. (eds.), *Multi-Technology Positioning*,
DOI 10.1007/978-3-319-50427-8_11

(consists of organic materials and living organisms, e.g. forests, animals, fossil fuels) and abiotic (consists of non-organic, non-living matter, e.g. land, water, air, metals/minerals/rocks and sunlight).

Pollution: Environmental contamination from man-made waste. Pollution can be in a form of chemical substance or energy (noise, light and heat). *Air pollution* is divided into primary (harmful chemicals which are directly emitted in the air during natural processes or human activities: CO, CO2, SO2, NO, NO2, CH4) and secondary pollutants (harmful chemicals react with each other and with other components of air to form new harmful chemicals: SO3, HNO3, H2SO4, O3, H2O2). It can be caused naturally (volcano eruption) or by human activates (transport, domestic combustion and industrial processes). *Soil pollution* is direct or indirect destruction of the Earth's land surface, including the change of soil's chemical and biological composition. This occurs due to man-made waste. In difference to natural waste (such as dead plants, animals and rotten fruits) which can be degraded naturally, man-made waste often consists of chemicals that are not originally found in nature (plastic) or not exposed in large amounts (toxic chemicals). Therefore it cannot degrade and it is changing the natural composition of the soil and causes various adverse effects. *Water pollution* is any change in the chemical, physical or biological properties of water, as well as the presence of external materials that naturally do not occur in water (for example, plastic waste), which results negatively on living creatures. The source of water pollution can be "point", "non-point" or "transboundary".

Soil erosion: natural process of removing topsoil by water, wind or excessive farming. Soil erosion process happens in three stages: soil detachment, movement and deposition. The process accumulates if land is not protected by vegetation, because plant roots do not hold it anymore. The topsoil is rich in organic matter and fertility. If it is relocated, the land eventually becomes desert-like and unable to support life. *Desertification*: persistent degradation of land by human activities (such as unsustainable farming, mining, deforestation and land misuse) and by climate change.

Remote sensing: data collection by detecting energy which is reflected from Earth. This is non-invasive technique based on *Reflectometry* principle. In remote sensing, sensors are not in direct contact with an object/event being observed; instead they are mounted on an aircraft or satellite. Passive and active sensors are two types of remote sensing. *Passive* sensors record natural energy (usually sunlight) that is reflected or emitted from the Earth. Human eye is an example of a passive remote sensor. *Active* sensors transmit a signal onto the object and measure the time that it takes for the laser to reflect back to sensor. An example of an active sensor is a dolphin's echolocation. Usually, remote sensors are installed on fixed-wing planes or helicopters, or on the satellites orbiting the Earth.

Geographic information systems (GIS): software packages, which are used to capture, analyse, manipulate and display of spatially referenced data.

Biodiversity: variation of all living organisms on Earth.

Invasive alien species: animals or plants, that are not native or originated from a specific region.

Marine debris: non-degradable fishing equipment, which is lost or dumped into the sea. It consists of fishing nets, buoys and ropes. Derelict fishing nets (ghost nets) can sink to the ocean's floor, float in the water column or on the surface. Usually ghost nets sink to the bottom of the seafloor, but sometimes they can be present in the water surface (with net area of a square metre or less), with a tangled knot of ropes, plastic netting and other materials descending downward in the water column for several metres [42]. Floating marine debris will be eventually washed up on shores, become entangled in the reefs endangering all marine animals as well as physically damaging the reefs.

11.3 Environmental Issues

The environment is a very complex, interconnected system. This means, that every human action, that interrupts natural environmental processes will have a direct or indirect consequence on nature and all living beings.

If we take **air pollution** as an example: today atmosphere concentration of carbon dioxide (CO2) is much lower than 4 billion years ago (today's 0.03% compared to 80%). Most of CO2 was removed from atmosphere by photosynthesis, e.g. it became locked in organisms and then into minerals such as oil, coal and petroleum inside the Earth's crust. By burning these materials, people "unlock" CO2 back to the atmosphere in a very short time and in large amounts [81]. Nature is not capable of recycling such high amounts of CO2 so quickly; therefore it is accumulating in the atmosphere. Together with other pollutants such as methane and harmful ground-level ozone (called greenhouse gases), they create a "blanket" around our planet that is capturing excessive sunlight and heat (which is supposed to be reflected back to space), causing ***global warming*** and earth ***climate change*** [81].

The loss of the forests further contributes to climate change, because trees are no longer capturing CO2. High amounts of CO2 pollution also cause ***ocean acidification***, which is very harmful to ocean environments and all creatures that live there [10]. Ocean acidification is a very serious ecological problem. It occurs because the ocean is constantly reacting with environmental gases, absorbing 1/3 of carbon dioxide from the atmosphere. An excess of CO2 in the atmosphere forces oceans to absorb larger amounts, which lowers the water PH value. As a result, fragile ocean ecosystems and biochemical cycles are disturbed. Ocean acidification has an extremely negative impact on reef building corals, shell marine animals, marine composite (calcifies) as well as changing the natural habitat of finfish.

Global warming is also "triggering" the ***release of methane*** from the ocean floor in the Arctic [69]. Naturally, methane carbon is stored in large amounts in the seabed due to high pressure and low temperatures, but when sea temperatures rise (as a consequence of global warming), the methane is released in the form of bubbles through the water into the atmosphere. This further adds to global warming.

Another problem arising from increasing temperatures is ***ice melting*** (apart from huge loss of fresh water resources) this negatively influences animals (e.g. polar bears have less time to accumulate fat due to shorter ice seasons [33]; and effects ecosystem productivity as well as timing of ecological events [51]. Additionally, it causes changes of the ocean salinity, currents and creates rising sea levels [51].

Climate change is "responsible" for a growing number of ***natural hazards***, including wild fires, floods, droughts, extreme weather, shifted rainfalls and others [84].

Global warming and climate change are not the only consequences of air pollution. The depletion of stratospheric upper *ozone layer* is an important problem. The ozone layer is around 50 km thick, but when it is compressed by sea-level pressure it will become only a few centimetres [21]. Never the less, it plays a very important role in shielding the Earth's biosphere against solar UV-radiation and filtering harmful sunlight. The ozone layer is thinning due to the man-made release of chemicals called chlorofluorocarbons (CFC). The adverse effects of ozone layer depletion include direct harm to human health (skin cancer, eye damage and immune system); adverse impacts on agriculture, forestry and natural ecosystems (as vegetation is UV sensitive); damage to marine life (especially plankton); adverse effects on animals (similar to humans, they get cancer and eye damage); even materials, such as wood, rubber and fabrics degrade from UV light [73] .

Acid rain is another negative consequence of air pollution and an example of pollution migration. Acid chemicals can be incorporated to the dust and fall to ground in the form of dry deposition. Alternatively, harmful air pollutants can be mixed with moisture in the atmosphere, than it drops down to Earth in the form of acid rain—a wet deposition. Although acid rain looks, feels and tastes like a clean rain, fine particles negatively influence human health, causing heart and lung disorders [17].

Acid rain also contributes to **soil pollution**, disturbing it's ecological balance. Other causes of the soil pollution are from agricultural activities (modern pesticides and fertilizers are full of chemicals, that are not produced by nature and cannot be broken down naturally), industrial activities (by-products from mining and manufacturing), waste disposal/landfills, mining, land misuse and accidental oil spills [41]. Most plants cannot adapt to rapid changes of the chemical structure of soil. Soil contamination decreases soil fertility, causes ***soil erosion*** and eventually leads to ***desertification***. Land pollution causes a loss of 24 billion tons of topsoil every year [23]. Pollution from soil maybe absorbed by plants or can migrate further into ground water, contaminating it. Polluted ground water will slowly move further into lakes and rivers, and eventually reach the sea.

Sea water pollution is a serious issue. Some toxic chemicals like persistent organic pollutants (POP) and some heavy metals do not break down in water but ***bio-accumulate*** inside fish and marine animals [38, 47]. Later, when people eat the fish, those toxins produce various adverse health effects, including cancer, endocrine disruption, immune and reproductive dysfunction [26]. Besides toxic chemicals, nuclear and industrial waste is also often discharged directly into the ocean [31].

Additionally, many tons of plastic is dumped into the sea every year. The latest research estimate over 150 million tons of plastic is currently (2016) floating the ocean. Studies show [79] that by 2050 there will be more plastic in the ocean than fish. Over 260 different species are known to suffer from entanglement or ingestion of marine plastic, this includes turtles, sea lions, seabirds, whales, seals and fish [2]. Plastic does not degrade in the water, instead it often breaks into smaller pieces and enters the food chain as micro plastic [27].

Fisheries add to the problem by losing/dumping their fishing gear. An incredible amount of lost fishing nets (ghost nets) continue to float in the ocean, trapping marine animals and killing them. This process is called ***ghost fishing*** [22]. According to [82], up to 1000 km of ghost nets are released into Pacific Ocean alone each year. Ghost fishing and ***overfishing*** strongly contribute to **natural resources depletion**. Fisheries, including illegal fishing, catch higher capacities of fish than can naturally reproduce. Often, huge nets catch other sea animals (like turtles, sharks and seabirds): this is called by-catch [54]. ***By-catch*** is often destroyed and thrown back at sea. Fishermen are using large fishing gear that can completely destroy the seabed and coral reefs, heavily damaging delicate marine environments.

Marine life is amazing in its beauty and has incredible economic value. By destroying it, people further contribute to ***biodiversity loss***. Many animal and plant species are irreversibly lost, together with its unique biological and genetic structure that could potentially be very valuable to humans (for example, when developing new medicine). *Poaching, invasive alien species and the loss of natural habitat are the greatest treats to biodiversity* [65]. When trees are cut down—the natural habitat of animals is destroyed; they have nowhere to go and no food source. Eventually the animals die of starvation. People also depend on forests directly (food, biomass and recreation) as well as indirectly (climate control, flood control, storm protection and nutrient cycling).

Additionally, trees and other plants bind the soil and hold its nutrients as well as protect soil from overheating (trees reflect 20% sun heat). When trees/plants are removed, e.g. the area is ***deforested***—the topsoil is exposed to direct sunlight and rain, the nutrients are washed away and soil erodes. Another precious natural resource is fresh water. All fresh water on earth make only less than 3% of the total water amount, most of this 3% is not easy accessible because it is trapped in snowfields and glaciers. Fresh ***water scarcity*** is a serious issue, especially in developing countries: clean drinking water shortage is leading to 250 million cases of water-related disease each year and 5–10 million death [29]. Apart from the direct negative health effects of this, water scarcity contributes to hunger problems by directly affecting the food chain because plants and animals need water to survive. Therefore, efficient water management should be a top priority nationally and internationally. A short summary of the environmental issues is presented in Fig. 11.1.

The majority of environmental issues, which was discussed in this current chapter, can be fully or partly addressed with GIS and remote sensing.

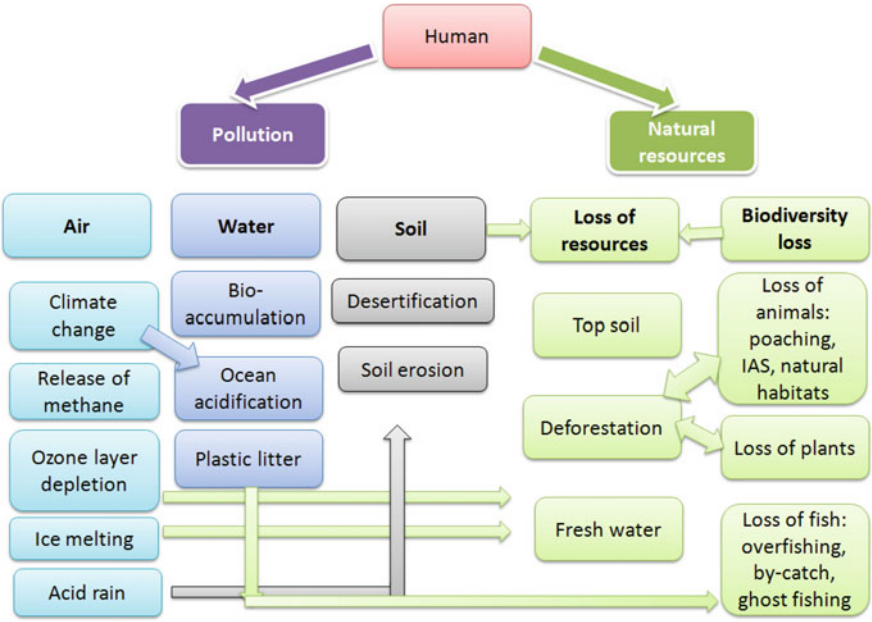

Fig. 11.1 Environmental issues

11.4 Geographic Information Systems

National Geographic society defines GIS as a computer system, which is designed to capture, check, store and display data related to position on Earth's surface [50]. Any kind of information which includes location can be entered into GIS in order to produce a map, which will highlight an "event of interest". GIS systems are able to evaluate information of different types, from different sources and of different original format. The main rule is to keep the same **scale** and **projection**. Scale presents a relationship between the distance on a map and the actual distance on Earth; while projection is a method of transferring information from Earth's curved surface to a flat piece of paper or computer screen (no projections are perfect, for example, a world map can show either correct size of countries or their correct shape, but cannot do both) [50]. The decision of how to collect and incorporate data into one system is a first step in creating GIS map. It is very important to decide which kinds of data are needed, for which purpose and how to obtain them with minimum cost and afford. This step is called ***pre-processing***. The next step of creating GIS map is actual ***data capture***. "GIS lounge" [16] proposes four methodologies for capturing GIS data:

- D*digitalizing*/scanning data from previously printed maps. Digitalizing process can be manual or "heads up". For details, see [16].

- *Surveying* methodology consists of three methods to capture GIS data: Coordinate geometry (COGO), Geocoding and Global Positioning Systems (GPS) method. For details, see [16].
- *Remote sensing* captures various GIS data. The sensors are usually attached to a satellite or aircraft.
- *Photogrammetric* technique is photo interpretation into GIS data.

Depending on GIS data presentation form, it can be divided into two categories: attribute tables (represented in tabular formats) and spatially referenced data (represented by vector and raster forms). The final part of data capturing is to provide data with the right coordinate system, projection and possible data changes. This final step is called ***data processing***. The process of removing different errors from the captured data is called Quality Assurance/Quality Control. Once all the data are entered into GIS system, it is possible to make different manipulations in order to produce the map "of interest" or different maps with individual data layers. Schematic representation of GIS data capture steps is presented in Fig. 11.2.

GIS and particularly remote sensing techniques are powerful and very efficient tools for Earth observation and environmental assessment studies due to the following reasons [1] and [77]:

- It records data from large areas (entire surface of Earth) in short time;
- It stores multidisciplinary data sets;
- It identifies complex interrelationship between environmental characteristics;
- It evaluates changes over time due to measurement repetition;
- It can be continuously updated and used for several projects;
- It serves as data set of various mathematical models;
- It stores and manipulates 2D and 3D files;
- It serves the interest of general public and technical analysts.

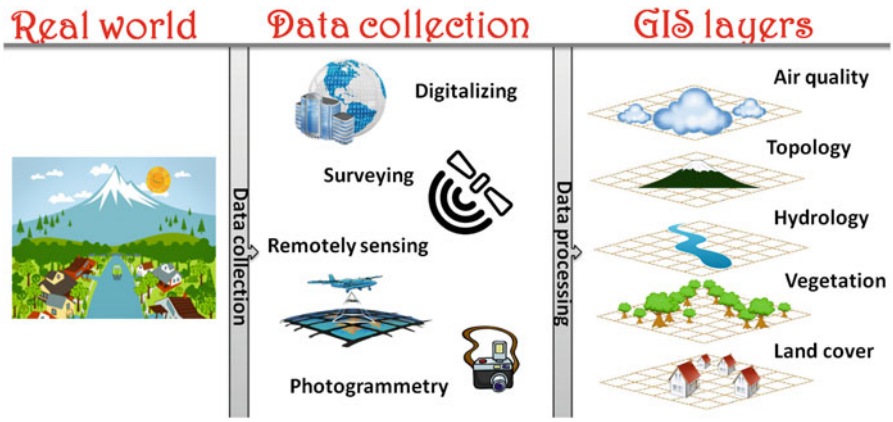

Fig. 11.2 Example of GIS layers

Satellite observation data over land, atmosphere and oceans provide valuable support to environmental policy makers to protect environment on the global scale, reduce losses from disaster and sustainably manage natural resources. The main objectives of GIS/remote sensing in environmental impact assessment are: to monitor and model Earth's surface processes as well as their interaction with atmosphere; to measure and estimate biological, geographical and physical variables; to identify details and materials on Earth's surface; as well as analyse spectral signatures which are obtained by satellite sensors [77]. Particular applications of GIS/remote sensing as well as data capture examples are described in the following paragraph.

11.5 Example of Environmental Applications of the Satellite Remote Sensing

11.5.1 Land Monitoring

The following information is taken from [32] unless stated otherwise. Satellite remote sensing provides consistent measurements of landscape conditions. Although field surveys are generally more accurate and provide better spatial/temporal/thematic resolution than remote sensing, they are more costly, slow, cover limited geographic area and not as frequent [32]. For local studies, which have a specific task over a small geographical area, field sampling may provide better results in terms of accuracy and data quality, but may be very expensive and/or not possible over large areas. Remote sensing can cover this gap. In fact, it is advised to use a combination of both techniques: field measurements and remote sensing in order to achieve the best results [32]. A big advantage of satellites is that they stay in the sky for several years, therefore it is possible to monitor landscape changes over time, e.g. makes repeated observations of the earth's surface. Thanks to remote sensing, resource managers can monitor landscape dynamics in a cost-effective way over large areas (including difficult to access or hazardous regions). It is possible to detect the stressors and drivers of the environmental changes, which allows managers to target causes of that change rather than simply observe consequences [36]. The process of interest can be addressed by measuring specific features that is causing or corresponding to this process. An example of correspondence between the resource indicator and the process of interest is presented in Table 11.1 (source: [36]).

Remote sensing change detection involves four major steps: data acquisition, preprocessing and/or enhancement, analysis and evaluation. For details, see [36]. The potential roles for RS in monitoring of protected areas, see Table 11.2 (source [32]).

Table 11.1 Natural resources management

Resource indicator	Process of interest
Variations of size and shape of patches or related cover types	Expansion of vegetation, infilling or penetration, erosion
Change in width or character of narrow, linear features	Visitor use of path or roads, flooding effects on stream vegetation; dynamics of terrestrial and submerged near-shore aquatic vegetation
Slow change in cover type or species composition	Succession, competition, eutrophication, exotic species invasion
Abrupt change in state of cover	Disturbance, human-mediated development, land management
Slow change in condition of a single cover type	Climate-related changes in vegetative productivity, slowly spreading forest mortality caused by insect or disease, changing in moisture regime
Changes in timing or extent of seasonal processes	Snow cover dynamics, vegetation phenology

11.5.2 Ocean and Coastal Zones

Remote sensing provide valuable data about ocean circulation/current systems, ocean temperature/wave heights and sea ice state. The data are used to better understand the oceans and how to best manage ocean resources [53]. Remote sensing can also help in mapping and assessing certain marine and coastal habitats, such as coral reefs, sea grasses, kelp beds, mangroves and sea ice. It is more beneficial to use RS in tropical areas in shallow waters, because the water is clearer. Observations usually include *coral reefs* (which provide protection from tropical storms, reef fisheries, tourism, pharmaceuticals), *sea grasses* (important source of food for various marine animals, provide habitat for many fish, filter coastal waters) and *mangroves* (reducing impacts of natural disasters, as tsunamis and hurricanes). It is possible to detect long-term threatening processes, for example, urban development in coastal areas, aquaculture, ocean temperature variations, river discharge; as well as short-term threats, for example, oil discharge from a vessel or algal blooms. The remotely observed indicators for marine and coastal habitats include: size of coral reef ecosystems, percent cover of living coral, coral bleaching, extent of sea grass and mangrove ecosystems, mangrove habitat conversion, natural hazards influence on mangrove extent, biomass of mangroves and others. For details, see [32].

Table 11.2 RS monitoring of protected areas

Resource indicator/focus of monitoring	Process of interest/potential role of remote sensing
Land cover change and encroachment	Monitoring of illegal activities on protected areas (for example, extraction of resources in and around protected areas)
Infrastructure development	Help reducing the impact of road expansion and any other developments in protected areas
Habitat fragmentation	Identify, quantify and monitor habitat of surrounding area; where feasible, identify fragmentation metrics relevant to target species or ecosystem
Livestock grazing	Monitor "greenness index" or some other indicator of grazing pressure and range condition
Invasive species	Identify (possibly), map and monitor distribution of invasives; track related changes in several stages (development stages of ecological communities) and loss of aquatic habitat
Special habitat	Monitor habitat area for threatened species
Plant succession	Monitor natural plant community succession after natural or anthropogenic disturbance
Recurring drought, catchment area changes, erratic release of water	Monitor alteration in habitats and invasion of woodland in wetland
Fire hazards	Identification, mapping and document extent of burns; map several stages of the habitat; MODIS Rapid Response alert system; fire history; burn mapping
Contamination of water with pesticides, fertilizers, sediment, etc.	Monitor agricultural practices and land use pattern in catchment area
Pollution	Monitor location and movement
Erosion	Identify, map and monitor riverbank position
Diseases and pests	Habitat and vector modelling and prediction (advanced)
Climate change	Monitor land area of low-lying islands; quantify coral bleaching extent

11.5.3 Biodiversity

Biodiversity can be assessed by applying one of two general approaches: *direct* remote sensing (individual organism mapping, species clusters or ecological communities by use of airborne or satellite sensors) and *indirect* remote sensing (analysing general land cover, elevation, geology, landform, human disturbance, or even mapping the threats (e.g. the areas which are avoided by species). The species population indicators, being measured by remote sensing, include: location and extent of species aggregation, as well as modelled/direct/or indirect extent

of specific habitats. In some cases it is possible to directly monitor individual species (like orangutan nests, elephants, some Australian mammals [32]); however, the majority of species cannot be mapped/identified or counted directly from the satellite or aerial imagery due to several reasons, for example, the size of species in relation to sensor resolution. It is more common to use telemetry-based methods to study individual species (see following chapter). Engaged animals can be better protected by mapping poaching risk areas. For example, [32, 46] produced rhinoceros poaching risk map by analysing the distance which hunters are travelling from settlements to the park, the frequency, the location of arrested illegal hunters and the rate of animals taken off the park.

Invasive alien species (IAS), which is one of the biggest treats to biodiversity, can also be mapped applying GIS and RS. It provides a cost-effective, large-scale and long-term documentation about specific indicators for IAS, such as: area, distribution and trends in particular IAS; prediction of distribution of IAS; indirect identification of areas vulnerable to invasion; identification of potential sources of invasion. For example, roads, off-road vehicle trails, hiking and horse trails are potentially vulnerable areas for IAS because seeds from the plats can be transported by vehicle tires or by foot. Distinguishing vegetation IAS is based on unique biochemical (specific spectral characteristics), phonological (timing of image acquisition) and structural (spatial pattern) characteristics.

11.5.4 Telemetry Techniques for Monitoring Animals

The International Union for Conservation of Nature [34] separates five major threats to biodiversity: Habitat loss and degradation, Invasive Alien Species, Over-exploitation of natural resources, Pollution and diseases, as well as Human-induced climate change. Positioning technologies have been widely applied to carry out wild life surveillance, health and mortality investigations, as well as mapping of the natural habitats [24, 43, 70]. In order to study wild animals, locating and tracking is of great importance. This can be done by attaching a special tag on the animal's body or inside the body, by implantation or ingestion (the tag is fed to the animal) [60]. Depending on the information from the tag is stored or sent, the tags are *archival* or *transmitting* (accordingly).

Archival tags are very small (available market solutions can be less than 0.39 g [64]) and can store several years' worth of data on a very fine scale. They contain a microcomputer, which record and store the data about the animal [such as heart rate, swimming speed (if in the water)] as well as its environment (such as light level and water salinity). In order to download the data from archival tags, they must be recovered, which usually means recapturing the animal, unless combined with satellite technology.

Transmitting tags are developed based on GPS/satellite technology, Very High Frequency radio tracking, acoustic tracking or Passive Integrated transponder (PIT). A short overview of these techniques is provided below.

11.5.4.1 GPS Tracking

Global navigation satellite system is a major positioning technology, which is well suited for outdoor applications. It has been a perfect tool to track wild birds and animals [61]. Global positioning system (GPS), sometimes in combination with other environmental sensors or automated data-retrieval technologies, allows remote observation of fine-scale movement or migratory patterns of wild birds and terrestrial/marine animals. A GPS receiver has to be placed on the animal. The receiver will calculate the location and movement of the animal. The collected data will be stored on the device until retrieved by recapturing the animal, or by remote data extraction applying the Argos satellite system (see below). Additionally, a drop-off unit may be deployed (a good option if recapturing the animal is problematic, for example, marine animals [44]).

FastlockTM: is a technology, allowing to capture GPS signals in a fraction of a second [40]. It is widely used for tracking marine and fresh water animals, which comes on the surface only for short time [64].

11.5.4.2 Argos Satellite Tracking

Argos is a scientific satellite system, which was designed for collection, processing and dissemination of environmental data from fixed and mobile platforms [5]. Typically, the transmitter of a radio signal is placed on the animal's beak. A transmitter sends animal movement information as radio signals to the satellite. In difference to GNSS (where location is calculated based on triangulation method, see [7]), Argos obtains objects location by calculating Doppler frequency shift, for details, see [5]. It was established in 1978 under an agreement between French Space Agency (CNES), National Oceanic and Atmospheric Administration (NOAA, USA) and National Aeronautics and Space Administration (NASA, USA). There are 6 operational Argos satellites (data from 2014, for details, see [4]): 4 × Argos-2 and 2 × argos-3 which process information from over 30,000 platforms tracked (which is about 3 M messages per day and 500 Tb data per month), from 8000 animal species, in over 100 user countries [5]. Besides wildlife tracking, Argos system is used for in-situ data collection in climatology, oceanography, marine fisheries and offshore (monitoring pollution and marine equipment) [5].

11.5.4.3 Very High Frequency Radio Tracking

In difference to satellite tracking methods, conventional radio-telemetry tracking has much lower costs and lightweight transmitters that allow studying almost all mammal species with minimum impact on the animal [35]. Additionally, VHF doesn't require line of sight, so trees and other obstacles are not a problem. VHF transmitters emit radio-frequency signals, which are detected by a researcher using a receiver and antenna. This is manual work, therefore it is limited by intensity and

scale that data can be collected, plus, the researcher has to be within a certain range of the animal. The accuracy of the method depends on local conditions, instrument used and skills of the operator [70].

11.5.4.4 Acoustic Tracking

Acoustic telemetry is very convenient for monitoring marine life because of good sound wave propagation in water. Acoustic tags are sound emitting devices, which are placed outside or inside the animal's body [62]. The signal is picked up by a "listening device", called a hydrophone. The main limitation of this technique is that it requires a relatively close distance between the tracked animal and the hydrophone.

11.5.4.5 Passive Integrated Transponder

PIT technology is commonly used for tagging pets. This type of tag contains a microchip (size of a rice grain) with an identification code. Usually it is implanted in the animal's body and a tag reader will identify the code. This code is registered in a computer database with information about the pet and its owner. The PIT technique is range limited.

11.5.4.6 Combination of Tracking Technologies

In order to enhance tracking performance and overcome different limitations, it is possible to combine different technologies into one receiver. Examples can be found in [64] and [62].

11.5.5 Illegal Fishing

Satellite technology and remote sensing has been widely used to estimate fish location [37]. It helps to save cost on fuel and time to find good fishing spots. Luckily, GIS and RS can also be applied to monitor limited fishing resources, especially illegal fishing activities. The following information is taken from [14]. There are several fishing rules, which are mandatory for fishing vessels in EU: fishing quotas, forbidden species, time of the year and certain fish size. Fishery inspections can take place in-situ, e.g. at sea (using patrol vessels or aircrafts) or at the port. Inspections at landing ports can detect presence of undersize fish, forbidden species and excess of quotas, while in-situ inspections are mainly used to detect illegal fishing. There are three ways to detect illegal vessels at sea:

1. ***Vessel monitoring system (VMS)***: includes GPS-position transmitting device, which is mandatory for all EU fishing vessels above 12 m in length. The position of the vessel is sent to national Fisheries Monitoring Center at regular intervals. The disadvantage of VMS is that this GPS device can be manually switched off, malfunction or manipulated to show wrong position.
2. ***Vessel detection system (VDS)***: is able to compensate the disadvantage of the VMS, because it allows independent identification of the vessel positions through the analysis of satellite synthetic aperture radar (SAR) imagery. Image delivery can vary from just a few minutes to 1 h, dependent on the capability and geographic location of the satellite receiving station. Since radar systems do not depend on weather conditions and light presence, it is an ideal tool to detect non-cooperative vessels under most conditions—day, night and through clouds. The vessel position detected by VDS is then cross checked with vessel position, which is provided by VMS, making it possible to flag "VMS-hidden" vessels or confirm vessel location.
3. ***Air patrol*** surveillance can be directed towards areas where suspected targets are identified.

Cicuendez Perez et al. [14] suggest that joined applications of all three approaches will not only improve effectiveness of fishery control, but also provide a higher cost efficiency than that of using air patrol surveillance alone (in terms of reducing fuel costs).

11.5.6 Poaching Combat

According to [45, 83], thousands of endangered animal species are killed by poachers every year. There is a high demand and market for exotic animals and parts of their body: thousands of species are captured or hunted illegally and sold as food, pets, leather and medicine. High revenues (particularly from elephant / rhinos ivory, as well as valuable large cat's skins) allow poachers to be very well equipped and organized. In order to monitor and minimize poaching activities, it is essential to use a combination of GIS, RS and spatial logistics. Sibanda et al. [63] present a detailed overview of the geographical areas with the highest poaching activates. The author discovered that areas with high vegetation cover and close distance to waterholes are the most attractive to poachers. Merchant [45] proposes different technologies, which are capable of detect poachers in the field:

- ***Heat-sensing plane***: deployed in Kruger National Park, South Africa, where it is estimated that over 500 rhinoceroses are killed annually. The heat-sensor place on the low-speed recon plane detects animals and humans.
- ***Hidden remote sensors***: electronic sensors, called TrailGuards, can be buried underground or hidden amongst trees. These sensors consist of metal detectors, which become activated if poachers (who are equipped with rifles and machetes) pass by. The signal with location information from the sensor will send a radio

signal to treetop antenna and will be detected by park rangers. While the rangers are equipped with special transponders, that tell detectors who they are (similar to identification "friend-or-foe" systems for military applications).

- ***Drones*** (unmanned aerial vehicles): is very efficient anti-poaching technology because a drone can provide real-time location information about the poachers and help to "scare" criminals, who know that they are being located and recorded. Additionally, drones are relatively inexpensive, fast recharging, lightweight and are able to travel a pre-programmed route. Applications of drones in Nepal helped to decrease rhino poaching from 1 rhino/per month to 1 rhino/per year.

11.5.7 Dealing with Natural Hazards

Geophysical natural hazards tend to occur in specific geographic "hotspots" [75], for example, earthquakes occur along active tectonic plate margins, volcanoes—along subduction zones, tsunamis—in the neighbourhood of active plate margins, tropical cyclones—in particular zones along coast lines and landslides—in hilly and mountainous regions. Geographic Information Systems and Remote Sensing play a crucial role in risk management and natural hazard assessment [56] by collecting, organizing and analyzing environmental, meteorological and geographical data units. The type of information to be collected by GIS for hazard management depends from the level of application, e.g. national, regional or local scale. Generally, GIS presents *natural hazard* information (including location, severity, frequency and probability of occurrence of hazardous events); information on *natural ecosystem* (estimating the effect natural hazards can have on the goods and services in this system); information about *population and infrastructure* (quantifying the impact natural events can have on existing and planned development activities) [56]. This generated "mapped" information helps to reduce/avoid the losses from hazards, insure fast and appropriate assistance to victims of disaster and effectively recover after the event.

11.5.8 Marine Debris

There is an incredible amount of derelict fishing gear (DFG) in the ocean. "Ghost nets" entangles marine life and kills it. According to [76], between 1996 and 2006, NOAA recovered a total of 511 metric tons of fishing gear from the reefs of the Northwest Hawaiian Island Marine National Monument, with an estimated accumulation rate of 52 metric tons. During 2009, clean up volunteers [55] found 18,000 derelict fishing nets in the United Kingdom alone, and 70,000 plastic bottles in Nicaragua. To the date (2016) there is no technique to detect ghost nets on the ocean floor, which the author is aware of. Therefore this paragraph will only address floating DFG which is on the ocean surface.

Detection of DFG is a very challenging task due to relatively small size (square metre or less on the surface), large area of the ocean and constant movement of the DFG (depending on weather conditions). No single technology or solution will enable detection and to clean it up. It is essential to apply an integrated system approach [42] as well as a combination of different sensors [76]. Mance from NASA D.F.R. Center [42] suggests implementation area sampling strategy to create a DFG distribution model. Spatial sampling can be systematic or stratified (samples are taken in each stratum in proportion to its area). It is essential to have prior knowledge of the DFG concentration areas in order to enhance sampling efficiency. It is possible to estimate the location of marine debris by using current and wind information. The total drift of an object is calculated from surface currents and wind using "wind-surface-speed coefficient" [58]. This is an example of *indirect* measurement of marine debris. After the DFG distribution model is created, it is necessary to confirm the location of DFG by *direct* measurement. Taking into account large-scale search areas, remote sensing is the only technique which is capable to carry out this task. Regarding DFG detection sensors, a successful combination of radar and imaging sensor is reported by [42, 58, 76]. In detail, T.H. Mace [42] is proposing to use models, satellite radar and multispectral data, followed by airborne remote sensing (particularly radar), to focus the search on eddies and convergent fronts in the open ocean. Pichel et al. [58] carry out marine debris detection in seven stages:

- Obtain insight from historical surveys and use drift models for survey planning;
- Study drifting buoy tracks to understand current patterns and areas of convergence;
- Develop GIS techniques for analysis of satellite, buoy and meteorological data;
- Use satellite imagery to locate regions of convergence;
- Test and refine aircraft sensors and observation strategies;
- Assess wind and wave conditions to select optimum flight tracks;
- Flying aircraft survey.

The final search stage includes visual debris detection from a small UAS, low-altitude aircraft or ship. While Veenstra et al. [76] propose to apply a multi-or hyperspectral imaging with an automated detection algorithm operating in real time and a lidar that would be aimed at targets detected by that algorithm.

11.6 European Space Agency: Earth Observation Programs

All information is taken from [25], unless sited otherwise.

In 1977 ESA launched its first meteorological satellite "Meteosat", thus, starting to observe the Earth's environment. Later, in mid-1990s, ESA started to develop a complex, full-scale earth observation program, known as "**The living planet**" program. "The living planet" is dedicated to address key scientific environmental challenges, which are identified by the science community. It consists of eight "*Earth Explorer*" missions and six "*Earth Watch*" missions (see Fig. 11.3).

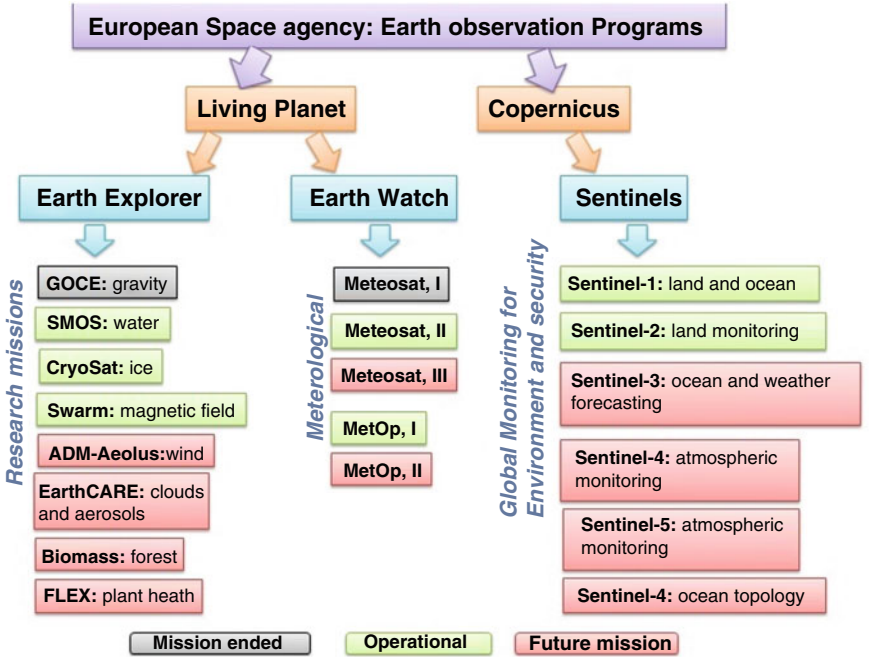

Fig. 11.3 ESA Earth Observation programs, 2016

The *"Earth explorers"* are scientific missions, which consists of:

- **Gravity field and steady-state Ocean circulation Explorer (GOCE) mission**, launched in 2009—ended in 2013. The aim of the mission was to detect gravity-field anomalies (force of gravity on the surface of earth is not a constant value, but varies due to its rotation, position of mountains, ocean trenches and variations of density of Earth's interior). Gravity force drives ocean circulation, sea-level change, variations of ice-cover and other natural processes. For example, ocean circulation is a major force, which forms Earth's climate, because it regulates heat exchange. Therefore, in order to understand Earth's climate and how its changes, it is essential to measure its gravity field. Another example of where GOCE data can be applied is studying earthquakes and volcanic activity. This is possible because gravity measurements reflect density variations in the Earth interior. Together with seismic activity data, it can be used to study processes in lithosphere and upper mantle.
- **Soil Moisture and Ocean Salinity (SMOS) mission**, launched in 2009—it was planned to end in 2012, but the mission is prolonged until 2017. SMOS makes global observations of soil moisture and ocean salinity. The collected data helps to understand water exchange processes between Earth's surface and atmosphere, thus, improving weather and climate models. Some examples of SMOS data applications are: measuring floating ice on the sea surface to plan ship routing;

providing global images of the surface-soil moisture, which can be used in agriculture; helping to understand climate change as warmer climate modifies water cycle; data about ocean salinity are needed to understand ocean circulation (because ocean salinity depends evaporation, precipitation, river runoffs and ocean dynamics); as well as other applications.

- **CryoSat mission**, launched in 2010, ends in 2017. It is dedicated to monitor cm-scale changes in floating sea ice, as well as ice sheet thickness in Greenland and Antarctica. In difference to icebergs (that are broken from edges of ice-shelves), sea ice is frozen sea water. Freezing and melting sea ice is a natural seasonal process (in the North Pole, the area with size of Europe melts and freezes every year). When sea ice is melting, large amounts of fresh water flow into the ocean, reducing water salinity and density. This influences global ocean circulation, consequently forming Earth's climate. Therefore, CryoSat data about sea ice thickness is essential for understanding climate change and policy-making.
- **Swarm mission**, launched in 2013, ends in 2017. Mission consists of three satellites, which accurately measure the strength and direction of Earth magnetic field. The magnetic field of Earth is formed from charged particles and acts as a protecting shield from cosmic radiation and charged particles from solar winds. It is estimated that every 200,000–300,000 years Earth's magnetic field changes its poles form South to North. Pole reversal is a natural phenomenon, which is supposed to happen again soon. Scientists confirmed that in the last 150 years, our magnetic field weakened by 15%. Also since 2001 magnetic North Pole is moving (away from "true North") with a speed of 65 km/year. This raises serious concerns. Therefore Swarm data is essential to understand Earth interior and how it works.
- **Atmospheric Dynamics Mission (ADM-Aeolus)** is due for launch in 2017. It will provide accurate global wind measurements, which can be used for weather forecasting. The mission will provide about 100 wind profile/h for the entire planet, which is an important data for weather prediction.
- **Earth Clouds Aerosols and Radiation Explorer (EarthCARE) mission** is due for launch in 2018. It is developed in cooperation with Japan Aerospace Exploration Agency, JAXA and dedicated to observe clouds, aerosols and measure radiation. The incoming solar radiation is partially absorbed by Earth and partially reflected back to space. Greenhouse gases, clouds and aerosols create a "blanket" around our planet, which traps heat. Excessive heat and radiation do not escape back to space. This is the main cause of global warming. The aim of the EarthCARE mission is to collect data about clouds, aerosols and the radiation at the top of the atmosphere. This data will help to predict climate and improve weather forecasting.
- **Biomass mission** will be launched around 2020. It aims to provide information about the state of the forests and how it's changing. Since plants absorb, store and release large amounts of carbon, the data from Biomass mission will help to understand Earth's carbon cycle. Additionally, it will provide data about habitat loss and its impact on biodiversity.

- **Fluorescence Explorer mission (FLEX)** will be launched after 2020. It will globally map the steady-state chlorophyll fluorescence in vegetation in order to quantify photosynthetic activity. This information will identify plants health and stress, which will lead to improved understanding of the carbon cycle and food security.

The "**Earth watch**" is created in partnership with European Organization for the Exploitation of Meteorological Satellites (Eumetsat). It addresses meteorological needs and consists of five missions: three generations of geostationary MeteoSat and two generations of polar-orbiting MetOp satellites.

The fist Meteosat aircraft was launched in 1977. Since then new aircrafts continuously provide meteorological data and are smoothly replaced by newer satellites. Second generation MeteoSat was launched in 2002 and became operational in 2004. Third generation MeteoSat will be launched in 2019.

Polar-orbiting MetOp satellites (from Meteorological Operational satellite programme) carry state-of-the-art imaging and sounding instruments, which provide valuable atmospheric data for meteorologists and climatologists. MetOp program is established in cooperation with US National Oceanic and Atmospheric Administration (NOAA). The first MetOp-A satellite was launched in 2006; MetOp-B in 2012, MetOp-C in 2016. This way, new satellites provide complimentary information or continuously replace old ones. Second generation of MetOp satellites will be launched after 2020.

In addition to the "Living Planet" program, ESA is developing "Copernicus" program, which is the most ambitious Earth observation program to the date. Therefore, "Copernicus" will be given detailed attention in the following paragraph.

11.7 The COPERNICUS Program

Environmental applications of earth observation images are difficult to overestimate. For example, a single satellite image can show air pollution across an entire continent. As Earth Observation satellites remain in place for long periods, it is possible to track environmental changes gradually. Copernicus is the newest and largest European Environmental program, previously known as Global Monitoring for Environment and Security, was primarily designed to manage and protect the environment and natural resources. Copernicus is designed to provide six main services:

(all information is taken from [25], unless sited otherwise)

1. **Marine environment** monitoring service provides information about physical, dynamical and biochemical states of the ocean. Applications include forecasting sea ice, oil spill combat, ship routing, marine resources/ fish management, water quality monitoring, pollution, coastal activities, seasonal and weather forecasting—climate monitoring and ice surveys.

2. **Land** monitoring service is divided into three components: *global* (provide real time, 10-day frequency information about the state of vegetation, the energy budget and water cycle); *Pan-European* (mapping land cover and land-change, additionally provides seasonal and annual vegetation parameters) and *local component* (provides detailed information for specific areas, for example, biodiversity "hotspots").
3. **Atmosphere** monitoring service aims to provide high-quality assessments, monitoring and forecasting of air composition at continental/regional/local scales. This focuses on climate change, UV-radiation and long-range transport of pollution observations. The focus is made towards monitoring carbon dioxide, methane, carbon monoxide and aerosols.
4. **Climate change** monitoring service will help to predict, mitigate and adapt to climate change. This will be realized through a network of observations (both, in-situ and satellite), as well as modelling capabilities. The climate influencing factors are: global temperature, sea level and ice cover.
5. **Emergency management** service aims to prepare, prevent, reduce the risk, response and recover from natural hazards, such as earthquakes, tsunamis, wildfires, floods, storms and industrial accidents.
6. **Security service** includes maritime surveillance and marine pollution control.

The Copernicus space component consists of two types of satellite missions: Sentinels 1–6 (designed for operational needs of Copernicus) and around 30 contributing missions (from other space agencies).

11.7.1 Satellite Equipment and Contributing Missions

Optical equipment—passive systems that measure sun's reflected energy across various wavelengths or thermal radiation emitted by Earth (based on reflectometry principle). Optical images are the most common instrument used for Earth observation. Spectroscopy allows for detailed monitoring of the Earth's surface, water and atmosphere by capturing object spectrum, e.g. each wavelength on the image corresponds to specific chemicals and elements. For example, ***Blue*** (450–515…520 nm) is used for atmosphere and deep water imaging; ***Green*** (515…520–590…600 nm) is used for imaging vegetation and deep water structures; ***Red*** (600…630–680…690 nm) is used for imaging man-made objects, water, soil and vegetation; ***Near-infrared*** (750–900 nm) is used primarily for imaging vegetation; ***Mid-infrared*** (1550–1750 nm) is used for imaging vegetation, soil moisture content and forest fires; ***Far-infrared*** (2080–2350 nm) is used for imaging soil, moisture, geological features, silicates, clays and fires; ***Thermal infrared*** (10,400–12,500 nm) uses emitted instead of reflected radiation to image geological structures, thermal differences in water currents and fires [78]. ***Multi- and hyper-spectral imaging*** is the main tool for earth observation. The important parameters are *spectral* (number of spectral bands available, which correspond to characterizing different materials)

and *spatial* resolution (minimum size of a detail, which can be observed from the image). Low spatial resolution is better suited for the study of regional vegetation coverage, or weather/cloud patterns. Intermediate resolution is good for studying agriculture, resource mapping and assessing impact of disasters. High-resolution sensors are able to highlight an object, for example, building or car. The main limitations of optical imaging are weather conditions and presence of daylight. Optical sensors are one of the most important land cover mapping tools. The data is generally delivered in three spectral regions: visible, near-infrared and middle-infrared. This is due significant differentiation of reflectance properties from the different surfaces (soil, leaves, wood, ash, water and snow). The combination of reflectance properties allows differentiate between vegetation types, which can be a good indicator of the land cover and habitat type [32]. Spectral information from the satellite imagery provides valuable knowledge about vegetation indices, which is quantitative data about vegetation productivity.

Optical Environmental missions (information about contributing missions is taken from [25] unless sited otherwise): *ERS-2/A TSR-2*, launched in 1995 (operational until 2011) is used to provide measurements of sea-surface temperature, cloud-top temperature and vegetation monitoring. *ENVISAR* launched in 2002 (until 2012) is the largest Earth satellite ever built. It measured sea-surface temperature. *EnMap* (Environmental mapping and Analysis program) to be launched in 2018. It will provide data about agriculture, forestry, soil, geological environments, costal zones and inland water. *Venμs*, to be launched in 2016, will provide data for vegetation monitoring. *Deimos-1* was launched in 2009, which provides data for agriculture and forestry, *Pleiades* consists of two satellites: Pleiades 1A (2011) and Pleiades 1B (2012). It can be applied for renewable resource management, hydrology, marine environment and mapping applications (and other non-environmental applications). *Prisma*, launched in 2011, deliver information for mapping of land cover and agricultural landscapes, natural hazards, quality of inland water, costal zones and Mediterranean sea, carbon cycle monitoring, atmospheric turbidity, land surface hydrology and water management, desertification. *Proba V*, launched in 2013, provides data about vegetation and land surface. This includes data about agriculture, forests/deforestation, deserts, fires, water management and others. *SEOS-at Ingenio*, launched in 2015, provides information for applications in cartography, land use, water management, environmental monitoring, and others. *SPOT* mission consists of seven optical satellites (SPOT 1 launched in 1986, SPOT 7 in 2014). Environmental applications of SPOT satellite include: Chernobyl disaster contamination monitoring [28], and vegetation monitoring [67].

Radar (radio detection and ranging) instruments—active system that transmits microwave pulses and measures its reflection. It is independent from light and heat therefore can provide day-and-night images. In Copernicus program, various satellites carry *Synthetic Aperture Radar* (SAR), which is designed to create images of an object (such land and ocean 2/3 D surface mapping). In SAR, the limitation of a microwave antenna (which is poor spatial resolution resulting from a large beam width or angular field of view) is turned into an advantage by precise measurement of phase and Doppler shift during movement of the satellite. After processing the

signal it is possible to construct (or synthesis) the aperture, which is equivalent to the distance that the physical antenna moves while the location remains in the beam. In other words, the physical antenna can have a small size, but the distance that the SAR is travelling while radar signal is bounced against an object and travel back can be turned into larger "synthetic" antenna. Similar as in photography, the larger the camera the better picture. SAR technique is applied to get high-resolution images with small physical antenna. Details about SAR can be seen in [13, 52]. Radar together with lidar (described below) can be used for mapping forest characteristics, such as age, density and biomass. A major advantage of the radar is its ability to penetrate clouds. A major disadvantage is that it is very difficult to use in vegetation mapping in hilly or mountainous areas because the topography dominates the patterns observed. Learn more in [32].

Radar Environmental Missions (information about contributing missions is taken from [25] unless sited otherwise). *ERS-2/SAR* (1995–2011 mentioned above) along optical sensors also carried SAR equipment for various environmental monitoring purposes, for details, see [6]. *Envisat/ASAR* (2002–2012, mentioned above) is the largest satellite ever built. It carried 10 instruments, including A-(advanced)-SAR. Sea ice mapping was one of the ERS-2/SAR applications. *Cosmo-SkyMed* (launched between 2007 and 2010) consists of four satellites; each is equipped with SAR sensor. The civilian/environmental use of the Cosmo-SkyMed includes monitoring of landslides, floods, oil spills and fires. *Radarsat-2* (launched in 2007), one of the applications is marine environment surveillance (oil seeps, bilge dumping) [68]. *TerraSAR-X*, launched in 2007, it was designed for hydrology, geology, climatology, oceanography, environmental and disaster monitoring, and cartography [18]. TerraSAR-X provides high-resolution SAR interferometry (overlapping two or more radar images over the same area to detect changes occurring between acquisitions), this is applied for vegetation mapping, ice drift detection, measurement of ocean currents and others.

Radar altimeters—are sensors that make precise measurements of time, which is needed for microwave or laser pulses to be reflected from the surface back to satellite. It makes accurate measurements of land and sea surface topology; sea ice and large iceberg monitoring; sea surface wind speeds and wave heights.

Altimeter Missions: (information about contributing missions is taken from [25] unless sited otherwise). *Envisat/radar altimeter-2* (2002–2012, mentioned above). It delivers valuable scientific measurements to examine changes on annual/decadal time scale of: global and regional sea level, dynamic ocean circulation patterns, significant wave height and wind speed, climatology, ice sheet elevation, sea-ice thickness, as well as monitoring ocean mesoscale, significant wave height and wind speed in near real-time, Marine geophysics—Polar oceans, Ice sheet margins—sea ice, Lakes, wetlands and river levels, Land, Ionosphere and water vapour [8]. *Cryosat*, launched in 2010, is able to determine the thickness of ice floating in the oceans and monitors changes in the vast ice sheets on land, particularly around the edges where icebergs are calved. *Jason-2 OSTM* (ocean surface topography mission), launched in 2008 and *Jason-3*, launched in 2016. Both deliver precise altimetry measurements of the global sea surface height (with accuracy of few

centimetres, every 10 days) in support of weather forecasting, climate monitoring (as sea-level rise is a critical factor in understanding Earth's dynamic climate) and operational oceanography. *Saral/Altika*, launched in 2013, contributes to operational oceanography and forecasting.

Atmospheric instruments, such as *lidars* (light detection and ranging): similar to radar (which transmit microwave pulses), it uses a laser to transmit a light pulse and a receiver with sensitive detectors to measure the backscattered or reflected light. Distance to the object is determined by recording the time taken between the transmitted and backscattered pulses and using the speed of light to calculate distance traveled. Atmospheric instruments also include radiometers and spectrometers (quantitatively measures the intensity of electromagnetic radiation in specific band of wavelengths in the spectrum) and *spectrometers* (discussed above). Atmospheric instruments detect light, heat or radio energy penetrating through the atmosphere. Applications include climate study and measuring air quality. Details about remote sensing methods can be seen in [30].

Atmospheric Missions: (information about contributing missions is taken from [25], unless sited otherwise). *Calipso*, launched in 2006, carries a cloud-aerosol lidar, an imaging infrared radiometer and wide field camera to measure clouds and aerosols. *Envisat* (discussed above) carries three instruments for atmospheric monitoring: GOMOS (Global Ozone Monitoring by Occultation of Stars)—medium resolution spectrometer, that uses star occultation measurement techniques to measure stratospheric ozone. For details, see [9, 57]; and MIPAS (Michelson Interferometer for Passive Atmospheric Sounding)—a Fourier transform spectrometer to measure trace gases in the middle and upper atmosphere [9]; *Sciamachy* imaging spectrometer measures trace gases in the troposphere and stratosphere. *Merlin*, launched in 2014, is dedicated to measure atmospheric concentrations of methane. *MetOP* (meteorological operational satellite)—A, launched in 2006; MetOP-B, launched in 2012, both provide data for weather forecasting and climate studies.

11.7.2 Sentinels

Sentinels are family of satellites, which are developed specifically for operational needs of COPERNICUS program. Sentinel project consists of six missions, which are scheduled for launch between 2014 and 2020; there are two satellites dedicated for each Sentinel-1, -2 and -3 missions. The Sentinel-4 and -5 will be carried on host satellite platforms. Sentinel-6 is designed to complement information from Sentinel-3.

Sentinel-1 [25]: first satellite (Sentinel 1A) was launched in April 2014, second (Sentinel -1B) is scheduled for lunch in 2016 [59]. It provides ***radar-based services for land and ocean applications***. Synthetic Aperture Radar (SAR) operations do not depend on sunlight or weather conditions and it can provide images with different resolution (down to 5 m) and coverage (up to 400 km). Applications of Sentinel-1 include:

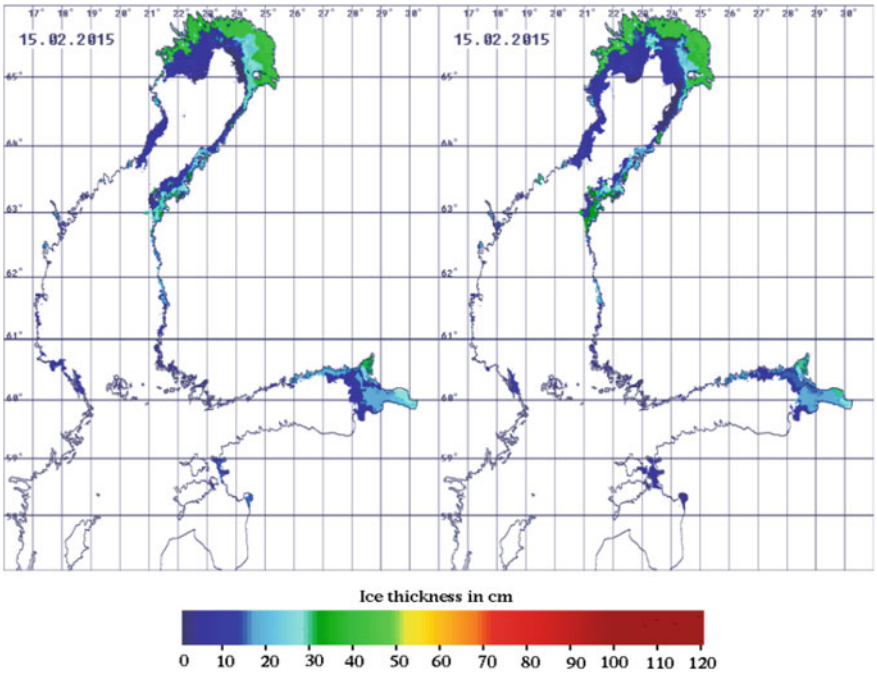

Fig. 11.4 Baltic Sea ice thickness chat from Radarsat-2 and Sentinel-1a, 2015. Copyright: Finnish Meteorological Institute (FMI). Image is reproduced with the permission of Dr. Juha Karvonen

- **Monitoring of sea ice** zones and the arctic environment. Radar images from Sentinel-1 generate timely maps of sea-ice conditions. It is possible to determine the ice thickness, which not only ensures shipping safety, but also in combination with earlier radar images, helps to understand long-term *impacts of climate change*. This include Arctic sea-ice cover, continental ice sheets and glaciers, as well as estimation of daily/annual ice velocities. For example, in Fig. 11.4, we can see the rate of ice cap elevation change between 2010 and 2014 observed by CryoSat, overlaid on an image acquired by Sentinel-1A (in 2014). The red colour on the image indicates that the ice surface is lowering faster than 2 m a year. Four smaller images in 1b show evaluation of ice velocity from 1995 to 2014. Literature about ice monitoring with satellite imaging can be found in [48, 74, 80].
- **Mapping of land surfaces and land surface motion**: forest, agriculture, water and soil. Satellites, equipped with optical and radar sensors, have great advantages to generate accurate land cover maps. Although optical imaging is a more preferable tool for area mapping, radar imaging is a complimentary data source when weather conditionals don't allow optical imaging [66]. Interferometric radar has potential to detect surface movements with very high accuracy. For example, in Fig. 11.5 we can see 28 radar images combined into single image

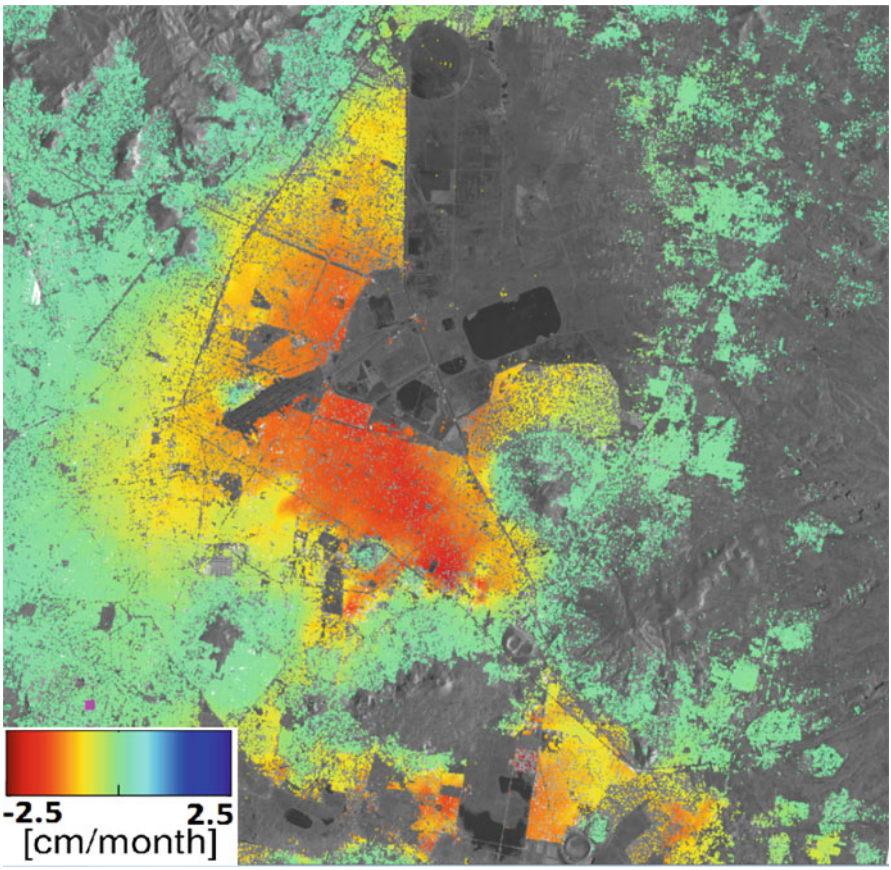

Fig. 11.5 Ground deformation of Mexico city. The result is produced from 28 radar scans of Sentinel-1a; 2016. Copyright: Copernicus data (2016)/ESA/DLR Microwaves and Radar Institute-SEOM InSARap study. Image is reproduced with the permission of Dr.Ing. Matteo Nannini

(from the Sentinel 1A), which shows ground movement in Mexico City with some areas sinking up to 2.5 cm/month due to ground water extraction [19]. Such technique is very useful for monitoring the effect of landslides, earthquakes, man-made activities, water resource monitoring [3], extraction of natural gas and mining.

- **Surveillance of marine environment** (oil spill monitoring). Sentinel-1 provides oil spill detection and surveillance services as part of Earth Observation *"CleanSeaNet"* program [20]. Synthetic Aperture Radar images are the main tool for remote oil spill detection because of various advantages (already mentioned above): 24 h operation capacity, independent from weather conditions, homogeneous dielectric property of water in microwave frequencies and high spatial resolution. An example of radar image of the oil discharge from oil

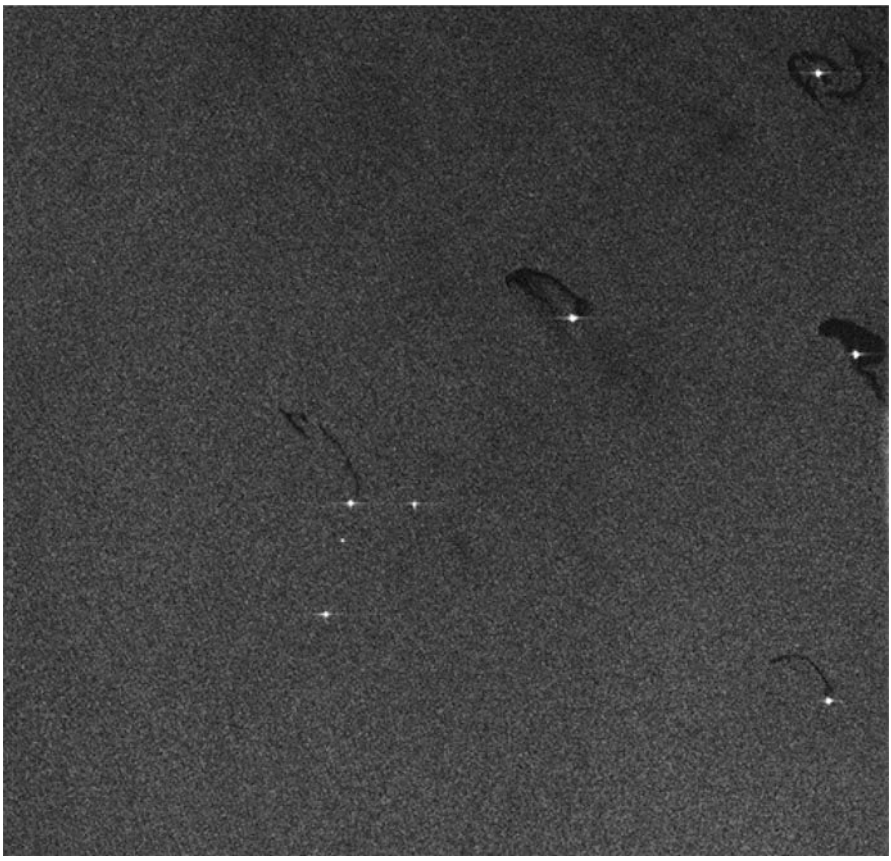

Fig. 11.6 Example of water contamination by oil platforms off the coast of Norway. Radar image is taken from Sentinel-1a, 2014. Copy right: European Space Agency. Image is reproduced with the permission of European Space Agency

platforms is presented in Fig. 11.6. The bright spots in the image are oil platforms off the coast of Norway. The black areas show where water is released by the platforms. This water is slightly oily, though not enough to be in violation of regulations. The radar on Sentinel-1A has the sensitivity to identify these events. Satellite imaging for oil spill monitoring is further presented in [11, 39]. Sentinel-1 also provides information about the sea state, e.g. wind, waves and current, which can be used to track the path of the oil spills and other pollutants.

- **Other** Sentinel-1 applications include maritime security (e.g. ship detection), support to emergency/risk management (for example, flooding) and humanitarian aid in crisis situations. For detailed overview of Sentinel-1 applications, see [15, 25, 66].

Sentinel-2 [all information is taken from [25], unless sited otherwise]: The mission is based on a constellation of two identical satellites. Sentinel-2A was launched on 23 June 2015 and Sentinel-2B will follow in the second half of 2016. It provides ***multi-spectral imaging services*** for land applications. The mission aims to cover four main areas:

- **Vegetation monitoring and Plant health**: Sentinel-2 can provide frequent, efficient and timely mapping of vegetation with 10–20 m resolution. When both satellites are operational, the missions re-visit time every 5 days, which means a very accurate vegetation growth monitoring. Sentinel-2 is a unique Earth observation mission, because it provides specific optical characteristics in the red band (it has three bands in the "red edge"). Depending on vegetation cover density and leaf chlorophyll content, the reflectance propertied in the red and near-infrared band region will vary. This is called normalized difference vegetation index [77]. More about spectral imaging can be seen in [12, 49, 71].

  The images from this mission can be used for distinguishing different crop types and characteristics of the plans (such as leaf area index, leaf chlorophyll and water content). This information is important to farm a good harvest (when deciding how much water or fertilizers are needed), as well as form strategies to address climate change.
- **Land services**: Sentinel-2 is applied for accurate mapping of the land cover (forests, crops, grassland, water or buildings), landscape and monitor changes over time. It helps to monitor deforestation, re-forestation, wildfires (which contributes to disaster management) and other natural resources.
- **Water**: Sentinel-2 allows tracking changes in water bodies and coastal environment. Additionally, the images can capture water quality parameters (but only on the surface), such as chlorophyll concentration, harmful algal blooms and measure water clarity. This is vital information for water management-related research.
- **Mapping of disasters**: the mission provides vital information for emergency response management/impact assessment in the case of extreme weather, volcano activities, wildfires, earth erosion, floods and others.

Sentinel-3: consists of two identical satellites, Sentinel 3A (launched in February 2016) and Sentinel 3B (to be launched in 2017). It carries:

- *ocean and land colour instruments* for vegetation mapping (it is complimentary to Sentinel-2 multispectral imager);
- *sea and land surface temperature radiometer*;
- *synthetic aperture radar altimeter*;
- *microwave radiometer* (measure sea/land surface temperature, as well as fires).

The mission is primarily dedicated to ocean and weather forecasting. But it will also provide services over land (such monitoring wildfires, land use and inland water, which are complimentary to Sentinel-2).

Applications include: marine ecosystem health monitoring (algal pigment concentration, total suspended matter, coloured dissolved organic matter, chlorophyll-

a), prediction of harmful algal blooms, sea state, sea-surface temperature, sea ice, water clarity and surface wind speed. Land applications include mapping vegetation and temperature measurements.

Sentinel 4–5: Both missions will be carried on host meteorological satellites. Sentinel-4 is scheduled for launch in 2019 and will carry an ***ultraviolet visible near-infrared (UVN) spectrometer***. Sentinel-5 is scheduled for launch in 2015–2020 and will carry an ***ultraviolet visible near-infrared shortwave (UVNS) spectrometer***. Both missions are dedicated to atmosphere monitoring, such as air quality control (greenhouse gases and aerosols), stratospheric ozone and solar radiation.

Sentinel-6: is a complimentary mission to Sentinel 3. It will carry a ***radar altimeter*** to provide information on global ocean topology.

11.7.3 Copernicus Data Access

Service providers will harmonize and analyse the raw data from the satellites. Depending on scientific objectives, satellite data will be combined with measurements from aircrafts, on the ground or in the ocean. For example, if the target is to provide air quality forecasts, the measurements of the air pollution from satellites will be combined with data from ground-based instruments and fed into a model. Open and free of charge data access from Copernicus can be found in at scihub.copernicus.eu.

11.8 Advantages and Limitation of GIS and RS

The general **limitation** of remote sensing is ability to monitor only features that can be viewed from above. An exception is radar and lidar, which can penetrate through clouds but they are limited by surface topology, high cost, lack of analytical monitoring standards and data availability [32].

When producing a map of vegetation, the individual features belonging to a particular class of interest must be larger with respect to the resolution of the imagery (example: 10-m wide stream will not be visible in an image composed of cells of 1 km spatial resolution).

It is very important that the feature being observed has a significantly unique spectral signature to be separated from other types of signatures (example: distinguishing between primary and secondary rain forests).

General limitation of optical sensors is cloud cover (tropical regions can produce problems for optical sensing), atmospheric conditions and mechanical problems often decrease measurement quality too.

The real **advantage** of using sensors over the human eye is the ability to automatically record data, to process data with various algorithms, will not tire with time and to automatically GPS/time stamp the data [58].

11.9 Summary

This chapter identified two main sources for almost all environmental problems: Man-made pollution of air/water/soil and Natural resource depletion. Geographic Information Systems and Remote Sensing are powerful and unique tools for non-destructive, repetitious, large-scale data collection for Earth observation and Environmental assessments.

This current chapter presented an overview about following issues:

- **Land monitoring** included observations of landscape conditions, landscape dynamics, monitoring of land usage, map wetlands and chart wildlife. This information is essential for natural resource managers to understand better how to protect the natural environment and minimize urban pressure.
- Information about **ocean and coastal zones** helps to assess marine and coastal habitats with a strong focus on protecting coral reefs, mangroves and sea grasses.
- In order to protect **biodiversity**, it is essential to know species population and identify areas of high risk of poaching. Telemetry tracking methods allows studying individual animal species by attaching special tag on/inside an animal. Depending on the information from the tag is stored or sent, the tags are archival or transmitting (accordingly). Transmitting tags correspond to: GPS-tracking, Argos satellite tracking, Very High Frequency tracking, Acoustics tracking or Passive Integrated transponder tracking.
- Successful combination of surveillance techniques for the **illegal fishing** combat is a joint application of: vessel monitoring system (on board the ship), vessels detection system (from the satellite) and air patrol surveillance.
- Drones are proven to be very efficient tools for minimizing **poaching** risks. Additionally, heat-sensor planes and hidden remote sensors can help to detect poachers in remote areas.
- **Natural hazards** can be effectively assessed applying RS and GIS. By collecting, organizing and analysing environmental, meteorological and geographical data units, it is possible to reduce/avoid the losses from hazards, insure fast and appropriate assistance to victims of disaster and effectively recover after the event.
- In order to detect **marine debris** it is essential to use a combination of techniques. For example: models debris movement (based on surface current/wind data) + satellite radar and multispectral data + airborne remote sensing. Another possibility is a combination of models + satellite observations and airborne radar + visual debris detection as a final stage. It is also possible to apply multi-or hyperspectral imaging with an automated detection algorithm operating in real time and a lidar that would be aimed at targets detected by that algorithm.

Additionally **THE LIVING PLANET** and **COPERNICUS** programs has been addressed. The living planet program consist of eight "earth explorer" research missions and six "Earth watch" methodological missions. While Copernicus is the largest European environmental space program, aiming to provide marine, land,

atmosphere, climate, emergency and security services. It consists of 30 contributing missions as well as six Sentinel missions:

- Sentinel-1 (2014/6): radar-based services for land and ocean applications;
- Sentinel-2 (2015/6): multispectral imaging services for land applications;
- Sentinel-3 (2016/7): ocean and land colour instruments; sea and land surface temperature radiometer; synthetic aperture radar altimeter and microwave radiometer. The mission is primarily dedicated to ocean and weather forecasting;
- Sentinel-4 and 5 (2019/20): ultraviolet visible near-infrared (UVN) spectrometer and ultraviolet visible near-infrared shortwave (UVNS) spectrometer services for atmosphere monitoring;
- Sentinel-6: radar altimeter will provide information on global ocean topology. Optical/hyperspectral imaging provides information about vegetation, flood, mining, pollution, vegetation stress, agriculture and rural developments. Radar and LIDAR record information about earth deformations, ground movement and landslides; Thermal images are dedicated to fire detection, geothermal activities and volcanoes.

Acknowledgements This work was financially supported by EU FP7 Marie Curie Initial Training Network MULTI-POS (Multi-technology Positioning Professionals) under grant nr. 316528.

References

1. I.I. Abbas, J.A. Ukoje, Application of remote sensing (Rs) and geographic information systems (Gis) to environmental impact assessment (Eia) for sustainable development. Res. J. Environ. Earth Sci. **1**(1), 11–15 (2009). ISSN:2041-0492
2. M. Allsopp et al., Plastic debris in the World's oceans. Tech. rep. Greenpeace, 2006, pp. 1–44
3. D. Amitrano et al., Sentinel-1 for monitoring reservoirs: a performance analysis. Remote Sens. **6**(11), 10676–10693 (2014). ISSN:2072-4292. doi:10.3390/rs61110676
4. X. Andre, B. Moreau, S. Le Reste, Argos-3 satellite communication system: implementation on the arvor oceanographic profiling floats. J. Atmos. Ocean. Technol. **32**(10), 1902–1914 (2015). ISSN:0739-0572. doi:10.1175/JTECH-D-14-00219.1
5. Argos, Worldwide tracking and environmental monitoring by satellite. Online www.argos-system.org (2014)
6. E.P.W. Attema, G. Duchossois, G. Kohlhammer, ERS-1/2 SAR land applications: overview and main results, in *IGARSS '98. Sensing and Managing the Environment. 1998 IEEE International Geoscience and Remote Sensing. Symposium Proceedings. (Cat. No.98CH36174)*, vol. 4 (1998), pp. 1796–1798. doi:10.1109/IGARSS.1998.703655
7. S. Baumann, W. Lechner, Global navigation satellite systems. Comput. Electron. Agric. **25**, 67–85 (2000)
8. J. Benveniste et al., The EnviSat radar altimeter system. Remote Sens. Ocean Sea Ice 2001 **4544**, 71–82 (2002). doi:10.1117/12.452745
9. J.L. Bertaux et al., Global ozone monitoring by occultation of stars: an overview of GOMOS measurements on ENVISAT. Atmos. Chem. Phys. **10**(24), 12091–12148 (2010). ISSN:1680-7324. doi:10.5194/acp-10-12091-2010
10. K. Boat, C. Turley, Ocean acidification. Tech. rep. UNEP, 2010, pp. 1–9
11. C. Brekke, A.H.S. Solberg, Oil spill detection by satellite remote sensing. Remote Sens. Environ. **95**(1), 1–13 (2005). ISSN:00344257. doi:10.1016/j.rse.2004.11.015

12. H.K. Burke, G.A. Shaw, Spectral imaging for remote sensing. Lincoln Lab. J. **14**(1), 1–28 (2003)
13. Y.K. Chan, V.C. Koo, An introduction to synthetic aperture radar (SAR). Progr. Electromagn. Res. B **2**, 27–60 (2008). ISSN:1937-6472
14. J. Cicuendez Perez et al., The efficiency of using remote sensing for fisheries enforcement: application to the Mediterranean bluefin tuna fishery. Fish. Res. **147** (2013), 24–31. ISSN:01657836. doi:10.1016/j.fishres.2013.04.008
15. M. Davidson, E. Attema, G. Levrini, Sentinel-1 ESA's new European radar observatory, in *The Future of Remote Sensing*, 2005
16. C. Dempsey, Data capture in GIS, 2012
17. S. Dubey, Acid rain-the major cause of pollution: its causes, effects and solution. Int. J. Sci. Eng. Technol. **2**(8), 772–775 (2013)
18. Earth Observation Portal Directory, Satellite missions, 2000
19. Earth Observation Portal Directory, Copernicus: Sentinel-1 - Satellite Missions - eoPortal Directory, 2000
20. Earth observation services, CleanSeaNet 2nd generation, 2014
21. G. Elert, The physics fact book: thickness of the ozone layer. Online e-book: www.hpertextbook.com (2000)
22. K. Erzini et al., Catches in ghost-fishing octopus and fish traps in the north-eastern Atlantic Ocean (Algarve, Portugal). Fish. Bull. **106**, 321–327 (2008)
23. ESchooltoday, What is water pollution for children, 2016
24. ESRI, GIS for wildlife management, Tech. rep. 2010, p. 40
25. European Space Agency, Copernicus/Observing the Earth/Our Activities/ESA (2017)
26. J. Forget, L. Ritter, K.R. Solomon, Persistent organic pollutants. Tech. rep., Canadian Network of Toxicology Centres, Guelph, 1995, pp. 1–145
27. T.S. Galloway, S.L. Wright, R.C. Thompson, The physical impacts of microplastics on marine organisms: a review. Environ. Pollut. **178**, 483–492 (2013)
28. GISGeography, SPOT Satellite Pour l'Observation de la Terre - GIS Geography, 2012
29. P. Gleick, J. Morrison, Freshwater resources: managing the risks facing the private sector. Tech. rep. Pacific institute, 2004, pp. 1–16
30. S. Graham, Remote sensing: feature articles, September 1999
31. Greenpeace, GREENPEACE and the dumping of waste at sea: a case of non-state actors intervention in international affairs. Kluwer Law Int. **4**(3), 1–17 (1999)
32. R. Hoft, H. Rtrand, J. Strittholt, *Sourcebook on Remote Sensing and Biodiversity Indicators, Convention on Biological Diversity Technical Series*, vol. 32 (CBD, Cambridge, 2007)
33. C.M. Hunter et al., Climate change threatens polar bear populations: a stochastic demographic analysis. Ecology **91**(10), 2883–2897 (2010). ISSN:0012-9658. doi:10.1890/09-1641.1
34. IUCN: International Union for Conservation of Nature (2017)
35. R. Kays et al., Tracking animal location and activity with an automated radio telemetry system in a tropical rainforest. Comput. J. **54**(12), 1931–1948 (2011). ISSN:0010-4620. doi:10.1093/comjnl/bxr072
36. R.E. Kennedy et al., Remote sensing change detection tools for natural resource managers: understanding concepts and tradeoffs in the design of landscape monitoring projects. Remote Sens. Environ. **113**(7), 1382–1396 (2009). ISSN:00344257. doi: 10.1016/j.rse.2008.07.018
37. V.V. Klemas, Advances in fisheries applications of remote sensing, in *2014 IEEE/OES Baltic International Symposium (BALTIC)*, May (IEEE, New York, 2004), pp. 1–21. ISBN: 978-1-4799-5708-8. doi:10.1109/BALTIC.2014.6887836
38. S. Koenig, Bioaccumulation of persistent organic pollutants (POPs) and biomarkers of pollution in Mediterranean deep-sea organisms. Ph.D. thesis, Universidad de Barcelona, 2012
39. O.Y. Lavrova, M.I. Mityagina, Satellite survey of the black sea. Int. Water Technol. J. **2**(1), 65–77 (2012)
40. G. Lennen, Method and apparatus for managing and configuring tracker components for enhanced sensitivity tracking of GNSS signals, November 2011

41. M.J. Maah, M.A. Ashraf, I. Yusoff, in *Environmental Risk Assessment of Soil Contamination*, ed. by M.C. Hernandez Soriano. InTech (2014). ISBN: 978-953-51-1235-8. doi:10.5772/57086
42. T.H. Mace, At-sea detection of marine debris: overview of technologies, processes, issues, and options. Mar. Pollut. Bull. **65**(1–3), 23–27 (2012). ISSN: 1879-3363. doi:10.1016/j.marpolbul.2011.08.042
43. B. Mate, Satellite tracking of sperm whales in the Gulf of Mexico in 2011, a follow-up to the deepwater horizon oil spill. Tech. rep. Hatfield Marine Science Center, 2011, p. 13
44. B. McConnell et al., Methods for tracking fine scale underwater movements of marine mammals around marine tidal devices. Tech. rep. Sea Mammal Research Unit, Scottish Ocean Institute of St Andrews, 2013, p. 38
45. B. Merchant, Anti-poaching tech: can heat-seeking planes, drones, and DNA mapping save the rhino? in *Motherboard*, 2012
46. K.L. Metzger et al., Using historical data to establish baselines for conservation: the black rhinoceros (*Diceros bicornis*) of the Serengeti as a case study. Biol. Conserv. **139**(3–4), 358–374 (2007). ISSN: 00063207. doi:10.1016/j.biocon.2007.06.026
47. J. Nacher-Mestre et al., Bioaccumulation of polycyclic aromatic hydrocarbons in gilthead sea bream (*Sparus aurata* L.) exposed to long term feeding trials with different experimental diets. Arch. Environ. Contamin. Toxicol. **59**(1), 137–146 (2010). ISSN: 1432-0703. doi:10.1007/s00244-009-9445-1
48. T. Nagler et al., The Sentinel-1 Mission: new opportunities for ice sheet observations. Remote Sens. **7**, 9371–9389 (2015). doi:10.3390/rs70709371
49. NASA, How Landsat images are made. Tech. rep., 2006
50. National Geographic Society, GIS (geographic information system), 2016
51. NOAA: National oceanic and atmospheric, Climate change and its effects on ecosystems habitats and biota. Tech. rep. Woods Hole, 2010, pp. 1–18
52. NOAA: National oceanic and atmospheric, *SAR Marine User's Manual* (NOAA/NESDIS, Washington, DC, 2004)
53. NOAA: National oceanic and atmospheric, What is remote sensing? EN-US. Tech. rep., 2015
54. S. Northridge, A.J. Read, P. Drinker, Bycatch of marine mammals in U.S. and global fisheries. Conserv. Biol. **20**(1), 163–169 (2006). doi:10.1111/j.1523-1739.2006.00338.x
55. Ocean Conservancy, Trash travels: from our hands to the sea, around the globe, and through time, Tech. rep., 2010
56. Organization of American States, Dept. of Regional Development and Environment, Organization of American States, Natural Hazards Project, and United States, Agency for International Development, Office of Foreign Disaster Assistance, Disaster, planning and development: managing natural hazards to reduce loss, 1990
57. T. Paulsen et al., The global ozone monitoring by occultation of stars (GOMOS) instrument on ENVISAT, in *IEEE 1999 International Geoscience and Remote Sensing Symposium. IGARSS'99 (Cat. No.99CH36293)*, vol. 2. (IEEE, New York, 1999), pp. 1438–1440. ISBN: 0-7803-5207-6. doi:10.1109/IGARSS.1999.774657
58. W.G. Pichel et al., GhostNet marine debris survey in the Gulf of Alaska–satellite guidance and aircraft observations. Mar. Pollut. Bull. **65**(1–3), 28–41 (2012). ISSN: 1879-3363. doi:10.1016/j.marpolbul.2011.10.009
59. P. Potin et al., Sentinel-1 Mission operations concept, in *2014 IEEE Geoscience and Remote Sensing Symposium*, July (IEEE, New York, 2014), pp. 1465–1468. ISBN: 978-1-4799-5775-0. doi:10.1109/IGARSS.2014.6946713
60. RIC British Columbia, Wildlife Radio-telemetry. Tech. rep., 1998
61. P.G. Ryan et al., GPS tracking a marine predator: the effects of precision, resolution and sampling rate on foraging tracks of African Penguins. Mar. Biol. **145**(2) (2004). ISSN: 0025-3162. doi:10.1007/s00227-004-1328-4
62. Sealtag, Tag types (2011)

63. M. Sibanda et al., Understanding the spatial distribution of elephant (Loxodonta africana) poaching incidences in the mid-Zambezi Valley Zimbabwe using Geographic Information Systems and remote sensing. Geocarto Int., 1–13 (2015). ISSN: 1010-6049. doi:10.1080/10106049.2015.1094529
64. Sirtrack limited, Wildlife tracking solutions - harness (2017)
65. A. Slingenberg et al., Study on understanding the causes of biodiversity loss and policy assessment framework. Tech. rep. European Commission, 2009, pp. 1–206
66. P. Snoeij et al., Sentinel-1, the GMES radar mission, in *2008 IEEE Radar Conference*, May (IEEE, New York, 2008), pp. 1–5. ISBN: 978-1-4244-1538-0. doi:10.1109/RADAR.2008.4720735
67. Spot-vegetation, Vegetation, 2015
68. G. Staples, D.F. Rodrigues, Maritime environmental surveillance with RADAR SAT -2, in *Anais XVI Simposio Brasileiro de Sensoriamento Remoto - SBSR* (2013), pp. 8445–8452
69. The response of methane hydrate beneath the seabed offshore Svalbard to ocean warming during the next three centuries. Geophys. Res. Lett. **40**(19), 5159–5163 (2013). doi:10.1002/grl.50985. http://dx.doi.org/10.1002/grl.50985
70. B. Thomas, J.D. Holland, E.O. Minot, Wildlife tracking technology options and cost considerations. Wildlife Res. **38**(8), 653 (2011). ISSN: 1035-3712. doi:10.1071/WR10211
71. S.E. Umbaugh, *Digital Image Processing and Analysis: Human and Computer Vision Applications with CVIPtools*, 2nd edn., vol. 19 (CRC Press, 2010), p. 977. ISBN: 143980205X
72. United Nations, World population prospects, Tech. rep. New York, 2015. doi:10.1613/jair.301. arXiv:9605103 [cs]
73. United Nations Environment Programme Environmental Effects Assessment Panel, Environmental effects of ozone depletion and its interactions with climate change: progress report, 2015. Photochem. Photobiol. Sci.: Off. J. Eur. Photochem. Assoc. Eur. Soc. Photobiol. **15**(2), 141–174 (2016). ISSN: 1474-9092. doi:10.1039/c6pp90004f
74. R.S.W. van de Wal et al., Large and rapid melt-induced velocity changes in the ablation zone of the Greenland Ice Sheet. Science (New York, NY) **321**(5885), 111–113 (2008). ISSN: 1095-9203. doi:10.1126/science.1158540
75. C.J. Van Westen, *Remote Sensing and GIS for Natural Hazards Assessment and Disaster Risk Management*, vol. 3, March (Elsevier, Amsterdam, 2013), pp. 259–298. ISBN: 9780080885223. doi:10.1016/B978-0-12-374739-6.00051-8
76. T.S. Veenstra, J.H. Churnside, Airborne sensors for detecting large marine debris at sea. Mar. Pollut. Bull. **65**(1–3), 63–68 (2012). ISSN: 1879-3363. doi:10.1016/j.marpolbul.2010.11.018
77. I. Vorovencii, Satellite remote sensing in environmental impact assessment: an overview. Bull. Transilvania **4**(53), 1–8 (2011)
78. Wikipedia, Multi-spectral image, 2015
79. World Economic Forum, The new plastics economy- rethinking the future of plastics, Tech. rep. January 2016, pp. 1–120
80. J. Wuite et al., Evolution of surface velocities and ice discharge of Larsen B outlet glaciers from 1995 to 2013. Cryosphere **9**(3), 957–969 (2015). ISSN: 1994-0424. doi:10.5194/tc-9-957-2015
81. WWF: World Wide Fund for Nature, Climate change explained, 2016
82. WWF: World Wide Fund for Nature, Fishing problems: destructive fishing practices — WWF, 2016
83. WWF: World Wide Fund for Nature, Illegal Wildlife Trade, 2016
84. Z. Zommers et al., *Reducing Disaster: Early Warning Systems For Climate Change* (Springer, Dordrecht, 2014). ISBN: 978-94-017-8597-6

Chapter 12
Context Awareness for Semantic Mobile Computing

Alejandro Rivero-Rodriguez and Ossi Nykänen

12.1 Introduction

This book has previously focused on positioning information and its obtaining in different environments. In this chapter, we move from positioning information to consider other types of information regarding the user. This information can be used to provide personalized services.

In this section, we motivate the research on Context-aware Systems (CAS) and provide some background information. Section 12.2 focuses on the Context Engine (CE), a software component responsible for dealing with user context. Section 12.3 includes two case studies to infer unknown user context based on available information. Section 12.4 discusses the potential areas of application. Conclusions and discussions are presented in Sect. 12.5.

12.1.1 Motivation

Mobile technologies have been developed tremendously in the last years. Telecommunications allow now mobile devices to be connected to Internet practically anytime and anywhere. Users use mobile (smart) phones for any purpose, from

A. Rivero-Rodriguez (✉) • O. Nykänen
Tampere University of Technology, Korkeakoulunkatu 10, 33720 Tampere, Finland
e-mail: alejandro.rivero@tut.fi; ossi.nykanen@tut.fi

J. Nurmi et al. (eds.), *Multi-Technology Positioning*,
DOI 10.1007/978-3-319-50427-8_12

traditional calling to more contemporary activities such as checking e-mail, reading the news or using social media. This is possible because of all specific purpose applications, typically referred as apps, that have been developed in the last years.

The app market is increasingly important: the industry is expected to grow to a $77 billion industry by 2017 [7]. There are mobile applications for almost anything, including those for e-mail, news, messaging, weather, navigation, online shopping, calendar, hotel booking and social media. These applications have been developing lately in terms of personalization, aiming at adapting to user needs in order to facilitate our daily tasks. Examples of personalized applications are weather applications, which adapt to user location; online shopping platforms, which provide user with recommendation of interesting items based on historical information and personal assistants, which provide users with, among other things, the best route to go home from current position.

In this chapter, we discuss and demonstrate the current means for applications to access user-related information, the foundations for this kind of services to exist. In a nutshell, we believe this information to be hard to obtain and utilize. We propose a framework and mechanisms to manage this information, which will be referred to as user context and contextual information, interchangeably.

12.1.2 Background

According to Baldauf et al., context can *refer to any information that can be used to characterize the situation of an entity, where an entity can be a person, place, or physical or computational object* [3]. Context-aware applications are those that adapt their behaviour based on contextual information, in this case the information related to the user. Context-aware applications are an increasingly important part of current mobile applications. However, looking at application volumes, one could say that contemporary mobile sensor frameworks, e.g., Android Location and Sensors API, establish the *de facto* technology driver for implementing context-aware applications.

Such technologies provide some contextual information; however, they are often limited to information coming from physical sensors, such as position, or some simple user attributes such as user languages. Position is the most used context, resulting in the so-called LBS [24]. Other contextual information is needed for providing users with tailored services. Some other context includes user interests, profession and user routines. The key is that applications would not have to ask this information from the user, but through context-aware systems that provide the information in a transparent manner. Such needs raised the popularity of context-aware systems for mobile devices. Several frameworks and systems have been proposed in the literature to manage user context, such as CoBra, CASS, CORTEX, Gaia, Context toolkit or the CE [1, 3, 23]. Briefly, these frameworks facilitate the development of context-aware applications. In some cases, the task of context information acquisition is delegated from the application to the context managers.

That means that, while applications would need to obtain user information to provided specific services, they could now ask this information from the context manager, which has access to data sources, inference tools, external services and user preferences, among other components, to provide relevant user information. In the next section one of these frameworks, the CE, is presented and discussed.

12.2 Context Engine

The CE is a software component responsible for dealing with (collecting, storing and distributing) modeling and reasoning with context [23]. The CE accepts the responsibilities and tasks of local context provider and logical context interpretation. Among other tasks, the CE provides contextual information to the applications, and manages user-related information and preferences.

12.2.1 Architecture

The architecture of the CE is depicted in Fig. 12.1 [25]. The end user performs an activity through a context-aware mobile application, which requests the needed information from the CE through the CE API using the context ontology. Ontology is defined in computer science as a model for sharing formal concepts and has been applied in field such as intelligent information integration or cooperative information systems [12]. The CE provides the requested information to the requesting application if information is available or can be obtained with existing information and processing methods. The CE also deals with user preferences, permissions, privacy and trust, uncertainty (e.g. consistency checking), and ontology-based inference, among others. Optional components include optimization engines and application assistance services to help other applications use the CE smoothly. In the sequel some of the components from Fig. 12.1 are described:

- **Terminological information** is the key for computers to understand information such as text. Thesaurus or taxonomies can provide information about relationships of words, e.g., how to relate the terms father and son. Note that the terminological information is used by the CE for its internal reasoning, and not for communication with the applications.
- **Factual information** provides information to provide users with tailored services, including user-related information such as user gender or interests, and general information such as weather information. (Weather information can be considered user-related when it relates to the user's current place.)
- **Context ontology** is the key for communication between the CE and the applications. It is similar to the terminological information but, since its purpose

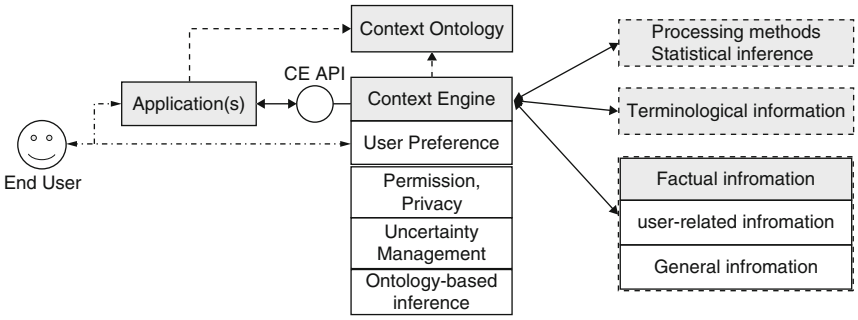

Fig. 12.1 Basic CE architecture [25]

is partially communicative, the applications need to understand the context ontology to communicate with the CE. It will further discussed in Sect. 12.2.3.

- **Statistical inference** is used to discover (previously unknown) contextual attributes based on available information. An example of an inference tool is activity recognition, to infer user activity status (walking, running, by car, etc.) based on available sensorial information [16]. Two statistical-based inference tools are presented in Sect. 12.3.

12.2.2 Responsibilities

The CE accepts the overlapping responsibilities and tasks of the local context provider and (logical) context interpretation, which typically exceed the boundaries of individual applications. The essential tasks of a CE include providing context information to the applications via various logical queries in terms of a standard I/O interface, and managing user preferences. When contextual information is unknown, the CE is responsible for obtaining this information by other means, using methods for acquiring contextual attributes, which will be discussed in Sect. 12.3.

12.2.3 Context Ontology

Although there are several means to model context, we choose ontology-based models since it offers several desirable properties such as information alignment, dealing with incomplete or partially understood information, domain-independent modelling, and formally working with context model of varying level of detail [5]. The Context ontology, therefore, plays a major role in context modelling. Many context ontologies have been proposed. We highlight the W3C Semantic Sensor Network (SSN) ontology, which was developed based on reviewing 17

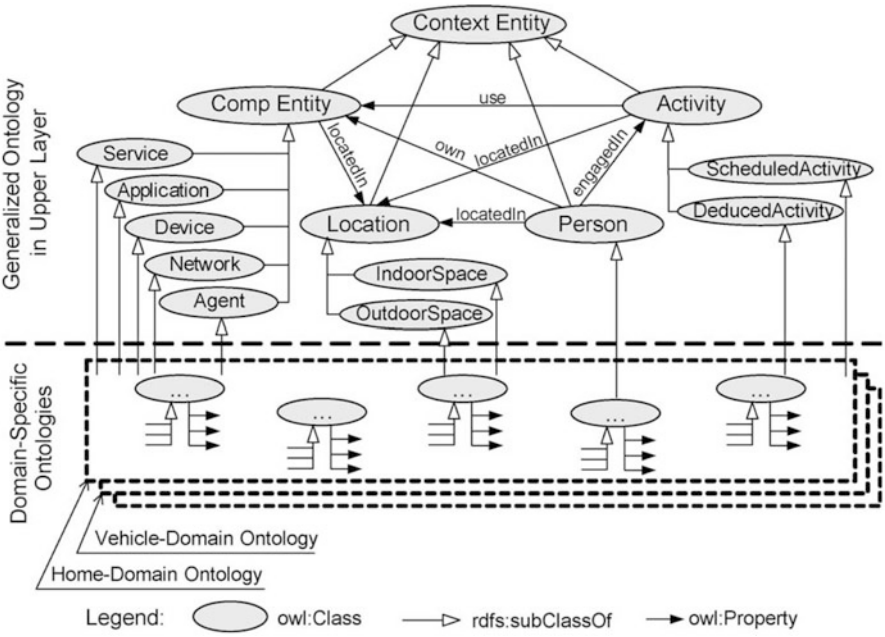

Fig. 12.2 Class hierarchy of the upper (SOCAM) ontology [13]

existing sensor or observation-centric ontologies [8], and the Service-Oriented Context-Aware Middleware (SOCAM) ontology, which acknowledges the general logical sensors context [13]. In SOCAM, context modelling is carried out in Web Ontology Language (OWL) ontologies based on two-level information architecture: the general context concepts are captured in the common upper ontology and application-specific concepts in domain ontologies, as shown in Fig. 12.2. The context engine typically includes general context as Location and Activities, which can also be integrated with other more specific ontologies, referred to as domain-specific ontologies in Fig. 12.2. This approach suggests using upper-level context ontology, in addition to general top-level alignment ontology, for integrating various kinds of domain ontologies, suitable for explaining their role in providing context.

12.2.4 Context Queries

Context and context queries can be classified according to different criteria. Two criteria for the classification are:

- **Information Vagueness**: Context information can be considered true in terms of crisp logic, e.g., user gender, while other information can be more uncertain, e.g., user interests.

Table 12.1 Classification of contextual information based on context query and information vagueness, showing an example

	Crisp information	Uncertain information
Atomic query	?age (I)	?favMusic (II)
Molecular query	?location(now,x), remind(x,y) (III)	?location(tomorrow,x), weather(x,y) (IV)

- **Complexity of query**: Context information can be obtained by using atomic queries for inferring single attribute context or molecular queries that combines at least two atomic queries.

Table 12.1 shows a simple classification of contextual queries based on the previous concepts. Crisp information can be obtained using atomic queries; for example, extracting user age from user profile (quadrant I). Using atomic queries we may discover uncertain information (quadrant II) using, for example, statistical inference: one could predict user favourite music based on his or her browsing history. Note that the age attribute has an objective and quantifiable value, while information attributes such as favorite music are more difficult to universally define. More complex contextual information acquisition processes require using molecular queries, in which we combine several contextual attributes. When dealing with factual information we may ask, for example, possible user reminders for the current location (quadrant III). The most challenging quadrant (IV) deals with molecular queries and vague information, but it is still very useful in practical applications. One such a context information instance would be the weather forecast for the user location the following day. (Note that, first, the user location on the next day has to be estimated and combined with weather forecast information.)

Since relevant context attributes are often unknown, inference tools are key components for the CE to reason with context. There are logical inference, based on the context ontology, its logical properties and relationships; and statistical inference, based on empirically extracting patterns from data. The latter infers contextual attributes statistically: therefore, inference method's validity and accuracy must be communicated properly to the CE to ensure the correct usage of such methods. For example, if an activity recognition method is used to determine the user motion status, the CE needs to know how this information was obtained and how accurate it is. For instance, the solution from the method can be something like *the user may be riding a bike, with probability 0.8. The information has been inferred from accelerometer information*. The CE may provide this information to the application that decides what to do with the contextual information. In the next section, Sect. 12.3, we present context inference tools to deal with atomic queries, corresponding with query types in quadrants I and II.

12.3 Methods for Acquiring Contextual Attributes

For providing user with personalized services, the user context needs to be known. We define contextual attribute as the smallest piece of context that is relevant to the user. The user has only one context, but many contextual attributes. User gender or location are examples of contextual attributes.

Specific context attributes are often unknown. Inference tools assist in inferring that information, allowing applications to use the inferred context. For instance, an activity recognition tool infers user activity, i.e., walking, sitting or in the bus, based on the inertial sensors on the mobile device [16]. Two case studies of inference tools to predict user information are presented.

12.3.1 Case 1: Inferring User Location Based on Phone Usage

The objective of this inference tool is to infer user place label based on user phone usage [26]. That is, to obtain logical information of the user location, not only in terms of position but also on the semantic of that position. We do so by automatically learning how mobile users use mobile phones in several places. Note that well-known classification techniques are used in this research. Most work consists of data processing and transformation to maximize classification accuracy. This section presents some insights and results from our previous work [26].

12.3.1.1 Dataset

The Mobile Data Challenge (MDC) dataset is used to learn the patterns. It was made available by Idiap Research Institute, Switzerland, and owned by Nokia Corporation [15, 17]. The dataset contains phone usage data from nearly 200 users over time periods from 6 to 18 months. Although the collected data was of 46 GB and contained information of several types, we extracted the most relevant data to solve the problem at hand and the data that could be used in practice without violating user privacy.

Frequent visits to places, defined as those where the user stays longer than 20 min, were detected automatically and users were asked to label these places manually. These label data were used to learn the relation between mobile phone usage and current user (semantic) location. The corresponding mobile usage information was recorded as well. The list of relevant stored information consists of

- **System data**
- **Call log**
- **Acceleration-based activity data**.

From each of these types of data, a list of potentially relevant features was computed, based on our insights on the problem. The system information list includes:

- **startHour** starting hour of the visit
- **endHour** finishing hour of the visit
- **duration** duration of the visit (in seconds)
- **nightStay** it captures the frequency of visits to the place between 6 pm and 6 am
- **sysActiveRatio** proportion of the duration when the system has been active in a visit
- **sysActStartsPerHour** number of status changes from system inactive to system active each hour during the visit
- **chargingTimeRatio** phone charging time in each visit
- **batteryAvg** average battery level during the visit.

Regarding calls, we consider but not distinguish incoming and outgoing calls, and we capture frequency and duration of calls.

- **callsPerHour** number of calls per each hour in a certain visit
- **callsTimeRatio** duration of calls per hours in a certain visit.

Accelerometer information was used to compute motion modes, e.g., idle, walking, etc. In the data, we calculated how portion of the time the user is in each motion status. The rationale behind that is that user motion may be correlated with user semantic location. The different modes were:

- **idleStillRatio** proportion of the visit that the user is idle/still
- **walkRatio** proportion of the visit that the user status is walk
- **vehicleRatio** proportion of the visit duration that the user status is either car/bus/motorbike or train/metro/tram
- **sportRatio** proportion of the visit duration that the user status is either run, bicycle or skateboard.

The possible places for the classification are *Home*, *Work* and *Others*. The methods should learn how users use their mobile phones in each place.

12.3.1.2 Data Processing

Two data-representation approaches were considered, as represented in Fig. 12.3. The **visit approach** uses the features computed for each visit as such, so that the data includes several samples of one user's visits to each of the user's places. Therefore, we consider one data point for each location-user-period. For instance, three user visits to the workplace would be capture in the system with three data points. We extract 55,932 labeled visits by 114 users.

The **place approach** considers all the visits of the user to the same place as a point, representing user place, but not time. In the previous example, three user visits to the workplace would results in one data point. The idea is that different users use their phones in similar ways in semantically similar places, for instance, users use phone similarly at home. This aggregation of visits to same place would make the

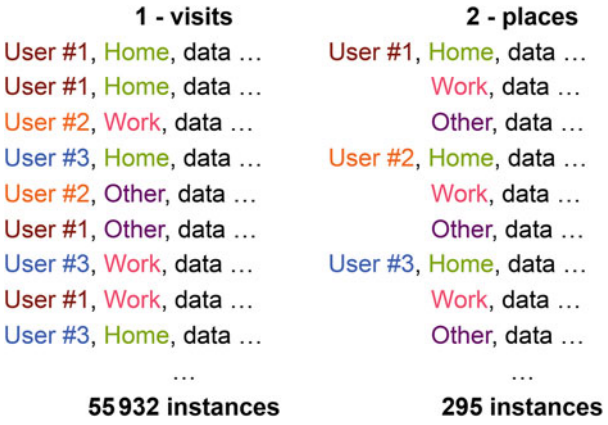

Fig. 12.3 Visit and place representation approaches

method robust to outliers, presumably improving the inference accuracy. From the database we extract 295 labeled places by 114 users. The difference between the approaches is illustrated in Fig.12.3.

12.3.1.3 Methods and Results

We learn patterns in the data using five well-known classification methods on the two data representation, namely Decision Tree, Bagged Tree, Naïve Bayes, Neural Networks and K Nearest Network (KNN). The results are presented in Figs. 12.4 and 12.5 for visit and place approaches, respectively. In the figures, the percentage of well-classified samples for each class is given above the bars and the overall percentage of well-classified samples for the classifiers is shown below the bars. The place approach yields better results than the visits approach. The Bayesian method and bagged trees outperformed the others in the places approach. More on the methods and its setting, the results and their interpretation can be found in [26].

The inference tool has been developed based on the MDC data. To use such tools in practical cases, one needs access to the user system information, call logs and acceleration-based activity data. This method may be used in practice by the CE.

12.3.2 Context Inference in Social Network Settings

This section presents a case study in which inference is made based on the users characteristic as a social being, i.e., relations with other users in the networks, using Social Network Analysis (SNA). In concrete we have a graph of 27 nodes representing student in a high-school classroom. Two users are considered similar

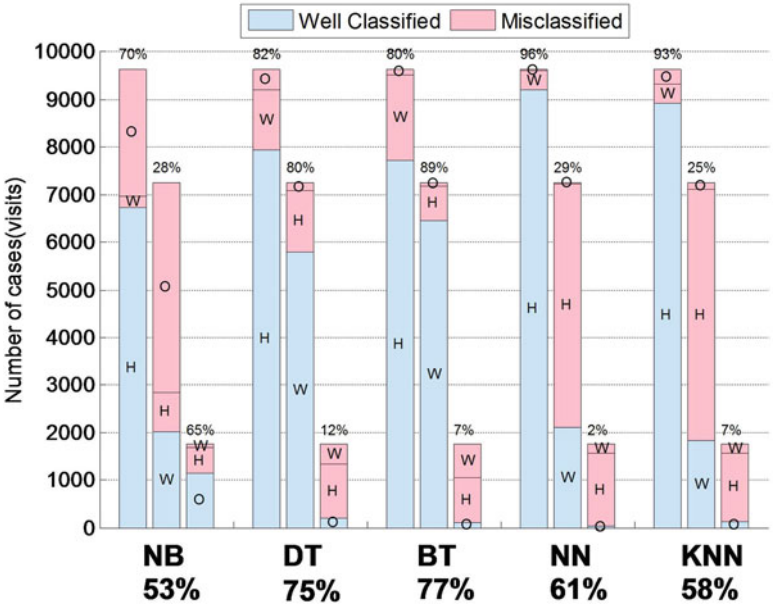

Fig. 12.4 Classification rates (%) for the visit approach

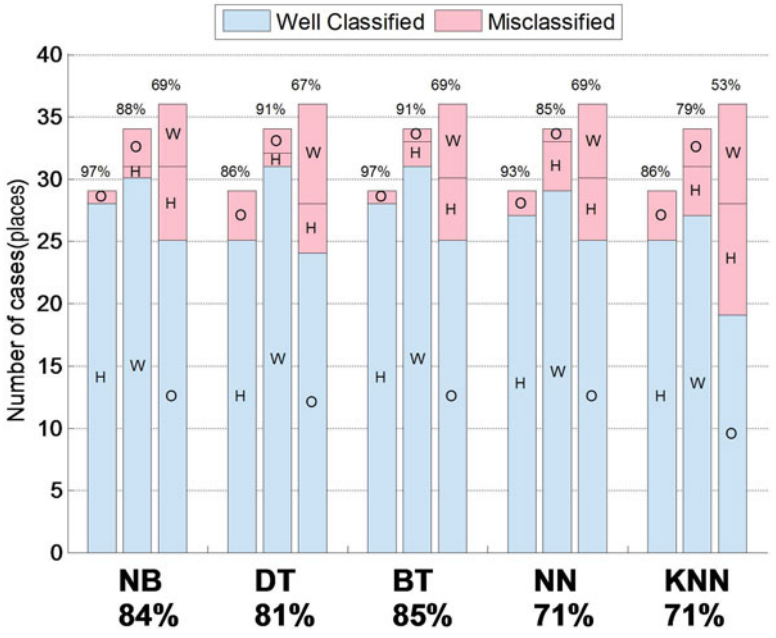

Fig. 12.5 Classification rates (%) for the place approach

if they have more than f common friends. Two users are connected if they spent certain time together (in Bluetooth proximity). First, we discuss social network analysis and present our homophily metric to be used in context inference. Second, in Sect. 12.3.2.3, we describe how the homophily metric can be used to improve context inference and present an experiment where the metric is used for inference.

SNA *focuses on the discovery and evolution of relations among entities (people, organizations, activities, etc.)* [4]. Intuitively, analysing the user as a social being using SNA can provide relevant (previously unknown) information about the user. In particular, the focus of the work lies on the phenomenon of homophily, defined as *the principle that a contact between similar people occurs at a higher rate than among dissimilar people* [11]. Homophily is shown to be ubiquitous in social networks [19], and can be used for improving context inference [27, 28]. In next section, we present the mathematical means to measure homophily in a network.

12.3.2.1 Quantifying Homophily

Introducing graph notation, $G = (V, E)$ denotes a finite undirected graph with nodes $V = \{v_1, v_2, \ldots, v_n\}$ and edges $E = \{e_1, e_2, \ldots, e_m\}$, where $n, m \in Z$ are the number of nodes and edges in G, respectively. E contains the edges or unordered pairs of nodes

$$e_k = (v_i, v_j) \quad \forall i, j \in \{1, \ldots, n\} \quad k \in \{1, \ldots, m\}$$

Given the previous graph definition, we characterize and measure the phenomenon of homophily. Next, we explain the derivation of the initial homophily indicator Hom, in order to quantify the potential to which homophily is exhibited in network G. Since homophily emerges from the context, attribute c_i is used in the formulation of its definition. Note that c_i can be any binary context attribute in the network such as user gender. Define two types of nodes in G, according to the binary value of c_i, namely types p and q, where $V_p(G)$ and $V_q(G)$ are the sets of each type of node. The number of elements in each set is given by n_p and n_q. We consequently also define two types of edges, where edges between nodes of the type and of different types are called homogeneous edges $E^+(G)$ and heterogeneous edges $E^-(G)$, respectively. Considering complete graph K, spanned from G, basic graph theory gives

$$|E^+(K)| = \frac{n_p(n_p - 1)}{2} + \frac{n_q(n_q - 1)}{2}$$
$$|E^-(K)| = n_p n_q$$

Next, we define r_G^+ and r_G^- as the ratios of homogeneous and heterogeneous edges present in G, respectively, with respect to the homogeneous and heterogeneous edges in K.

$$r_G^+ = \frac{|E^+(G)|}{|E^+(K)|}, \qquad r_G^- = \frac{|E^-(G)|}{|E^-(K)|}$$

The rationale behind is that if ratio of homogenous edges is significantly larger than the ratio of heterogeneous edges, the network exhibits homophily. For instance, suppose a network representing relationships between people, categorized as either men or woman. If the ratio of homogeneous edges is higher than heterogeneous edges, it means that people with same gender are more likely to connect than with different genders.

Following that rationally and assuming at least one edge is present in G, we define our homophily indicator Hom for graph G as

$$\text{Hom}(G) = \frac{r_G^+ - r_G^-}{r_G^+ + r_G^-}$$

The range for the homophily indicator Hom is $[-1, 1]$. Positive values indicate that the network exhibits homophily, while negative values indicate that the networks exhibit heterophily, i.e., users have relationships with other different to them. When Hom is close to 0, between $-\varepsilon$ and $+\varepsilon$, the system does not exhibit homophily. ε is thus the homophily threshold and it varies in different networks, depending on the size of the graph and the density of edges. The threshold is the way in which one deals with translating the theoretical definition of homophily into a practical working definition, i.e.,

$$G \text{ exhibits } \begin{cases} \text{homophily if } \text{Hom}_G > \varepsilon \\ \text{no homophily if } |\text{Hom}_G| < \varepsilon \\ \text{inverse homophily if } \text{Hom}_G < -\varepsilon \end{cases}$$

as mentioned by Easley and Kleinberg [10].

12.3.2.2 Using Homophily to Improve Predictions

Once homophily has been measured in the system, it can be used to improve our predictions. This work assumes the homophily indicator to be constant in time, which will assist in inferring missing information [27].

In concrete, we consider the network over time period D. By discretizing G into L periods, each of duration W, we obtain the sequence of successive graph states

$$G = (G_1, G_2, \ldots, G_L),$$

such that

$$LW = D.$$

We assume homophily to keep constant in time, mathematically expressed as

$$\mathrm{Hom}_s(G_1) = \mathrm{Hom}_s(G_2) = \cdots = \mathrm{Hom}_s(G_L).$$

As result, we present two methods to infer links in a social network:

1. The **Random Method (RM)** does not consider the effect homophily for link prediction
2. The **Structural Homophily Randomized Method (SHRM)** considers homophily in the network and makes its prediction assuming it constant.
3. The **Deterministic Homophily Method (DHM)** considers homophily in the network and makes its prediction assuming it constant. However its values are crisp, there are no levels of homophily, it can take values -1, 0 and 1.

These methods are mathematically derivated. Although these derivations are out of the scope of this book, they can be found in [27].

12.3.2.3 Experiment

We apply the aforementioned methods for context prediction to the Nodobo dataset. Nodobo is an open and publicly available dataset that contains social data of 27 senior students in a Scottish high school [8]. The data consist of cellular tower transitions, Bluetooth proximity logs and communication events, including calls and text messages.

From Nodobo dataset, we construct a series of graph: We split the data into L different periods of size W. For each period i, we construct the graph G_i, obtaining the whole graph

$$G = (G_1, G_2, \ldots, G_L)$$

For constructing each graph G_i, the users are the nodes of the graph. We include an undirected edge between nodes if they have been in proximity for an average of 60 min a day, decision made based on previous analysis of the data. In our case, an edge (v_i, v_j) meets the homophily condition if nodes v_i and v_j have at least f common friends.

For experimentation, we have chosen different scenarios or ways to partition the data where the variables have the following values:

- Experiment A: $W = 15, f = 4$, Hom$= 0.38$ (low)
- Experiment B: $W = 15, f = 2$, Hom$= 0.48$ (medium)
- Experiment C: $W = 35, f = 2$, Hom$= 0.67$ (high).

Once the graph was built with all the information, we tried to infer the evolution of the graph, that is to infer the status of the graph G_{t+1} based on its former status G_t. The graph was built from the data in many different ways. In all of the experiment, the predictions considering homophily, using SHRM were better than without using

homophily, achieving improvements ranging from 20 to 119% with respect to the random method. The table with further details about calculation, the results and further interpretation can be found in [27].

This section has shown how homophily can be used in social networks to improve information inference. Although the methods presented for inference aimed at inferring contextual information in different slots of time, we can use the same idea to infer missing information in the present. This is indeed very powerful. For instance, other research has shown that homophily can be used to infer relevant user information such as geographic location, interests and schools attended [21].

12.4 Applications Areas

CAS for mobile computing has myriads application areas [25]. Its importance is as well increasing, since sensors are becoming more widespread everywhere due to their low cost and last technological developments. The quantity of information for one user are plenty: several user devices such as computer, smart phone and tablet; sensors everywhere, in electronic devices, at home and even in public transportation, weather information; access to social networks where information can be pulled through APIs, etc. Some applications that can benefit from CAS are:

- **Location-based services** that adapt to user position [24]. Geo-fencing has become also a hot topic in LBS [22].
- **Information providers** that would help users process the overwhelming amount of web information. AmbiAgent is agent-based infrastructure for context-based information delivery [18].
- **Recommender systems** work similarly to the information providers. The specific task is to estimate user-specific rating for specific items based on user information [2], e.g., Amazon recommender system.
- **Education** may benefit as well from CAS. There is a trend to opt for personalized learning services instead of one-size-fits-all solutions, such as UoLmP, a context-aware adaptive and personalized mobile learning system that supports semi-automatic adaptation of learning activities. User information can be used to boost learner motivations or objective setting [29].
- **Sports**. There are plenty of Context-aware Services in this area, from sport tracking applications such as PureRunner, to sport partner finding applications such as buddyup.
- **Health domain** benefits as well. For example, the home-care context-aware computing (HoCCAC) multi-agent system is designed to maximize task-planning schedule discovery and react autonomously according to changes in the hospital environment [30].
- **Traveling and Tourism**. Some tourism-related context-aware systems and applications can be found in the literature [6, 20]. The context-aware GUIDE is an intelligent electronic tourist GUIDE that presents to visitors information tailored to both their personal and environmental contexts [6].

- **Logistic** applications are expected to come, specially given the trend of opening public transportation data [14]. Some applications already provide information of how to go from point A to B in public transportation or information related to road congestion for those driving their cars.
- **E-democracy**. Studying opinion and sentiment of people in social media is a means for governments to perceive their citizens' insights, and worries. This information is of value to assist decision making. The project Advanced Large-Scale Language Analysis for Social Intelligence Deliberation Support (ALL-SIDES) [9] uses language analysis to understand citizens' opinions.
- In **smart cities**, a variety of networked sensor-based systems and devices are deployed on the scale of cities [31]. The development of CAS is crucial for *smart homes* to succeed [32].
- **Crowd-based applications** are those matching needs of some users and available resources of other users. *airbnb* is an example of this type of application. Crowd-based application can be used as well for car-sharing, participatory involvement in local activities, voluntary works, e-commerce, crowd-based logistics and peer-expert services to mention a few.

12.5 Conclusions

Context-aware Systems are of great importance when dealing with user information, providing it to third parties to build personalized services. The current *de facto* technologies allow services to be built based on physical position, LBS.

To achieve (more) personalized services, CAS need some component to deal with the user context and provide it to third-party applications if permissions are granted. Our implementation of this component is the Context Engine.

There have been a lot of research made on CAS. These CAS should be able to manage and reason with different types of contextual information. Some user information can be considered factual. This information can be extracted from user profiling information in mobile phones or social networks. Other information is considered vaguer in the sense that it has not been asserted by the user, but inferred based on user information such as usage logs. There are already plenty of these inference tools.

Future work may focus on more complex context queries that allows for deeper understanding of the user context, i.e., combining several contextual attributes.

Understanding user context ultimately allows us to understand the user's situation and needs. In a world with plenty of user information about anything, personalization seems to be the key for reducing the overload of information to which we are exposed. On top of that, we expect more personalized services to appear in the markets in the near future [25].

Acknowledgements This work was financially supported by EU FP7 Marie Curie Initial Training Network MULTI-POS (Multi-technology Positioning Professionals) under grant nr. 316528.

References

1. G.D. Abowd et al., Towards a better understanding of context and context-awareness, in *Proceedings of the 1st International Symposium on Handheld and Ubiquitous Computing*, HUC '99 (Springer, London, 1999), pp. 304–307. ISBN: 978-3-540-66550-2
2. G. Adomavicius, A. Tuzhilin, Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. IEEE Trans. Knowl. Data Eng. **17**(6), 734–749 (2005). ISSN: 1041-4347. doi:10.1109/TKDE.2005.99
3. M. Baldauf, S. Dustdar, F. Rosenberg, A survey on context-aware systems. Int. J. Ad Hoc Ubiquitous Comput. **2**(4), 263–277 (2007). ISSN: 1743-8225. doi:10.1504/IJAHUC.2007.014070
4. N. Belov, J. Patti, A. Pawlowski, GeoFuse: context-aware spatiotemporal social network visualization, in *Proceedings of the 13th International Conference on Human Computer Interaction*, 2009
5. H. Chen, T. Finin, A. Joshi, An ontology for context-aware pervasive computing environments. Knowl. Eng. Rev. **18**(3), 197–207 (2003). ISSN: 0269-8889. doi:10.1017/S0269888904000025
6. K. Cheverst et al., Developing a context-aware electronic tourist guide: some issues and experiences, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '00 (ACM, New York, NY, 2000), pp. 17–24. ISBN: 978-1-58113-216-8. doi:10.1145/332040.332047
7. C. Clifford, By 2017, the App Market will be a $77 billion industry (Infographic), 2013
8. M. Compton et al., The SSN ontology of the W3C semantic sensor network incubator group, in *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 17 (2012). ISSN: 1570–8268
9. DFKI, DFKI LT - ALL-SIDES, 2016
10. D. Easley, J. Kleinberg, *Networks, Crowds, and Markets: Reasoning About a Highly Connected World* (Cambridge University Press, Cambridge, 2010)
11. N.E. Friedkin, *A Structural Theory of Social Influence* (Cambridge University Press, Cambridge, 2006), 254 pp. ISBN: 978-0-521-03045-8
12. W. Gao, T. Xu, Stability analysis of learning algorithms for ontology similarity computation, in *Abstract and Applied Analysis*, vol. 2013 (Hindawi Publishing Corporation, London, 2013)
13. T. Gu, H. Keng Pung, D. Qing Zhang, A service-oriented middle-ware for building context-aware services. J. Netw. Comput. Appl. **28**(1), 1–18 (2005). ISSN: 1084-8045. doi:10.1016/j.jnca2004.06.002
14. HSL, Open data - Home page, London, UK, 2016
15. N. Kiukkonen et al., Towards rich mobile phone datasets: Lausanne data collection campaign, in *Proceedings of the 7th International Conference on Pervasive Services*, 2010
16. J.R. Kwapisz, G.M. Weiss, S.A. Moore, Activity recognition using cell phone accelerometers. SIGKDD Explor. Newsl. **12**(2), 74–82 (2011). ISSN: 1931-0145. doi:10.1145/1964897.1964918
17. J.K. Laurila et al., The mobile data challenge: big data for mobile computing research. http://research.nokia.com/files/public/MDC2012_Overview_LaurilaGaticaPerezEtAl.pdf (2012)
18. T.C. Lech, L.W.M. Wienhofen, AmbieAgents: a scalable infrastructure for mobile and context-aware information services, in *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems*, AAMAS '05. (ACM, New York, NY, 2005), pp. 625–631. ISBN: 978-1-59593-093-4. doi:10.1145/1082473.1082568
19. M. McPherson, L. Smith-Lovin, J.M. Cook, Birds of a feather: homophily in social networks. Ann. Rev. Sociol. **27**, 415–444 (2001). ISSN: 0360-0572
20. K. Meehan et al., Context-aware intelligent recommendation system for tourism, in *2013 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)*, March 2013, pp. 328–331. doi:10.1109/PerComW.2013.6529508

21. A. Mislove et al., You are who you know: inferring user profiles in online social networks, in *Proceedings of the Third ACM International Conference on Web Search and Data Mining*, WSDM '10 (ACM, New York, NY, 2010), pp. 251–260. ISBN: 978-1-60558-889-6. doi:10.1145/1718487.1718519
22. D. Namiot, GeoFence services. Int. J. Open Inf. Technol. **1**(9), 30–33 (2013). ISSN: 2307-8162
23. O.A. Nyknen, A. Rivero Rodriguez, Problems in context-aware semantic computing. Int. J. Interact. Mob. Technol. (iJIM) **8**(3), 32–39 (2014). ISSN: 1865-7923
24. B. Rao, L. Minakakis, Evolution of mobile location-based services. Commun. ACM **46**(12), 61–65 (2003). ISSN: 0001-0782. doi:10.1145/953460.953490
25. A. Rivero-Rodriguez, O. Antero Nyknen, Mobile context-aware systems: technologies, resources and applications. Int. J. Interact. Mob. Technol. (iJIM) **10**(2), 12–20 (2016). ISSN: 1865-7923
26. A. Rivero-Rodriguez, H. Leppkoski, R. Pich, Semantic labeling of places based on phone usage features using supervised learning, in *Ubiquitous Positioning Indoor Navigation and Location Based Service (UPINLBS), 2014*, November 2014, pp. 97–102. doi:10.1109/UPINLBS.2014.7033715
27. A. Rivero-Rodriguez, P. Pileggi, O. Nyknen, Social approach for context analysis: modelling and predicting social network evolution using homophily, in *Modeling and Using Context*, ed. by H. Christiansen, I. Stojanovic, G.A. Papadopoulos. Lecture Notes in Computer Science, vol. 9405. doi:10.1007/978-3-319-25591-0_41 (Springer International Publishing, 2015), pp. 513–519. ISBN: 978-3-319-25590-3 978-3-319-25591-0
28. A. Rivero-Rodriquez, P. Pileggi, O. Nykanen, An initial homophily indicator to reinforce context-aware semantic computing, in *2015 7th International Conference on Computational Intelligence, Communication Systems and Networks (CICSyN)*, June 2015, pp. 89–93. doi:10.1109/CICSyN.2015.26
29. L. Shi et al., Contextual gamification of social interaction towards increasing motivation in social E-learning, in *Advances in Web-Based Learning ICWL 2014*, ed. by E. Popescu et al. Lecture Notes in Computer Science, vol. 8613. doi:10.1007/978-3-319-09635-3_12 (Springer International Publishing, 2014), pp. 116–122. ISBN: 978-3-319-09634-6 978-3-319-09635-3
30. B. Skov, Th. Hegh, Supporting information access in a hospital ward by a context-aware mobile electronic patient record. Pers. Ubiquitous Comput. **10**(4), 205–214 (2006). ISSN: 1617-4909. doi:10.1007/s00779-005-0049-0
31. USA Information Resources Management Association, *Computer Engineering: Concepts, Methodologies, Tools and Applications*, 1st edn. (IGI Global, Hershey, PA, 2011). ISBN: 978-1-61350-456-7
32. D. Zhang, T. Gu, X. Wang, Enabling context-aware smart home with semantic technology. Int. J. Hum.-Friendly Welf. Robot. Syst. **6**(4), 12–20 (2005)

Chapter 13
The Impact of Galileo Open Service on the Location Based Services Markets: A Review on the Cost Structure and the Potential Revenue Streams

Anahid Basiri, Elena-Simona Lohan, and Terry Moore

13.1 Introduction

Location Based Services generate the biggest portion of revenue within all GNSS applications; more than half of all Global Navigation Satellite System (GNSS) revenue is from LBS and it is predicted to get 62% of the whole global GNSS market revenue by 2020 [8]. Mobile phones and smartphones, tablets, portable computers, and fitness devices are some of platforms to receive/run location based applications and services. There are more than 2.5 Billion mobile devices with the Global Positioning System (GPS) capabilities globally [17]. However, the wide range of advantages of receiving multi-GNSS signals has made manufacturers and also users willing to produce/have multi-GNSS enabled devices (i.e. equipped with multi-GNSS antennas, the front-ends, baseband processing and navigation solution software). This has resulted in a competition between GNSS systems to be the second in line, following GPS, in the market. On the other hand, many users all around the world have got used to the free-of-charge GPS signals. The success of GPS in generating a worldwide size market may put the financial success of another GNSS system under question; Does the GNSS market really needs another satellite navigation system besides the current two fully functional ones, namely GPS and GLObal NAvigation Satellite System (GLONASS)? Can another GPS-like system be financially successful and generate reasonable amount of revenue

A. Basiri (✉) • T. Moore
The Nottingham Geospatial Institute, The University of Nottingham, Triumph Rd, Nottingham NG7 2TU, UK
e-mail: anahid.basiri@nottingham.ac.uk; terry.moore@nottingham.ac.uk

E.-S. Lohan
Tampere University of Technology, Korkeakoulunkatu 10, 33720 Tampere, Finland
e-mail: elena-simona.lohan@tut.fi

J. Nurmi et al. (eds.), *Multi-Technology Positioning*,
DOI 10.1007/978-3-319-50427-8_13

while free-of-charge GPS signals are available globally? Can another GNSS can have a significant impact on the LBS markets and/or improve quality of such services? The European Union has been developing and deploying a civil-based GNSS; however, the Galileo programme has already faced several delays, which resulted in the increase of cost and also possibly losing some parts of the global markets, such as potentially big market in China and India as they are now deploying their own Satellite Navigation System (the Indian system is regional), the GNSS market is becoming more and more crowded, as more satellite navigation systems are being developed and deployed within Galileo's delays. On the other hand, beside all the political motivations, the Galileo program is based also on many technological and social justifications [14]. EU views Galileo as a technologically complement/backup to GPS not a rival [4]; the 2004 agreement between EU and the USA regarding compatibility and interoperability of GPS and Galileo makes it easier to benefit from availability of more satellites in view. Many LBS applications can benefit from availability of more satellites that may result in better continuity and accuracy of positioning services. This can help emergency, security and safety related applications, which directly increase the quality of life of people. Also more accurate and continuous navigation services, such as in car navigation services, can help more people to save more fuel and time. In addition, being the only civil-based GNSS, as all the others are military-controlled systems, puts Galileo in a more reliable position for many other critical and sensitive applications such as timing for bank transactions and oil industry-related applications, as many countries can rely on American or Russian military-controlled systems now. These are only a few examples of future applications where Galileo can play an important role and therefore there is a global market to be shared with Galileo. In addition to the complementary role of Galileo to GPS, Galileo will be a complete standalone GNSS and if it only target people within Europe as its future users, i.e. mandated by the European Commission (EC), the annual core revenue from device shipment and services in a few years can be almost equal to the whole cost of development, validation and deployment of the whole system of Galileo. Galileo also plans to deliver a Commercial Service (CS) possibly at a fee for professional users, which may change the landscape of the revenues that Galileo could generate. The Galileo CS foresees to offer precise point positioning and code encryption services for authentication, complementing the open service authentication which would be provided for free. However, as the Commercial Service specifications have not formally been endorsed by the Galileo programme at the time being, they are not quantified in this chapter.

This chapter reviews the impact of Galileo, the European GNSS, on the LBS market by considering both positive and also skeptical economic views on Galileo. The chapter studies the cost structure of Galileo and then estimates economic impact of it on the global LBS market and also its own generated revenue by 2022. This chapter assesses the cost structure and the potential revenue streams of the Galileo. The chapter estimates the impacts on Galileo, when becomes fully deployed and functional, on LBS, as the largest revenue generator among all the GNSS

applications. Next section reviews the cost structure, the questionable features and the skepticism regarding the financial aspects of Galileo system. Then in section three the potential markets and opportunities, unique features and enabling services of Galileo will be discussed. Also some suggestions and recommendations are made.

13.2 Galileo Cost Structure and Skepticism

The necessity of having another GNSS system, in addition to GPS, has been under question; many people all around the world, including the EU taxpayers who have been paying for full costs of Galileo, and have got used to free-of-charge GPS signals, may ask whether launching and maintaining another GPS-like system is financially justifiable. This section reviews views against and for the financial aspects of Galileo including its investments and cost structure and potential revenue streams when it gets to its full functionality level. The total cost of Galileo, including development and validation (€2.4 billion), deployment (€3.4 billion) and also 20 years of support, operating costs (20 * €800 million p.a.) and deployment phase of The European Geostationary Navigation Overlay Service (EGNOS) (€1.1 billion) has been reported as €23 billion. However, these figures and numbers and all cost estimations have been subjected to many changes since 2000 when the initial estimates from EC were released [3]. For example, deployment cost was initially estimated about 2.15 billion euro while in EC report in 2013 it was more than double folded. Table 13.1 shows the detailed cost estimations so far. Facing delays in

Table 13.1 Galileo cost estimation [1, 5, 6, 16]

Phase	Initial estimates	After PPP failure estimation	EC estimate report	EC MEMO report
Year	2000	2007	2010	2013
Development and validation (Million EUR)	1100	2100	2100	2400
Deployment (Million EUR)	2150	3400	5000	4500
Operation p.a. (Million EUR)	220	312	750	800
20 years of operation and maintenance (Million EUR)	4400	6240	15,000	16,000
Total (Million €)	7730	11,820	22,180	22,900
Public funding resource	33.63% from EU taxpayers	100% from EU taxpayers	100% from EU taxpayers	100% from EU taxpayers

different phases of the project, public–private partnership funding negotiations and scheme transformation [13], delays in plans and launches and most fundamentally time-consuming process of reaching agreements by all EU member states are some of the reasons of increasing Galileo costs. In below some of these reasons are discussed.

One of the fundamental reasons of increasing the cost of Galileo has been the EU decision-making process in general. This decision making is influenced by a large number of member states (27 EU countries) in comparison with other systems which are being run by military authorities of only one country. Consequently, the management and decision-making process for Galileo are much longer than for other similar systems. This has already resulted in several delays, change in funding schemes and even changes in technical plans for Galileo. Reaching an agreement has been a quite time-consuming process as there are more negotiations and discussions on different member states' concerns and issues. For example, member states had long discussions and negotiations over where the system's activity centre (including ground infrastructure and headquarters) would be located. Obviously for other GNSS systems, which are run by military/air force authorities of one country, this was not an issue [20]. Such negotiations have resulted in delays and changes, and this also increased the Galileo deployment cost. One of the best examples of this can be having reached the agreement on financial supports of Galileo, which leaded the whole project to a pause and consequently added €103 million to the project cost; Between July 2005 and December 2005, discussions and negotiations between the member states and the private investors regarding an agreement on how to financially move forward, led the whole project to a 6-month of pause with no progress. According to the European Court of Auditors this period of doing nothing but negotiations, by itself, added €103 million to the cost of the project, i.e. additional €17 million per month to the cost of the project.

The GNSS markets are becoming more and more crowded as other satellite navigation systems, such as Russian GLONASS, Chinese BeiDou 2, and also some regional systems including Indian Regional Navigation Satellite System (IRNSS) and Japanese Quasi-Zenith Satellite System (QZSS), are being or have been developed and deployed. While other GNSS systems are making the market crowded, Galileo is still facing some delays. In addition to having more competitors in the GNSS market, Galileo has lost some parts of its to-would-be potential markets, such as quite big markets in China and India.

The continuous delays in Galileo deployment has arisen questions regarding its future markets and profitability of Galileo services when it gets to its full functionality level. In addition to have more market competitors, the delays in the Galileo programs have damaged its management robustness, reputation and expectations regarding economic and also technical sustainability. These damages and threats of having more intimidating market competitors may become more and more serious issues if Galileo faces more delays and unfortunately some of these issues, such as losing the market in some regions where a well-established competitor of Galileo is in control of a full functional system, which could hardly be compensated.

There are many who have had deep reservations about the cost of Galileo from the outset and, in particular, the uncertainties that exist about what the precise end-cost will be. Beside the increasing cost of Galileo, there are some skepticism regarding future market of Galileo. On the other hand, there are many potential areas from which Galileo can generate revenue or have indirect impacts. Next section studies Galileo future market and also discusses its potential markets, the revenue streams, and the impact of Galileo on the LBS market.

13.3 Galileo Potential Revenue Streams and Opportunities

Developing future markets for Galileo encounters some challenges and skepticism and enjoys some potentials and opportunities. As the GNSS market gets more and more crowded, Galileo cannot play a major role in the big markets of GNSS in China, Russia and India, however, being the only GNSS system run and controlled by non-military sector makes it easier for other countries to trust [15]. Being the only civil-based system can potentially make Galileo as the third or even the second GNSS system being used by almost all countries as there is a trust issue due to current political situation between the key GNSS players, i.e. the USA, Russia and China [2]. In this regard, the interoperability, compatibility with GPS based on official agreements with the USA can significantly enhance Galileo's role as the second system in the GNSS market in Europe, the USA, Africa and Middle East, and as the third system in Russia and far east. For the summary of Galileo strength, weakness, opportunities and threat, see SWOT analysis below (Fig. 13.1).

The GNSS market share in the USA, Europe, Africa and Middle East and Asia (excluding China) generates more than 80% of the whole GNSS revenue (see Fig. 13.2) and being the second in such a big and well-established market, i.e. with 60 billion euro of core revenue and 200 billion euro of enabled revenue, could still be a good target for Galileo [10].

The agreement between the USA and Europe recognises that civil GPS and Galileo with compatible and interoperable radio frequency at the user level can increase the number of satellites visible from any location on the Earth and aid the accessibility to navigation signals for civil users worldwide [4]. In addition, due to the political situation and sanctions against Russia, if continued, Galileo can potentially come even as the second system to support by American and European receiver manufacturers [14]. The Europe and USA based manufacturers are 91% of all LBS chipset producers globally, by the number and also by the revenue [8]. In addition, no particular import–export control policy in the EU makes it easy for the Galileo-enabled devices to be shipped. The global annual GNSS core revenue (such as shipment of GNSS receivers and devices) is currently €60 billion [8] and is expected to have a market size of €110 billion in 2022 [8]. Even ignoring the growth as the result of the agreement between the USA and EU regarding GPS and Galileo signal compatibility, and also unfairly assuming that only EU

Strengths
- Having higher positional and temporal accuracy
- Being the only civil-based GNSS
- Having a Europe-size default market (mandate)
- Interoperability, compatibility based on official agreements with GPS (with 100% of GNSS market capability)
- Having very broad range of well-developed potential applications, which are currently based on available GNSS (potentially similar selling points)
- Potentially secure and robust market at least in Europe
- No particular import-export control of Galileo-enabled devices

Weaknesses
- Lack of obvious competitive strength
- Having encountered delays which have damage its reliability in management
- Facing delays resulted in having more competitors, loosing a big part of its potential markets in China and India
- Multi-national management procedure which makes the decision making process time-taking
- Having collapsed PPP has arisen scepticisms regarding financial benefits and returns (even between EU states)
- Application predictability
- Need of having plan transparency

Opportunities
- Political and military management of other competitors (as all are military based, its harder to be trusted and also reach agreements)
- Minimum Europe-wide market size (mandate)
- No particular import-export control and barriers for Galileo-enabled devices
- Fast growing GNSS market with new applications
- Well-developed augmentation system

Threats
- GPS technical threats (jamming, spoofing) remain Galileo's too.
- No obvious market push/demand
- Facing several delays results in increase in cost and loosing some big markets (specially in China, India)
- Upcoming/already developed technologies (e.g. BLE, 4G, Wi-Fi 5G) can be used for many LBS applications (as the major GNSS market segment)
- Having GNSS market more and more crowded
- Having well-established competitors with mostly free of charge services

Fig. 13.1 Galileo—SWOT analysis

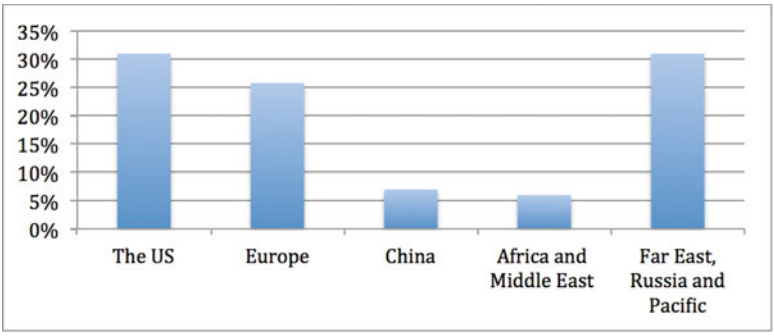

Fig. 13.2 The GNSS market share by region [8]

citizens will only use all the Galileo devices, then giving EU population respect to the rest of the world, it is estimated that Galileo can have an annual market of €5.3 billion. This is only the direct GNSS market revenue and all other enabled, indirect and long-term revenues have not been taken into account yet. The global annual revenue from enabled services and products of GNSS is currently more

than €180 billion and this will get up to €240 billion by 2022 [8]. Again if, unfairly, assume that Galileo-enabled services market is only limited to EU, then Galileo-enabled services and products will have a Europe-wide market of about €30 billion. In addition to the political situation, which can bring a potential opportunity for Galileo to be used in many countries and for many sensitive and commercial applications [2], there is another unique feature of Galileo, which can help Galileo to establish its role in the GNSS markets; the signal authentication can be another motivation for receiver manufacturers to produce Galileo-enabled chipsets. Although according to the current service layers of Galileo [9], i.e. Open Service, Search and Rescue and Public Regulated Service, authenticated signals will not available for open service, it would be possible for Galileo to make most of such unique and enabling service if it became available for everyone. There are several revenue streams and financial justifications, particularly targeting a replacement for mandating Galileo, behind this suggestion. The authors believe that providing the authenticated signal of Galileo for free would be a good opportunity to attract a huge market within the EU and even beyond the EU borders [19]. This can be done even without mandating Galileo, which might be legally and politically challenging task. According to market studies, 6–7% of the value of all the finished goods and services produced within EU borders in 2013 (GDP), which is estimated about €800 billion, currently relies on satellite navigation signals provided by the GPS [18]. This dependency equals about €850 billion (European Commission report, 2014) (based on 6.5% of the EU28 GDP of €13.075.000 million in 2013). A disruption of the GPS signal would therefore have a major impact on the European economy. The following range of services rely on GNSS: Location Based Services (LBS) are a class of information services that use location/geographic data—the LBS market and number of users and devices have faced enormous growth over the last years. Its revenues are estimated to €735 million in 2013, expecting to grow up to € 2.3 billion in 2018.

The road segment consists of many different applications, such as Advanced Driver Assistance Systems (ADAS), Road User Charging (RUC), Pay-per-use insurance (PPUI) and road traffic monitoring. It is a market in transition; over 50 million units are in use within the European Union [8] and its relative share in the GNSS market is expected to continuously grow. The navigation systems used by the aviation sector demand the highest robustness and integrity. The use of GNSS within all aviation segments is expected to increase over the next decade reaching a penetration of over 90% by 2022. GNSS devices in the rail segment are used by major train companies to track all of their trains. The use of GNSS is expected to grow significantly within the next years. This is a good sign for European manufactures, because the market is dominated by Europe when it comes to the shipment of Rail GNSS devices, more than 5000 units in 2012 [8]. One of the first adopters was the maritime segment. Global shipment of GNSS devices within this market hovers around 100,000 in 2012, just over 25% of Maritime devices serviced the European market. The economic importance of GNSS to agriculture is modest yet advancing. With technological adoption established at a high level in Western Europe major growth is foreseen in Central and Eastern Europe.

European Space Agency (ESA) expects the value of businesses relying on satellite navigation to grow by roughly 11% a year, and to be worth €244 billion by 2020 [8]. This dependency is increasing since more web and mobile transactions, services and applications have become commonly available. This dependency is the case for almost all the regions and countries. Due to sensitivity of the applications and services, such as bank transaction synchronisations, and also relatively large market relying on potentially vulnerable GPS signals, there is a high demand to have a reliable (back up) system. Providing authenticated signals by Galileo for free (i.e. as a part of the Open Service of the only civil-based GNSS) would be a great motivation for many countries to switch to and rely on Galileo. This large market may not bring many revenue streams directly but indirectly makes many systems and services use Galileo signals and build the trust on Galileo system for other sort of services, which may bring more money.

In addition to enabling services, there are some direct revenues; many sensitive services, such timing and synchronisation, aviation and maritime may require to receive authenticated signals and this only will be provided by Galileo. Although timing and synchronisation, aviation and maritime applications are less than 3% of the global GNSS market share [8], about 30% of component manufacturers in their value chain are based in Europe. The authentication of the signal and also having access to local, Europe-based, companies make it possible to make these valuable and sensitive segments of the GNSS markets dominantly Galileo-based.

Galileo is the only civil-based satellite navigation system, therefore it would be more reliable for many countries to use Galileo's signals for their sensitive and critical applications such as power grid synchronisation, electronic trading and banking, mobile phone network, air traffic management. Galileo will always be controlled under non-military sectors [7]; this will bring secure signals, which are required for such critical application. In this regard, there would be a welcoming market for Galileo. In addition to business side, Europe should not rely on GPS signals anymore due to the sensitivity of such applications, political issues and also economic reasons [11, 20]; EU annual GDP, which relies on GPS signals in 1 year, is enough for whole Galileo system deployment. Therefore Galileo is a politically and financially justifiable system as it can bring bigger markets and also it is in line with EU security policies [11].

European Commission officials have publicly and recently stated that they are considering how to stimulate Galileo use; in particular through regulatory measures requiring that navigation equipment be installed on aircrafts, cars and other platforms. However, according to the US government representative at the international satnav forum (GPS World, 2015) such mandating use of specific GNSS services for applications such as emergency calling, road tolling and LBS applications could violate the terms of World Trade Organisation (WTO) agreement, including promoting open market access and the agreement on Technical Barriers to Trade (TBT), and the general agreement on Trade in Services (GATS), that many nations including all six satellite navigation service providers (i.e. USA, EU, China, Russia, India and Japan) have signed [6, 14].

In order to avoid such conflicts and agreement violations, it is recommended to discuss and assign technology-neutral, platform-based standards [6]. For example, the US E911 rules specify positioning accuracy requirements while allowing to choose the best technical solutions accordingly.

As it shown in Table 13.1 and explained previously, Galileo's cost has been increasing due to changes in plans, delays and management issues. This has caused scepticism regarding investments and future turnover.

One of the most important reasons of having increasing cost is the collapse of the Private–Public Partnership (PPP) in 2007, see Table 13.1. In addition to cost increase, having investment structure changed raises some questions about stability and profitability of Galileo when it gets to the full functionality level [13]; If would-to-be committed funding bodies and private investors doubted they can get back their investments why other shouldn't?

Looking at the GPS investment structure may help to answer this question. GPS was built at a considerable cost by the US taxpayers and the returns to the American economy have shown that such investments have been rapidly paid back, actually many times over. GPS market forecasted in the infancy years of GPS (1984–1988) barely got to one billion USD; in 2001 GPS market estimated to get 5.2 billion USD, this doubled in only 4 years [12]. Nowadays core revenue of GPS is more than 50 billion USD [4]. This can be considered as a good model for Galileo; if publicly funded GPS has repaid its debt to the US taxpayers even more than once, why Galileo cannot do so.

It has estimated that in 15 years after full functionality of Galileo, Galileo can pay back its debt to the public sector from revenues generated through the sale of guaranteed Galileo signal accuracy, mainly to governments. But most of these profits will be made toward the end of this period, this might be the justification why PPP had collapsed; private investors may need the return of their money directly and off course earlier than this time and more importantly due to repeating delays this time could postpone even more.

Of course, the same revenue stream and structure of GPS is not expected from Galileo due to having more crowded market competitors. Early GPS entrepreneurs are now dollar billionaires, but how much room is left in the GNSS market for others? More accurate timing results in more accurate and continuous positioning service. Galileo higher level of accuracy, better signal penetration, more satellite in view, continuity and other parameters of quality of the positioning service can contribute in many LBS applications, including navigation. Galileo is a fully standalone system; however, the market trends, applications requirements and freely available similar service (competitors) push it not be only used as a standalone system. Galileo better signal penetration, for indoor and urban canyon positioning, and Galileo signal compatibility and interoperability with GPS, will bring opportunity to have more satellites in view, can provide LBS users with more accurate and continuous positioning service, see Fig. 13.3. It has been estimated that more accurate and continuous positioning services in conjunction with real time traffic data, only in car navigation, may save up to 460 billion euro by 2020. McKinsey Global Institute (2011) estimated that people all around the world will save more than €460 Billion

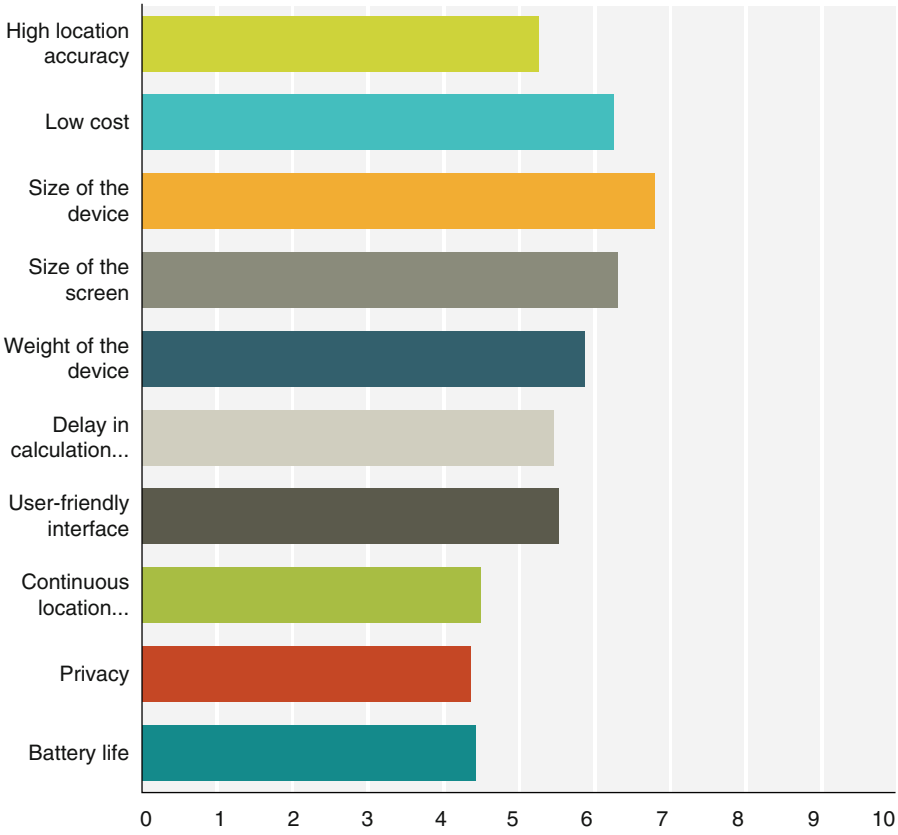

Fig. 13.3 Quality of LBS service/device from the perspective of the users (a conducted survey by the authors)

annually by saving in "time" and "fuel" by 2020 as a result of use of personal location data captured by either personal smartphones or In-Car navigation systems. One of the most obvious examples of this is time and fuel saving of avoiding congestions and suggestion of alternative routes. Galileo can definitely contribute in regards of more accurate and continuous positioning data capture, however, this contribution will save more money rather than generate revenue. Just in Europe currently there are more than 500 million smartphones and tracking devices in Europe, which is more than 25% of global share.

Savings from less fuel consumption and less time spent on travels can be beneficial for industry, although the economic contribution of Galileo is not as obvious as its direct revenue, or at least is not appreciated by investors in the same way. The contribution of Galileo in having less carbon dioxide due to less fuel consumption (which is due to more accurate and continuous car navigation services) would be quite significant as well. Considering all of savings from

Galileo contribution in navigation, discussed above, Galileo can contribute about 25% of the whole €460 Billion, annually. This would be far more than its initial investments. However, this contribution is more in the form of savings rather than revenue. Another example of such savings and long-term returns is spill-over effect. A research project by Oxford Economics shows that the technical advances that come about as a result of research and development investments in the space industry are transferred to firms in other sectors in the form of "spill-over" effect, which is surprisingly large; the social return of around 70% in long-term is expected. If only development and deployment phases' costs (€2.4 billion + €3.4 billion) are considered as R&D investments, then an increase of €4 billion in long term in other sectors (such as health, transport and computer science) is expected. As it was shown by two simple examples, there are many applications, verticals and segments, which can benefit from Galileo services; however, these benefits are not all in the form of revenue. They are also in forms of cost saving, and spill-over effects, which are not the most favourable form of revenue streams as return of the investment will be long -term or even it spills over in other sectors. So Galileo would be able to return its initial investments and even far more than that, however, this returned money would go to the pocket of its "real" investors, i.e. EU taxpayers, in the form of savings. According to a recent study, this window of opportunity gives Galileo an estimated worldwide GNSS market size of 1800 million users in 2010 and 3600 million users in 2020. Nevertheless, other studies support the case that further benefits will arise from route guidance, improved personal emergency, management of taxis and ambulances [16], less pollution by reduction of travel times and creation of 140,000 jobs. An estimate of all benefits for the period 2000–2020 is illustrated below: Economic benefits of €62,000 million; Social benefits of €12,000 million; Total benefits of €74,000 million. The total investment cost for the Galileo operable system is some 2300 million Euros (EU MEMO-11-717). From 2008 onwards the annual cost will be around 220 million Euros including operations, maintenance and replenishment. The above-mentioned opportunities would be generated if there will be no more delays in deployment and maintenance of Galileo. This assumption is not too ambitious as the findings are now secured and secondly there will be no design-changing negotiations over the technical or financial aspects of the system. In addition, there is an essential need to promote Galileo service, unique features and advantages of using its signals. This should target almost all market chain and user types, including chipset manufacturers, devices and application platforms, app developers, stakeholders, map and other spatial content providers, policy makers and researchers. Also the relevant policies, legislation and standards, such as import–export control, positional requirements and policies regarding upcoming E112, outreach and public engagement policies and plans need to be promoted as soon as possible to ease the Galileo market penetration. These can be done by providing R&D funding to research institutes and SMEs, having workshops, seminars, hackathons and conferences to make developers and researchers aware of the unique features and multi-GNSS capabilities of Galileo, legal aspect done through industry, public and academia engagement and outreach initiatives, research and development funding and grants, export promotion and facilitation policies.

13.4 Conclusions

The success of GPS in generating such a big worldwide market may sound intimidating to another similar GNSS system and financial justification of the new system might be criticised by many. This chapter reviewed the impacts of Galileo, the European GNSS, on LBS markets, the Galileo cost structure and potential threats and also its strength and opportunities to generate revenue for all EU tax payers, who have invested in Galileo. The chapter has comprehensively reviewed the financial aspects, costs, threats and weakness of Galileo, and then estimates its economic impacts, strength, and opportunities particularly for LBS market when Galileo gets to its full functionality level. Having faced several delays in development and deployment phases, being run and managed by larger number decision makers, i.e. EU state members, increase in cost of the whole project and also change in the cost structure over last years, have risen some questions and concerns, such as joining a more crowded market as more satellite navigation systems are being deployed by other regions such as Russia, China, India and Japan. On the other hand, unique features of Galileo, such as authentication of the signals, higher accuracy and signal penetration, and being the only civil-based GNSS bring several market opportunities that can only be provided by Galileo. In addition there are even more opportunities if Galileo could establish itself as a multi-GNSS system in combination with GPS and similar systems; Galileo full interoperability and compatibility with GPS, more satellites in view and politically stability of the system management (the only non-military based GNSS) can generate more possibilities and opportunities. This chapter made some suggestions and recommendations, based the reviewed market reports, research papers, conducted surveys, interviews and personal and experts opinion, on making most of Galileo's opportunities and opening new revenue streams in the GNSS markets. They include making authenticated signals free for all users to attract more enabling services and sensitive applications, providing funding and grants for R&D projects, public engagements and outreach initiatives, export control facilitation, workshops and meeting to promote and advertise Galileo multi-GNSS potentials and unique features, having no more delays in deployment and maintenance phases and finally furthering the negotiations with other GNSS systems to make it as compatible as possible to establish its third or even second place in the GNSS market.

Acknowledgements This work was financially supported by EU FP7 Marie Curie Initial Training Network MULTI-POS (Multi-technology Positioning Professionals) under grant nr. 316528.

References

1. S. Abbondanza, F. Zwolska, Design of meo constellations for Galileo: towards a design to cost approach. Acta Astronaut. **49**(12), 659–665 (2001)
2. G.P. Ammassari, M.C. Marchetti, Innovation process in the European union: the case of the Galileo project, in *European Socio-Economic Integration* (Springer, Berlin, 2013), pp. 129–146

3. J.P. Bartolome et al., Overview of Galileo system, in *GALILEO Positioning Technology* (Springer, Berlin, 2015), pp. 9–33
4. S.W. Beidleman, *GPS Versus Galileo: Balancing for Position in Space* (Lulu.com, 2012)
5. EC (2016). http://ec.europa.eu/enterprise/policies/satnav/galileo/index_en.htm
6. B. Giegerich, Navigating differences: transatlantic negotiations over Galileo. Camb. Rev. Int. Aff. **20**(3), 491–508 (2007)
7. M.P. Gleason, Galileo: power, pride, and profit. The relative influence of realist, ideational, and liberal factors on the Galileo satellite program. Technical Report DTIC Document (2009)
8. GNSS GSA. Market Report, Issue 4 (2015)
9. G.W. Hein, T. Pany et al., Architecture and signal design of the European satellite navigation system Galileo-status Dec 2002. Positioning **1**(02), 73–84, 2009
10. I.F. Hernandez, Galileo receiver research in Europe, in *GALILEO Positioning Technology* (Springer, Berlin, 2015), pp. 249–271
11. T. Hoerber, Creating ESA, in *Yearbook on Space Policy 2014* (Springer, Berlin, 2016), pp. 243–254
12. L. Jacobson, *GNSS Markets and Applications-GNSS Technology and Applications.* Artech House Publishers, 2007
13. E. Killemaes, Study of public-private partnerships in the European space industry. Ph.D. thesis, Faculty of Science, Universiteit Gent, 2012
14. D. Kong, Shaping a uniform governance structure over Global Navigation Satellite System (GNSS): the way of risk management, in *Uniform Law Review-Revue de droit uniforme*, unw019 Oxford University Press, (2016)
15. J. Lembke, The politics of Galileo. Center for West European Studies (2001)
16. MEMO/11/717, Galileo will boost economy and make citizens' lives easier (2016)
17. Pew Research, Location-based services, smartphone ownership 2013: three-quarters of smartphone owners use location-based services (2013)
18. P. Stephenson, Talking space: the European Commission's changing frames in defining Galileo. Space Policy **28**(2), 86–93 (2012)
19. P. Walker et al., Galileo open service authentication: a complete service design and provision analysis, in *Proceedings of the 28th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2015)* (2015), pp. 3383–3396
20. S.-C. Wang, Surviving the crises: the changing patterns of space cooperation among the United States, Russia, Europe, and China (2010)

Chapter 14
Location Based Services Analysis Through Analytical Hierarchical Processes: An e-Health-Based Case Study

Elena-Simona Lohan, Pedro Figueiredo e Silva, Anahid Basiri, and Pekka Peltola

14.1 Introduction

14.1.1 Motivation

Wireless localisation has become a valuable information for a large number of location based or location-aware services [2]. Significant economic value can be generated with an increased granularity of the location information (i.e., knowing the location at block or room level instead of street or district level) as shown, for example, in [2]. LBS are generally offered to the users by an LBS provider, using one or more localisation technologies, which can belong or not to the LBS provider. For example, an LBS provider can use the information coming from the WiFi network owned by a Mall owner and provide position location information to the user. In this example, the LBS provider and the technology provider (i.e., the owner(s) of the WiFi network) are two distinct units. In another example, a cellular operator can use the information provided by its own cells or base stations to provide the user location. In this example, the LBS provider and the technology provider are the same unit, namely the cellular operator.

E.-S. Lohan (✉) • P. Figueiredo e Silva
Tampere University of Technology, Korkeakoulunkatu 10, 33720 Tampere, Finland
e-mail: elena-simona.lohan@tut.fi; pedro.figs.silva@gmail.com

A. Basiri • P. Peltola
The Nottingham Geospatial Institute, The University of Nottingham, Triumph Rd, Nottingham NG7 2TU, UK
e-mail: anahid.basiri@nottingham.ac.uk; pekka.peltola@nottingham.ac.uk

J. Nurmi et al. (eds.), *Multi-Technology Positioning*,
DOI 10.1007/978-3-319-50427-8_14

Today's Location Based Services have four main characteristics according to [7]:

- they are proactive instead of reactive, meaning that they can be automatically invoked or generated by the network or the application, without a special request from the user;
- they are multi-target instead of single target, meaning that a significant focus from the point of view of the LBS provider is on estimating and inter-relating the positions of several targets rather than focusing on the position of a single target;
- they are application-oriented rather than content-oriented, meaning that the offered LBS is tailored dynamically and interactively to both the user's position and the user mobile applications that are running on his/her mobile;
- they are cross-referenced, meaning that the information collected by the LBS provider about one user's location is typically used to enhance or/and provide location information and location based content to other users who are using the same LBS provider.

Such characteristics have both benefits and drawbacks for both the users and the LBS providers. For example, a proactive LBS requires a continuous user tracking and a continuous updating of the services, which puts more burden on the LBS provider's design, implementation and computational resources. A proactive and cross-referenced LBS also can threaten the users' privacy, as many decisions are taken by the LBS provider without the users' specific request. On the benefit side, in a proactive LBS, a user doing sightseeing in a foreign town can receive pertinent and timely information about all the touristic sites and eating places nearby, without doing a specific search for them, and only by letting the LBS running on his/her mobile. This can improve the overall user experience. A multi-target approach to LBS facilitates a lot of proximity-based applications, such as advertising the offers in the nearby store to all users passing nearby or offering bonus points to first users to check in a certain place.

LBS can be basically split into outdoor LBS and indoor LBS, as shown in Fig. 14.1. Some examples of indoor LBS are: finding your way in a commuting hall, such as airport of railway station, fall detection alarm from the home of your elder relative [30], finding fast a desired shopping item on the shelves of a shopping centre [20], home exercising through gamification with a Kinect camera and a movement detection engine [18], finding lost persons in a partially collapsed diamond mine, or finding fast a parking place in a crowded parking lot [32]. Examples of outdoor LBS are: in-car navigation engine, plane guidance during landing and take-off, under-water asset monitoring or precision farming to increase crop yields. A large pool of LBS are applicable both in indoor and outdoor scenarios, such as geo-fencing (e.g., setting boundaries on the allowed zones of the patients of a hospital), pedestrian navigation such as tourist guidance when visiting an unknown town, helping blind people to navigate or offering health advices based on the monitored level of daily physical activity.

The LBS are typically dependent on the technical parameters of the available localisation engine, such as the accuracy, the availability and the integrity (or

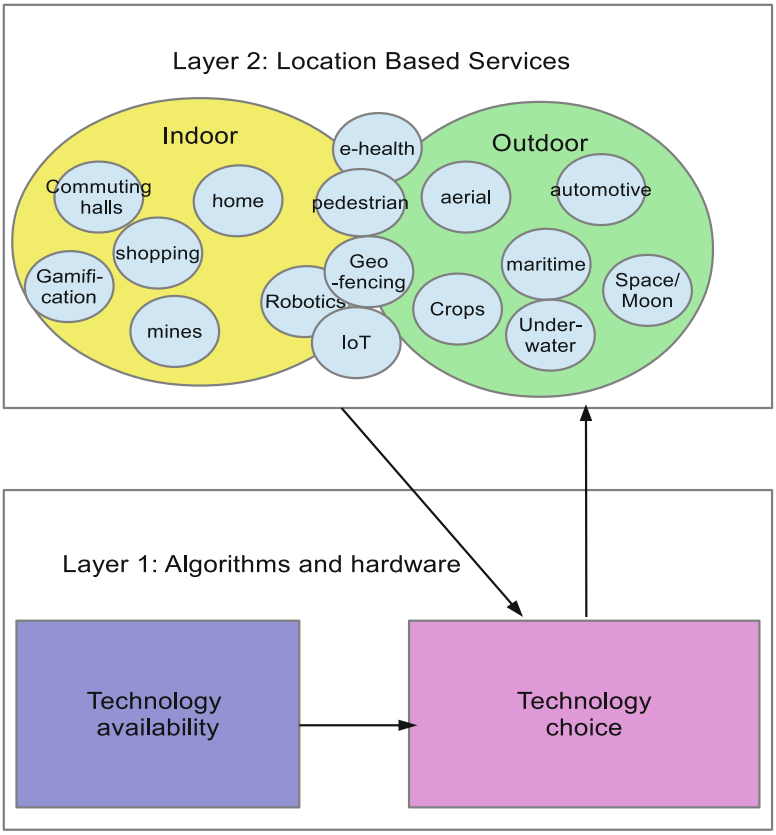

Fig. 14.1 Example of LBS types and decision levels related to LBS

'trustfulness') of the estimated position. For example, an accuracy of few tens of metres can be sufficient for an LBS pointing out towards the nearby touristic monument, but it may be totally inadequate for helping you in finding a desired item on the shelves of a supermarket. The technology choice thus highly influences the feasibility and the functioning of an LBS, as shown in Fig. 14.1. The reverse is also true: an LBS provider with a certain LBS in mind can design the localisation engine and choose the supporting localisation technology in such a way to serve best the target LBS. In addition, the choice of the technology is of course dictated by the availability of that particular technology. If you have on your phone only a WiFi-based positioning engine and you are lost in the middle of a forest in Lapland, such WiFi-based positioning engine will be of little service.

A key question related to LBS is how to design and create services and applications based on available localisation technologies and by taking into account users' needs and various constraints.

This is where the AHP comes into equation, by providing a framework to solve some of these design questions. AHP, developed by Saaty [34], belongs to the category of Multi Criteria Decision Making (MCDM) processes [28]. AHP can help decision makers to choose between various options by taking into account both quantitative and qualitative factors, and it is thus a powerful tool in many of research areas. Indeed, many engineering and research areas are faced with the problem of deciding upon a certain action and deploying resources in such a manner that certain objectives are met and certain constraints are fulfilled. A top-down (hierarchical) approach is often useful to split the target problem into several smaller sub-problems with respect to clearly articulated objectives. Once a hierarchy is achieved, an overall solution can be found by building a multi-attribute cost function and evaluating it according to the set of the objectives. AHP is a 30-year-old concept, used so far in various areas such as project management and organisation [28], mathematical programming [22], Geographic Information Systems (GIS) [40], logistics [24], vascular surgery [35], and incentive to use Body Area Networks (BAN) [1].

When a decision is to be taken, many times different sources of information are available, sometimes conveying conflicting information. Also, knowing more or having more information available is not always a warranty that a better decision can be reached. One of the most comprehensive surveys of the state of the art in decision-making processes modeled in a mathematical framework is found in [11]. The main advantage of using an AHP framework over a one-objective decision process is its flexibility, namely the fact that it allows to deal with multiple conflicting objectives [22] and with subjective information. The main idea of an AHP analysis is described in Sect. 14.2. The e-health domain was chosen to illustrate a case study based on AHP because e-health LBS is highly relevant in today's world, by focusing on how to improve the quality of life of an increasingly elderly society, how to prevent diseases and help millions of persons with physical or mental disabilities and how to motivate people towards a more active and healthier life.

14.1.2 Key Questions Related to Location Based Services

As it will be discussed in Sect. 14.3, there are currently a multitude of wireless localisation technologies that can be adopted for indoor scenarios. Nevertheless, each has its advantages and shortages and there is currently no-fit-for-all solution indoors. Each location based service or application has its own positioning requirement and quality of service priorities. Moreover, for a specific LBS application, a designer can focus on several important criteria, such as availability, positioning and navigation accuracy, cost or power consumption, privacy-preserving level, required infrastructure and device modifications in software or hardware, etc. One of the key questions in designing suitable LBS and supporting positioning technologies is how

to choose the best fit-to-purpose technology. This is a challenge which can be in fact addressed through AHP, as illustrated in Sect. 14.4.

Some of the generic questions that need to be addressed for any LBS are:

- Do we use wearable or device-free solutions?
- Do we use existing infrastructure or create a new one?
- What kind of positioning accuracy is needed for that particular application or LBS?
- Are the users comfortable to use the selected technology? It is easy enough to install, use and maintain?
- What is the reasonable/maximum cost users are happy to pay to get the application?
- Is the positioning technology robust to attacks and interferences?
- What level of privacy users may need?

If the focus is on couple of particular LBS application areas, such as logistics, e-health or emergency, then there are additional questions to be addressed, depending on the target application. Examples are shown in Table 14.1.

Table 14.1 Potential questions to address when designing a specific LBS

LBS area	Example of LBS area-tailored specific questions
Logistics	What is the level of the technology vulnerability (jamming, spoofing, interference, privacy, etc.)? Is this level OK for the target application? Is the technology easily scalable?
Ride sharing or friend finder	What are the nearest persons traveling to or being at the same location as the user? Is the power consumption of the mobile LBS low?
Security or person/vehicle identification	Can the location information be robustly used to identify an user or to make some transactions more secure?
	Is the privacy of the user preserved in the process?
E-health	Are positioning devices easy to carry and easy to use by elderly and children?
	Is the positioning service is contentious and reliable to make sure it always can provide the quality of the service?
	Is the technology supporting fall detection?
	If yes, at which alarm rates?
	Is the technology able to estimate accurately the user mobility patterns?
Emergency	Does the technology offer fast on-line responses or Does it work well only in off-line mode?
	Is the technology valid in both indoor and outdoor scenarios or Is it limited to certain cases only?

14.1.3 Some Definitions and Introductory Concepts

For a better road map into the chapter, the main concepts dealt with in this chapter are enumerated and briefly defined below:

- The Location Based Services (**LBS**) refer to the set of applications on user devices (typically mobile and wireless devices) that need the navigation information in order to function. There are a variety of LBS, ranging from person and pet finding to infotainment, gamification and social networking;
- The Multi-Criteria Decision-Making (**MCDM**) process is the process of taking a decision pertaining a certain problem, by taking into account different conflicting criteria or constraints and different sources of information. One of the most challenging problems in MCDM is to achieve a consensus when many judgments from many decision makers are available;
- The Analytical Hierarchy Process (**AHP**) is a tool for solving a problem or reaching an objective, by defining a multitude of criteria and stating the alternatives possible under each criterion, then taking a decision based on a certain cost function. The mathematical formulation of AHP is given in Sect. 14.2;
- The **pair-wise comparisons**, which are key elements in AHP, are comparisons between any pair of two criteria, studying the relative importance of one criterion with respect to the other one. For example, if we compare the cost of a mobile wireless product with its ease of use and with its battery life, a judgment can be made about which criterion is more important between the two and with how much. Such a judgment can be made either subjectively or based on user surveys or on other empirical data. As an example of a pair-wise comparison in this example, it can be said that the cost is twice more important than the ease of use, and as important as the battery consumption, while at the same time, the battery consumption is three times more important than the ease of use. Such apparent contradictory judgments are possible to be addressed via the AHP process, as it will be explained later in this chapter. One important aspect in AHP is the consistency of subjective decisions;
- The **consistency** in the context of AHP refers to the coherence or logic of a decision process. If the subjective judgments or opinions are too contradictory, then the result of the decision will not be consistent. This may happen, for example, when the judgments are done based on incorrectly collected data, but some amount of inconsistency can also be inherent due to the inherent inconsistency of the human nature [39];
- The **e-health** is a term introduced about a little bit more than a decade ago to encompass a variety of health-related medicine, illness prevention and remote intervention for health problems via Internet or other form of wireless or wired communication. The e-health encompasses nowadays all the health services and health-related information that are delivered remotely to an end-user, such as a patient, a care-taker or a family member.

14.2 What is an Analytical Hierarchical Process?

A decision regarding a certain problem solving can be reached by taking into account several criteria and by dividing the problem into a hierarchical process. There can be multiple hierarchy levels. A block diagram of a two-level AHP decision tree is shown in Fig. 14.2: the first-level hierarchy shows the criteria according to which a decision is reached, and the second-level hierarchy shows the options to be analysed. Each level has a certain priority factor or weight associated with it, here denoted by $w_i, i = 1, \ldots, M$ (first level) and $v_{ij}, i = 1, \ldots, M, j = 1, \ldots, N$ (second level), where M is the number of criteria that need to be analysed when taking a decision and N is the number of options or decisions that can be taken. The meaning of the w_i factors is the importance allocated to the ith criterion, when trying to reach a decision. The v_{ij} factors carry the weight or the 'importance' of the ith criterion when the jth decision is reached. In order to compute the w_i and v_{ij} factors, some intermediary weights at each hierarchical level h are introduced. Let's call them **decision weights** $d_{kl}^{(h)}, k = 1, \ldots, N_h, l = 1, \ldots, N_h, h = 1, 2$. They have the following significance: $d_{kl}^{(h)}$ shows how many time the kth criterion or kth decision from level h is more 'important' than the lth criterion or lth decision from level h. N_h is the number of criteria or decisions or options at the hth level. The issue of the 'importance' will be addressed in more detail after a brief mathematical description of how the decision weights $d_{kl}^{(h)}$ and the **hierarchical weights** w_i and v_{ij} are inter-related. The hierarchical weights are illustrated in Fig. 14.2.

Once the main decision criteria are chosen (as shown in Fig. 14.2) and the decision weights for each hierarchical level are selected (e.g., based on subjective or

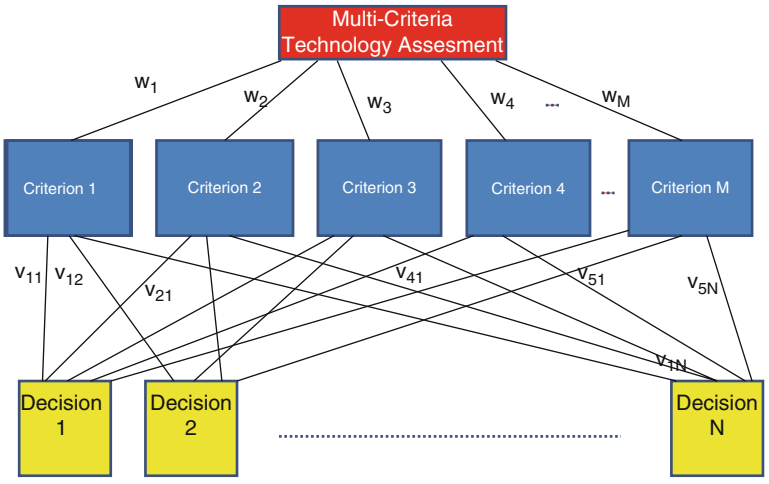

Fig. 14.2 Illustrative block diagram of an AHP

objective information or a combination of both), the pair-wise comparison matrices at each hierarchy level can be built.

A ***pair-wise comparison matrix*** for hierarchy level h (here, $h = 1, 2$) is built as given below,

$$A^{(h)} = \begin{bmatrix} 1 & d_{12}^{(h)} & d_{13}^{(h)} & \dots & d_{1N_h}^{(h)} \\ \frac{1}{d_{21}^{(h)}} & 1 & d_{23}^{(h)} & \dots & d_{2N_h}^{(h)} \\ \frac{1}{d_{31}^{(h)}} & \frac{1}{d_{32}^{(h)}} & 1 & \dots & d_{3N_h}^{(h)} \\ \dots & \dots & \dots & \dots & \dots \\ \frac{1}{d_{N_h1}^{(h)}} & \frac{1}{d_{N_h2}^{(h)}} & \frac{1}{d_{N_h3}^{(h)}} & \dots & 1 \end{bmatrix}. \tag{14.1}$$

Let's denote the elements of pair-wise comparison matrix $A^{(h)}$ as $a_{ij}, i, j = 1, \dots, N_h$. The matrix $A^{(h)}$ can be normalised by the sum of each of its columns, and a normalised matrix $M^{(h)}$ can be obtained. The elements $m_{kj}, k, j = 1, \dots, N_h$ are built as $m_{kj} = \frac{a_{kj}}{\sum_{i=1}^{N_h} a_{ij}}$. Once the normalised $M^{(h)}$ is obtained from the pair-wise comparison matrix, the so-called **priority vector** $V^{(h)}$ at the hth hierarchical level can be defined as:

$$V^{(h)} = \frac{\sum (M^{(h)})^T}{N_h} \tag{14.2}$$

where ^{T} stands for the transpose of the matrix, and the $\sum(\cdot)$ operator is the sum over the columns of the matrix.

After the priority vectors are built starting from the hierarchical weights and by following the above-mentioned steps, the decision weights at each level are the elements of the vector $V^{(h)}$. When the hierarchical weights at each hierarchical level are available (e.g., for a two-level hierarchy, the hierarchical weights are w_i and v_{ij}), the **priority levels** $t^{(i)}$ corresponding to each decision i can be computed as:

$$t^{(i)} = \sum_{j=1}^{N_1} w_i v_{ij}, 1 = 1, \dots, N_2 \tag{14.3}$$

The highest priority level $t^{(i)}$ will correspond to the best suited decision to the considered optimisation problem and the lowest priority level will correspond to the least suitable decision. The AHP decision flow is illustrated in Fig. 14.3.

A simplified example of applying AHP in the context of e-health Location Based Services is shown in Sect. 14.4.

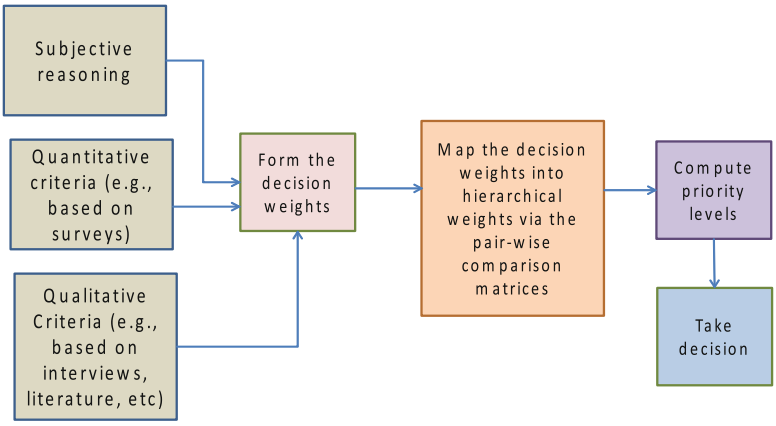

Fig. 14.3 AHP decision steps

14.3 Technologies Which Can Be Used for Positioning

A basic classification of the technologies which can be used to position a mobile node, such as a person or a moving car, is the following:

1. *Wearable technologies*: they require that the users carries a tag, or a transmitter with her/him, e.g., embedded in clothes, as a bracelet or as a portable device such as the mobile phone
2. *Device-free technologies*: certain wireless transceivers, tags and sensors are installed in the area where the users are to be located (typically at home, in hospitals or in shopping malls), but the user does not need to carry on any device. Of course, the positioning accuracy in device-free technologies is typically lower than the one achieved with the wearable technologies.

14.3.1 Wearable Technologies

The wearable technologies are those that can be carried on by the user, either on a mobile device, or embedded in the clothes. The most known that employ some form of wireless localisation or motion detection are:

1. RFID tags: Radio Frequency Identification (RFID) tags are becoming an attractive alternative for e-health applications, due to their tracking and positioning capabilities that can implemented at low-cost [12, 14]. Radio Frequency Identification (RFID) tags can be active or passive [29]. Passive RFID tags can be embroidered in human clothes and they 'communicate' through back-scattered power measurements with an in-room reader. RFID ranges are typically small

(few metres), allowing thus for proximity positioning. For example, RFID-based fall detectors have been studied in [21, 31].

2. BLE tags or transmitters: Bluetooth Low Energy (BLE) is a Bluetooth version meant for low power applications. Wearable BLE tags are already available. BLE-based solutions for fall detection and indoor positioning have been investigated only to a small extent so far [9].
3. WiFi or WLAN tags or transmitters: WLAN, also referred to as WiFi, are one of the most popular and widespread wireless technologies nowadays. They are heavily present around us: in houses, hospitals, universities, commuting halls, etc. Most portable wireless devices have nowadays a incorporated WiFi chipset and clothing embedded transceivers are becoming a reality [27]. Indoor localisation and fall detections are nowadays possible based on wearable WiFi tags [5].
4. Accelerometer-based wearable devices: Accelerometers are devices that measure the human body acceleration along certain axes. 3D digital accelerometers are widely used in e-health monitoring [21]. Positioning estimation via an accelerometer typically requires few additional sensors, such as gyroscopes (measuring the direction change), barometers (measuring the height change, in 3D positioning) or even camera-based sensors [30].
5. UWB tags: Ultra Wide Band (UWB) technology is based on sending short time pulses over a very wide bandwidth. This allows to achieve centimetre-level accuracy in positioning. UWB is still a rather expensive technology and wearable solutions are still rather scarce [17, 38].
6. Assisted-Global Positioning System (GPS) wearable receiver: the availability and accuracy of the Global Positioning System (GPS) indoors is quite limited. However, with assistance from the network and high-sensitivity receivers, and Assisted-GPS [(Assisted Global Positioning System Assisted Global Positioning System (A-GPS))] solution can also be used indoors to some extent (e.g., in areas with many windows or thin or wooden walls). Indoor localisation solutions, however, are likely to rely on some hybridisation of GPS or A-GPS with other sensors [3].
7. Doppler radar: Radar systems can be used to detect the distance and displacement. Together with digital gyroscopes, they can be used to compute human trajectories indoors [37] or for fall detection of elderly persons [25].
8. Light Emitting Diodes (Light Emitting Diodes (LED)) tag: with the widespread use of Light-Emitting Diodes (LED) for lighting, the visible light communications have also started to be used in the context of positioning [46]. Indeed, a commercial LED-based positioning solutions has been deployed by Philipps in 2015 in a Carrefour supermarket in Lille, France. In order to use this solutions, the user has to install an application on his or her mobile device, that is used to identify the light patterns of the fixed LED light bulbs on the ceiling of the supermarket.

14.3.2 Device-Free Technologies

In a device-free localisation technology, the user does not have to carry any device with her or him. Some fixed inter-connected receivers forming a wireless network together with a central controller unit are able to localise or sense the user based on ray reflections or changes in signal propagation due to the presence of the human body. Typically, such technologies offer a lower accuracy than their wearable counterpart, since most of them—with the exception of vision systems–are based on the human body influences on the signal strength fluctuations between the in-house tags and in-house receivers, when the person crosses the wave path. Some of them, e.g., vision systems, are quite privacy invasive, and thus they have a lower acceptability by the end users than the privacy-preserving techniques.

1. RFID in-house systems: The same principles as for the RFID tags apply, but this time the tags are to be scattered all through the house, instead of carried on by a person. The localisation is based on the observation that the presence of the human body changes significantly the way the signal propagates. Therefore, an RFID reader can detect the human presence and could, in principle, estimate the human movements [19, 33].
2. BLE in-house systems: The technology is the same as with the case of the BLE tags apply, with the main difference that the tags are on fixed places inside the house (not carried on). The fall detection and indoor localisation principles are the same as for RFID tags.
3. WiFi in-house systems: Similarly with RFID and BLE in-house systems, the WiFi in-house systems are contact-less systems, where the Access Point and the WiFi tags or transmitters are installed externally to the human wearable fabrics or pockets [47]. A device-free WiFi fall detection scheme with average precision of 94% detection probability at a 15% false alarm rate has been recently shown in [43].
4. UWB in-house systems: UWB in-house systems differ from UWB tags in the fact that there are no wearable devices in the in-house UWB solutions. The fall detection and user position are based on the time of arrival of multipath reflections due to presence of the human body [8]. Device-free UWB solutions are still scarce at the moment.
5. ZigBee in-house system: ZigBee signals can be used in a similar way with other types of indoor wireless signals. In [42], the authors proposed to use the shadowing effect created by the presence of user on wireless links in order to estimate the user location and to infer his or her activity.
6. Vision systems or camera-based systems: The vision systems are those systems requiring at least one surveillance camera in users' homes. The surveillance cameras capture continuously the images of the users and analyse their movement patterns and behavioural changes based on vision navigation and pattern matching techniques.

7. Tactile floors: Tactile floors or smart floors can also offer a device-free solution for user status monitoring indoors (e.g., user position or fall detection). The estimation accuracy depends on the density of the pressure sensing nodes, and the information can be sent to the central server through the WiFi network. Tactile floors are, however, too expensive and disruptive technologies for the use in elderly homes, and thus they are not included in our analysis.
8. Acoustic and ultrasound solutions: such solutions have been partially studied in the context of elderly automatic monitoring of activities and indoor localisation. There are rather few studies yet regarding the performance of such solutions for indoor positioning and fall detection. In a device-free ultrasonic localisation solution, some wall-mounted ultrasonic transducers capture the echoes that form the basis of the decision about a person's location [41].
9. Infrared localisation: in infrared communications, the thermal radiation of the human body can be exploited and no additional tag is required to perform the localisation [15]. The principle of infrared communications, briefly, is that every object with a temperature above the absolute zero emits energy via electromagnetic radiation. The human skin temperature leads to a maximum emission of radiation, as shown, for example, in [15]. A thermal detector, such as a pyroelectric sensor, a microbolometer array or a thermopile, converts the received infrared radiation into heat and it is able to detect this radiation. The thermopiles are the least expensive thermal detectors. By placing such thermopiles at the edges of a room, they are able to detect the human presence and human movements.

14.3.3 Examples of Indoor Positioning Technologies

It is well known that people spend most of their times indoor. In the last decade, the focus has moved from the outdoor positioning towards the indoor positioning. Some of the most promising technologies for indoor tracking are discussed, for example, in [10, 26]. The main technologies are enumerated in Table 14.2, with cross-references for further reading.

14.4 Simplified Example of Applying AHP in the e-Health LBS Context

This section focuses on a simplified example of how AHP can be used in order to address the following question: among a set of selected positioning technologies, which is the best-fit-to-purpose positioning technology to be used in indoor e-health applications. For a better understanding of the AHP analysis, this example takes

Table 14.2 Summary of the positioning technologies for indoor localisation and tracking, movement detection and fall detection

Wearable technologies	*References for further readings*
Wearable BLE tag + indoor receivers	[9, 12]
Wearable RFID tag + indoor readers	[12, 31]
Wearable WiFi tag + indoor receivers	[5, 12]
Wearable UWB tag + indoor receivers	[38]
Wearable accelerometer tag + indoor receivers	[21, 30]
A-GPS wearable receiver	[3, 4]
Doppler radar	[25, 37]
Device-free technologies	*References for further readings*
Indoor BLE system (user is device free)	[12]
Indoor RFID system	[12, 19, 33]
Indoor WiFi system	[12, 43, 44, 47]
Indoor UWB system	[8]
Indoor ZigBee system	[42]
Vision-based system/video cameras	[30]
Tactile or smart floor	
Acoustic and ultrasound solutions	[41]
Infrared localisation	[15]

into account only four of the technologies shown in Table 14.2 and it considers only three decision criteria. More comprehensive and realistic examples can be found in [5, 6, 26].

Let's consider that a designer would like to design an indoor Location Based Service for elderly people in order to detect falls and unusual patterns in the elderly mobility. Let's assume also that the designer is interested in three main constraints: the cost of implementing such a system, its performance and the end-user acceptability of the technology. In this example, but is assumed also that, for various reasons, only three supporting technologies are available for the designer, namely: (1) wearable BLE tags, (2) video camera based solutions or (3) in-house device-free WiFi solution.

In a first step, the designer can build the pair-wise comparison matrix of the three considered design criteria. Of course, the diagonal elements of the pair-wise comparison matrix are equal to 1 [see Eq. (14.1)]. For the rest of the elements of the level-one pair-wise comparison matrix, the designer uses a mixture of subjective and objective information, collected from surveys and literature searchers and draws the conclusion that the Cost criterion is four times more important than the performance and twice less important than the acceptability of the technology by the end-user. He or she also assumes that the acceptability criterion is seven times more important than the performance. With these values, the level-1 pair-wise comparison matrix can be built, as shown in Table 14.3.

Table 14.3 Level-1 pair-wise comparison matrix for the considered example

	Cost	Performance	Acceptability
Cost	1	4	1/2
Performance	1/4	1	1/7
Acceptability	2	7	1

Table 14.4 Level-2 pair-wise comparison matrix (cost efficiency criterion)

	Wearable BLE	Vision	In-house WiFi
Wearable BLE	1	5.87	4.73
Vision	0.1704	1	0.813
In-house WiFi	0.2114	1.23	1

Table 14.5 Level-2 pair-wise comparison matrix (performance criterion)

	Wearable BLE	Vision	In-house WiFi
Wearable BLE	1	0.08	3
Vision	12.5	1	20
In-house WiFi	0.33	0.05	1

In order to check if the data 'makes sense' or is logical, Saaty, the inventor of the AHP concept [34], introduced the idea of a consistency check based on the eigenvalues of the pair-wise comparison matrix. Details on computing the consistency ratio can be found, for example, in [34, 45]. The consistency ratio of Table 14.3 matrix is 0.2%. Typically, a consistency ratio below 10% shows a good consistency of the data.

For the level-2 pair-wise comparison matrices, we will have three of them, each corresponding to one of the design criterion. The Cost level-2 pair-wise comparison matrix is shown in Table 14.4, the performance level-2 pair-wise comparison matrix is shown in Table 14.5, and the acceptability level-2 pair-wise comparison matrix is shown in Table 14.6. All the weights in Table 14.6 are chosen based on 3 the literature results regarding the properties of different positioning technologies with respect to their performance, on users' surveys (e.g., as reported in [26]) on acceptability level of a certain technology, and on off-the-shelf commercial solutions currently available on the market for the four considered technologies in here.

In Table 14.4, a higher value means that a technology is more cost efficient with respect to another one.

In Table 14.5, the performance criterion refers to how good fall detection probabilities (at low false alarm rates) and how good indoor positioning accuracy can be achieved with one of the considered technologies.

In Table 14.6, a higher value means a more acceptable technology to have inside home from the user point of view. Users typically want their privacy preserved inside their homes and also they prefer non-disruptive technologies.

With the data from Tables 14.3, 14.4, 14.5 and 14.6, the priority vectors $t_i, i = 1, 2, 3$ can be computed as explained in Sect. 14.2 and their values are 53.9, 17.7 and 28.4. This means that the most suitable technology (among the considered ones

Table 14.6 Level-2 pair-wise comparison matrix (acceptability criterion)

	Wearable BLE	Vision	In-house WiFi
Wearable BLE	1	4	1.5
Vision	0.25	1	0.25
In-house WiFi	0.67	4	1

and according to the considered design criteria in this example) is a solution based on wearable BLE tags. The least suitable technology in this example turns out to be a vision-based solution, no doubt due to the low levels of acceptability associated with it. Such low levels of acceptability are due to the fact that most users want to preserve their home privacy and are uncomfortable with intrusive technologies such as video cameras.

14.5 Multi-Level and Joint Decisions

One of the advantages of using AHP is the possibility of comparing multi-level alternatives or factors. An AHP hierarchy can consist of an overall goal, a group of options or alternatives, and a group of factors or criteria, which can be broken down into subcriteria, sub-subcriteria. The AHP hierarchy can be used in many levels of problem [16, 23]. The use of an AHP hierarchy can depend on the nature of the problem and also on the knowledge, judgments and opinions of the participants in the decision-making process. Some factors and criteria can be broken down based on the opinion of one expert while another may group them into only one criterion, if it is considered less important. The inconsistency becomes even more complicated issue if there is an AHP hierarchy. The pair-wise comparison process can be repeated internally before weighting the criterion on the higher level.

In addition, multiple agents, such as people and companies with interests, are involved in joint decision making. Identities, situation, mutual trust and power interrelations affect the joint decision making. Relationships in the joint decision process evolve. The sense of identity, responsibility, cognition and feelings change during the process. The whole process can be described as an occasion for decision making. A model for targeting marketing can be found, for example, in [13]. Depending on the market and the target group, the advertiser can estimate whether it is preferable to target all of the individuals, subgroup or just one key person within the group. Modeling helps in assessing the effectiveness of marketing. Moreover, target group heterogeneity can be estimated.

Variable probabilities resulting in a decision or another are rarely fully independent in joint decisions when concerning people. This of course depends on the random variable that is being discussed upon. Sophisticated Bayesian network approaches can be in dealing with this issue of conditional dependency between decision makers. An example of such an approach can be found in Hartmann and Yildiz's approach [13], focusing on investigating two agents in decision-making

process. Another model on cooperative or collaborative marketing using AHP can be found in [36]. In a cooperative marketing, two or more units or enterprises join their forces to create a long-term alliance built upon their complementary of resources.

This chapter does not consider the multi-level criteria or alternatives, neither the joint decision making; however, if the combinations of positioning technologies are being studied, then an AHP hierarchy can be a very good solution as there are multi-level alternatives.

14.6 Conclusions

This chapter presented the steps involved in an Analytic Hierarchy Process and applied them to an example taken from Location Based Services. An overview of the main supporting technologies of LBS nowadays has been given and the main questions to be addressed by a designer of LBS have been emphasised. The purpose of this chapter was not to answer all the LBS design challenges, but rather to present a useful tool, namely the AHP, which can be employed to address some of these challenges. It has been shown that AHP is a flexible and useful tool which can be used in many contexts, because the final decision or ranking is obtained on the basis of the pair-wise relative evaluations of both the criteria and the options provided by the designer. An example from e-health LBS was selected to illustrate the applicability of AHP, motivated by the fact that this domain is highly relevant to the society nowadays. AHP is only one implementation of an MCDM process. Extensions of AHP are, for example, the Analytic Network Processes (ANP) or fuzzy set theories.

Acknowledgements This work was supported by the following projects: Academy of Finland (projects 250266 and 283076), EU AAL NITICS, Mobile@Old, PN-II-PT-PCCA-2013-4-2241 No 315/2014 and EU FP7 Marie Curie Initial Training Network MULTI-POS (grant nr. 316528).

References

1. S. Aflaki et al., Evaluation of incentives for body area network-based health-care systems, in *2013 IEEE Eighth International Conference on Intelligent Sensors, Sensor Networks and Information Processing* (2013), pp. 515–520. doi:10.1109/ISSNIP.2013.6529843
2. F. Baccelli, J. Bolot, Modeling the economic value of the location data of mobile users, in *2011 Proceedings IEEE INFOCOM* (2011), pp. 1467–1475. doi:10.1109/INFCOM.2011.5934934
3. Y. Bai et al., Designing a wearable computer for lifestyle evaluation, in *2012 38th Annual Northeast Bioengineering Conference (NEBEC)* (2012), pp. 93–94. doi:10.1109/NEBC.2012.6206978
4. K. Bakhru, A seamless tracking solution for indoor and outdoor position location, in *2005 IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications*, vol. 3 (2005), pp. 2029–2033. doi:10.1109/PIMRC.2005.1651796

5. A. Basiri et al., Indoor positioning technology assessment using analytic hierarchy process for pedestrian navigation services, in *2015 International Conference on Localization and GNSS (ICL-GNSS)* (2015), pp. 1–6. doi:10.1109/ICL-GNSS.2015.7217157
6. A. Basiri et al., Overview of positioning technologies from fitness-to-purpose point of view, in *2014 International Conference on Localization and GNSS (ICL-GNSS)* (2015), pp. 1–7
7. P. Bellavista, A. Kpper, S. Helal, Location-based services: back to the future. IEEE Pervasive Comput. **7**(2), 85–89 (2008). ISSN:1536-1268. doi:10.1109/MPRV.2008.34
8. C. Chang, A. Sahai, Object tracking in a 2D UWB sensor network, in *Conference Record of the Thirty-Eighth Asilomar Conference on Signals, Systems and Computers, 2004*, vol. 1 (2004), pp. 1252–1256. doi:10.1109/ACSSC.2004.1399342
9. Z. Chen, H. Hu, J. Yu, Privacy-preserving large-scale location monitoring using bluetooth low energy, in *2015 11th International Conference on Mobile Ad-hoc and Sensor Networks (MSN)* (2015), pp. 69–78. doi:10.1109/MSN.2015.38
10. D. Dardari, P. Closas, P.M. Djuri, Indoor tracking: theory methods, and technologies. IEEE Trans. Veh. Technol. **64**(4), 1263–1278 (2015). ISSN:0018-9545. doi:10.1109/TVT.2015.2403868
11. J. Figuera, S. Greco, M. Ehrgott (eds.), *Multiple Criteria Decision Analysis, State of the Art Surveys* (Springer, New York, 2005)
12. Y. Gu, A. Lo, I. Niemegeers, A survey of indoor positioning systems for wireless personal networks. IEEE Commun. Surv. Tutorials **11**(1), 13–32 (2009). ISSN:1553-877X
13. W.R. Hartmann, V.T. Yildiz, Marketing science conference, in *A Structural Analysis of Joint Decision-Making* (2006), 38 pp.
14. M. Hasani et al., Hybrid WLAN-RFID indoor localization solution utilizing textile tag. IEEE Antennas Wirel. Propag. Lett. **14**, 1358–1361 (2015). ISSN:1536-1225. doi:10.1109/LAWP.2015.2406951
15. D. Hauschildt, N. Kirchhof, Advances in thermal infrared localization: challenges and solutions, in *2010 International Conference on Indoor Positioning and Indoor Navigation (IPIN)* (2010), pp. 1–8. doi:10.1109/IPIN.2010.5647415
16. G. Hu, W. Li, Y. Li, Evaluation of teacher's performance in independent colleges based on AHP and multi-level matter element extension measurement models, in *2010 International Conference on Computational Intelligence and Software Engineering (CiSE)* (2010), pp. 1–4. doi:10.1109/CISE.2010.5676753
17. Z. Irahhauten, H. Nikookar, M. Klepper, A joint ToA/DoA technique for 2D/3D UWB localization in indoor multipath environment, in *2012 IEEE International Conference on Communications (ICC)* (2012), pp. 4499–4503. doi:10.1109/ICC.2012.6364603
18. M. Jalobeanu et al., Reliable kinect-based navigation in large indoor environments, in *2015 IEEE International Conference on Robotics and Automation (ICRA)* (2015), pp. 495–502. doi:10.1109/ICRA.2015.7139225
19. A.R. Jimnez, F. Seco, Combining RSS-based trilateration methods with radio-tomographic imaging: exploring the capabilities of long-range RFID systems, in *2015 International Conference on Indoor Positioning and Indoor Navigation (IPIN)* (2015), pp. 1–10. doi:10.1109/IPIN.2015.7346937
20. T. Kagawa, H.B. Li, R. Miura, A UWB navigation system aided by sensor-based autonomous algorithm - deployment and experiment in shopping mall, in *2014 International Symposium on Wireless Personal Multimedia Communications (WPMC)* (2014), pp. 613–617. doi:10.1109/WPMC.2014.7014890
21. D.M. Karantonis et al., Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring. IEEE Trans. Inf. Technol. Biomed. **10**(1), 156–167 (2006). ISSN:1089-7771. doi:10.1109/TITB.2005.856864
22. R. Khorramshahgol, H. Azani, Y. Gousty, An integrated approach to project evaluation and selection. IEEE Trans. Eng. Manag. **35**(4), 265–270 (1988). ISSN:0018-9391. doi:10l.1109/17.7449

23. Y. Kuang, M. Hu, Q. Wu, Multi-level evaluation model for intangible cultural heritage status based on fuzzy set theory, in *2015 8th International Symposium on Computational Intelligence and Design (ISCID)*, vol. 2 (2015), pp. 572–576. doi:10.1109/ISCID.2015.170
24. J.S. Li, X.L. Miao, Research on mode selection for developing port logistics in China based on analytical hierarchy process; Taking Tianjin port logistics as an example, in *2010 IEEE 17th International Conference on Industrial Engineering and Engineering Management (IE EM)* (2010), pp. 1383–1387. doi:10.1109/ICIEEM.2010.5646005
25. L. Liu et al., Doppler radar sensor positioning in a fall detection system, in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (2012), pp. 256–259. doi:10.1109/EMBC.2012.6345918
26. E.S. Lohan et al., Analytic hierarchy process for assessing e-health technologies for elderly indoor mobility analysis, in *5th EAI/ACM International Conference on Wireless Mobile Communication and Healthcare - Transforming Healthcare Through Innovations in Mobile and Wireless Technologies, Mobihealth 2015* (2015). doi:978-1-63190-088-4
27. M. Mantash et al., Dual-band textile hexagonal artificial magnetic conductor for WiFi wearable applications, in *2012 6th European Conference on Antennas and Propagation (EUCAP)* (2012), pp. 1395–1398. doi:10.1109/EuCAP.2012.6206238
28. P.K. M'Pherson, A framework for systems engineering design. Radio Electron. Eng. **51**(2), 59–93 (1981). ISSN:0033-7722. doi:10.1049/ree.1981.0010
29. S. Nakao et al., UHF RFID mobile reader for passive- and active-tag communication, in *2011 IEEE Radio and Wireless Symposium* (2011), pp. 311–314. doi:10.1109/RWS.2011.5725441
30. K. Ozcan, S. Velipasalar, Wearable camera- and accelerometer-based fall detection on portable devices. IEEE Embed. Syst. Lett. **8**(1), 6–9 (2016). ISSN:1943-0663. doi:10.1109/LES2015.2487241
31. D.C. Ranasinghe et al., Towards falls prevention: a wearable wireless and battery-less sensing and automatic identification tag for real time monitoring of human movements, in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (2012), pp. 6402–6405. doi:10.1109/EMBC.2012.6347459
32. S.F.A. Razak et al., Interactive android-based indoor parking lot vehicle locator using QR-code, in *2015 IEEE Student Conference on Research and Development (SCOReD)* (2015), pp. 261–265. doi:10.1109/SCORED.2015.7449337
33. W. Ruan, Unobtrusive human localization and activity recognition for supporting independent living of the elderly, in *2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)* (2016), pp. 1–3. doi:10.1109/PERCOMW.2016.7457085
34. T.L. Saaty, *Decision Making for Leaders* (Lifetime Learning Publications, Belmont, CA, 1982)
35. G.E. Slutsker, Decision making in vascular surgery, in *Intelligent Systems and Semiotics (ISAS), Proceedings of Intelligent Control (ISIC), 1998. Held Jointly with IEEE International Symposium on Computational Intelligence in Robotics and Automation (CIRA)* (1998), pp. 489–492. doi:10.1109/ISIC.1998.713710
36. X. Sun, The effect of online product reviews motivation on collaborative marketing, in *2012 IEEE Symposium on Robotics and Applications (ISRA)* (2012), pp. 525–528. doi:10.1109/ISRA.2012.6219240
37. Y. Tang, C. Li, Wearable indoor position tracking using onboard K-band Doppler radar and digital gyroscope, in *2015 IEEE MTT-S 2015 International Microwave Workshop Series on RF and Wireless Technologies for Biomedical and Healthcare Applications (IMWS-BIO)* (2015), pp. 76–77. doi:10.1109/IMWS-BIO.2015.7303785
38. X.F. Teng et al., Wearable medical systems for p-health. IEEE Rev. Biomed. Eng. **1**, 62–74 (2008). ISSN:1937-3333. doi:10.1109/RBME.2008.2008248
39. R.C. Trundle, Paradoxes of human nature. Etica & Politica/Ethics & Politics IX 1 (2007), pp. 181–186
40. M.H. Vahidnia et al., Fuzzy analytical hierarchy process in GIS application, in *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* (2008), pp. 593–596

41. E.A. Wan, A.S. Paul, A tag-free solution to unobtrusive indoor tracking using wall-mounted ultrasonic transducers, in *2010 International Conference on Indoor Positioning and Indoor Navigation (IPIN)* (2010), pp. 1–10. doi:10.1109/IPIN.2010.5648178
42. J. Wang et al., Device-free simultaneous wireless localization and activity recognition with wavelet feature. IEEE Trans. Veh. Technol. **PP**(99), 1 (2016). ISSN:0018-9545. doi:10.1109/TVT.2016.2555986
43. Y. Wang, K. Wu, L.M. Ni, WiFall: device-free fall detection by wire-less networks. IEEE Trans. Mob. Comput. **PP**(99), 1 (2016). ISSN:1536-1233. doi:10.1109/TMC.2016.2557792
44. C. Wu et al., Non-invasive detection of moving and stationary human With WiFi. IEEE J. Sel. Areas Commun. **33**(11), 2329–2342 (2015). ISSN:0733-8716. doi:10.1109/JSAC.2015.2430294
45. W.J. Xu, Y.C. Dong, W.L. Xiao, Is it reasonable for Saaty's consistency test in the pairwise comparison method? in *2008 ISECS International Colloquium on Computing Communication, Control, and Management*, vol. 3 (2008), pp. 294–298. doi:10.1109/CCCM.2008.136
46. W. Xu et al., Indoor positioning for multiphotodiode device using visible-light communications. IEEE Photonics J. **8**(1), 1–11 (2016). ISSN:1943-0655. doi:10.1109/JPHOT.2015.2513198
47. B. Zhou, N. Kim, Y. Kim, A passive indoor tracking scheme with geometrical formulation. IEEE Antennas Wirel. Propag. Lett. **99**, 1 (2016). ISSN:1536-1225. doi:10.1109/LAWP.2016.2537842

Chapter 15
DroneAlert: Autonomous Drones for Emergency Response

Luis Bausá López, Niels van Manen, Erik van der Zee, and Steven Bos

15.1 Unmanned Air Vehicles

Unmanned Air Vehicles (UAVs), a.k.a. drones, are causing a revolution as an emerging technology with many different fields of application. Although initially UAVs were mostly used for military applications, they have been recently adopted by the leisure sector, thus becoming cheaper and more feature-rich. This has opened the doors for researchers and businesses to find applications of this technology to a variety of purposes [16]. One of the most important fields of application is Crisis Management (CM) and Emergency Response [11], because an autonomous flying drone can reach places faster [21] than a human asset, and it can also reach to places where access by land is deemed impossible at the current circumstances. Also, the possibility to use UAVs that can fly autonomously reduces the risk of accidents caused by an error of the human pilot.

15.1.1 Types of UAV

There is no international standard classification for UAVs. Still, governments and users have developed several UAV classifications by functional purpose, weight, structure and range, among others. Some of the most important ones are detailed next.

L.B. López • N. van Manen (✉)
Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081HV Amsterdam, The Netherlands
e-mail: luis.bausa@uji.es; n.van.manen@vu.nl

E. van der Zee • S. Bos
Geodan, President Kennedylaan 1, 1079 MB Amsterdam, The Netherlands
e-mail: erik.van.der.zee@geodan.nl; steven.bos@geodan.nl

J. Nurmi et al. (eds.), *Multi-Technology Positioning*,
DOI 10.1007/978-3-319-50427-8_15

According to the functional purpose, UAVs can be classified in the following categories:

- Military: designed for military applications, such as battlefield reconnaissance, logistics, combat or training operations.
- Logistics: designed for logistic operations.
- Research and development: used for the development and testing of UAV technologies.
- Civil and commercial: designed for leisure and commercial purposes. The drones that can be used for the use case presented in this document fall into this category.

According to the structure, UAVs can be classified into two major groups:

- Fixed-wing: airplane-like UAVs (Fig. 15.1b).
- Rotorcraft: copter-like UAVs, also called multicopter UAVs when they have more than two rotors. They can receive a specific name depending on the number of rotors, e.g. a quadcopter (Fig. 15.1a) has four rotors.

UAVs can be categorised, according to their weight, in:

- Micro unmanned Air Vehicle (MAV): weight from a few grams up to 2 kg [4].
- Miniature or small UAV: weight between 2 and 25 kg (Fig. 15.1a).
- Medium UAV: weight between 25 and 200 kg.
- Heavy UAV: weight above 200 kg (Fig. 15.1c).

The type of UAV that best suits the purpose of the system presented by the authors of this document is a rotorcraft UAV of small size. The main reasons behind choosing rotorcraft instead of fixed-wing UAVs are two. The first reason is that fixed-wing UAVs of small size and weight require a person to throw them for the take off and also someone to catch them during the landing, but rotorcraft UAVs can take off and land without human assistance. The second reason is that rotorcraft UAVs require a smaller area for the landing and takeoff than fixed-wing UAVs. Moreover, rotorcraft UAVs can also stay still in the air maintaining their position, which can be of great advantage in case it is necessary to obtain imagery [4] of the area for a sustained period or from a closer distance. However, rotorcraft UAVs have one great disadvantage compared to fixed-wing UAVs: they can achieve lesser speed and have less autonomy, which means they can cover smaller distances [7].

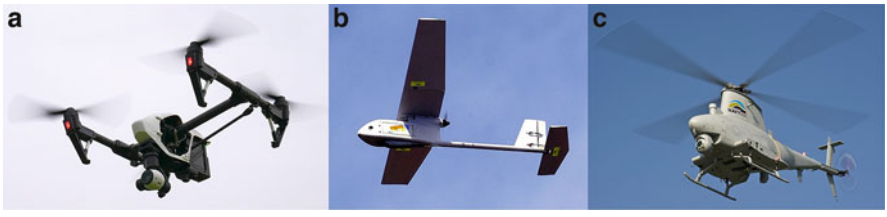

Fig. 15.1 Types of UAV. From *left* to *right*: (**a**) Quadcopter, (**b**) Fixed-wing and (**c**) Biofuel copter

The reasons to choose UAVs of small size are two: first, the regulations for UAVs weighting less than 25 kg at the moment of the take-off are less strict and provide more freedom of use; and second, small aircrafts are more versatile, thus more useful in complex environments such as urban environments, parks or forests.

15.1.2 Laws and Regulations for Unmanned Air Vehicles

Each country is responsible for developing their own laws and regulations regarding UAVs. The laws and regulations concerning the use of UAVs for civil and commercial purposes are, in the best of cases, at an early stage, and in some countries nonexistent. The USA currently has the most comprehensive regulations regarding the use of UAVs. The leading organisation in charge of the control and regulation of unmanned aircrafts in the USA is the Federal Aviation Agency (FAA). The European Aviation Safety Agency (EASA) started in 2015 to gather information to develop a common normative to regulate the usage of UAVs, and aims to develop common laws in 2016 and 2017. In the meantime, each EU country is responsible for defining their own laws. In Spain the agency in charge is the Spanish Air Safety Agency (AESA), in Finland it is the Finnish Transport Safety Agency (Trafi) and in the Netherlands it is the Human Environment and Transport Inspectorate (ITL). A comprehensive list of the laws and regulations of UAVs in each country can be found at http://uavcoach.com/drone-laws/. Further, currently there are still no regulations in any country regarding the civil and commercial use of autonomous UAVs.

15.1.2.1 Privacy Considerations

The FAA has been working for some time on the regulations of Unmanned Aerial System (UAS) [6]. While most of the current laws and regulations are concerning the restrictions of flight and usage of UAVs, an important aspect that is currently under consideration are the privacy issues of using UAVs with mounted cameras in public spaces. For this reason, the FAA has developed the regulations of flying UAVs, differentiating between the usage for recreational purposes, for business purposes and for public entities. There is also an initiative, Know Before You Fly [14], that aims to inform the users of drones of the current procedures and regulations.

15.1.2.2 Airspace Flight Restrictions

One important regulation of UAVs is the establishment of UAV No-Fly Zone (NFZ) and UAV High Restriction Fly Zone (HRZ) around the airports (Fig. 15.2). Airports have been classified into two categories according to the size and importance of

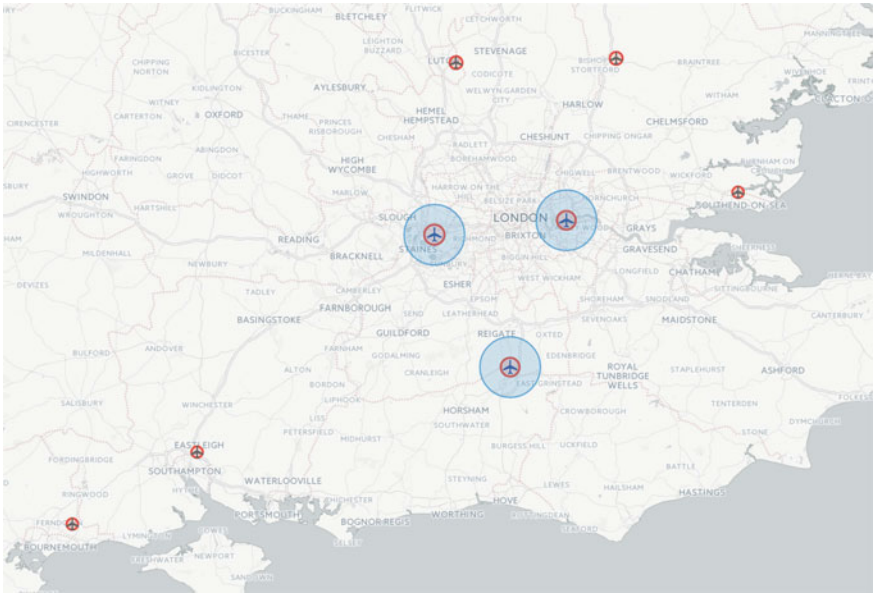

Fig. 15.2 Example of UAV No-Fly Zones and UAV High Restriction Fly Zones around airports. The image shows three Category A Airports and Category B airports around London, UK

the airport. Category A airports have an NFZ of 2500 m and an HRZ of 8000 m. Category B airports have an NFZ of 1600 m. The UAV No-Fly Zone, as its name indicates, is an area where UAVs cannot fly, whereas the high restriction fly zone is an area where UAVs can only fly with a permission and always below 100 m. Some drone providers have already developed features in their drones so they will not fly in those areas [3]. The FAA has also led an initiative for the safe and responsible use of drones, and has designed logos to identify no drone zones [5]. Finally, there are whole cities and areas that have been declared as no drone zones, such as the city of Washington, DC, USA.

Cross-border international regulations are another important aspect to be considered. In most cases there are regulations regarding the flight restrictions of UAVs establishing that they cannot cross country borders. This regulation was established for multiple reasons, such as avoiding the usage of UAVs for criminal purposes like drug smuggling. Also, this regulation is meant to avoid problems originating from the differences between the laws and regulations of each country.

15.1.2.3 Flying License and Aircraft Registration

Many countries have established regulations regarding the piloting of UAVs. In most cases it is necessary to acquire a special license, expedited by a national organisation, in order to be able to fly unmanned aircrafts. In the USA, this license is

called Certificate Of Authorization (COA) and can be obtained from the FAA or an equivalent accreditation agency. FAA regulations also specify that the UAVs have to be registered before they can fly.

15.1.3 Current and Future Fields of Application

UAVs are being used for a wide variety of applications in many different fields. Some examples are:

UAVs with LiDAR cameras are being used for mapping, to perform 3D spatial models of buildings, structures and geographic areas [1], to perform surveying measurement tasks, to count and classify crops and trees [20], etc. Currently there are companies offering these services on demand, and there are even brokering services to find drone companies, pilots and operators [2].

UAVs with video cameras are used to acquire video footage of public events such as music concerts or sport competitions, to perform surveillance tasks such as the police [15] for following suspects or the lifeguards to watch the beaches, for recreational purposes such as race competitions.

There are several initiatives to use UAVs for transportation and logistic. One well-known example is the intention of Amazon to use drones in the future to deliver packages. There are also initiatives for using drones for humanitarian and medical tasks, such as delivering medicines to villages that get isolated during certain seasons [9] or delivering a defibrillator [23].

15.1.4 Components

The specifications of the UAV components can vary considerably depending on the type of UAV. The following components can be found in most common UAVs [13]:

1. Body:
   The body is the main structure of the UAV. It is composed by the frame and the wings (for planes) or the arms and rotors (for copters).
2. Power supply:
   Small UAVs are usually powered by a lithium battery, but bigger UAVs can be powered by fuel or solar power.
3. Computing
   Depending on the functional purpose and size of the UAV, the computational capabilities can vary considerably. Bigger UAVs have high computational capabilities, medium UAVs have reduced computing power and small UAVs have embedded systems based on micro-controllers. The main functions of the computing part are flight control and communications. The flight control

tasks are supported by actuators, such as stabilisers and speed controllers. The actuators are in charge of the various control tasks, such as controlling the RPMs of each rotor, the inclination of the aircraft and the speed, and executing predefined automatic action based on the input.

4. Sensors
   UAVs can be equipped with a wide variety of sensors. The most common sensors that can be found on UAVs, categorised by purpose, are:
   a. Communication: radio wave antenna/s of different frequency, such as Wi-Fi and 3G.
   b. Positioning and guidance: Digital compass, positioning sensor (GPS/GNSS/Glonass, etc.), inertial sensor and ultrasound sensor.
   c. Observation and measurement: video camera, LiDAR sensor, microphone, speaker, barometer, hygrometer, magnetic sensor, proximity sensor, IR sensor, etc.

5. Software
   The simplest UAVs do not make use of any external software whatsoever, and are piloted from a remote controller. Advanced UAVs are controlled using a UAV Ground Control System (GCS). In the simplest cases, this system only provides tools to pilot the drone directly [8]. More advanced UAV Ground Control Systems can provide extended functionality [12], such as: access to the telemetry of the sensors; real-time streaming of the telemetry, such as video streaming [4]; recording sensor outputs; creation and loading of flight routes; creation and execution of macros defining complex actions, e.g. perform a loop or take a series of pictures while flying around a specific position; post-processing of the sensor telemetry, such as creating a 3D model of the area or mapping the flight path; performing on-flight corrections to the predefined flight route, etc.

   Although in some cases the extended functionality is embedded into the GCS, strictly speaking many of those tasks are not within the responsibilities of the GCS, and thus not part of it. In these cases, additional software can be used to extend the functionality of the GCS [8]. The additional software can be classified into two types: the first type is composed by those applications that work tightly with the GCS as "plugins", providing extra functionality that can be directly accessed from the interface of the GCS; the second type are those applications that work separately from the GCS, providing a separate interface to access the new functionality. For example, a flight route planner can be embedded into the GCS interface, and thus classified as type 1, or it can be an external application, and thus classified as type 2. On the other hand, an application to upload a flight route to the UAV will usually be of the first type.

   The GCS is installed on a separate computing device, i.e. a PC, a laptop, a smartphone, etc. Additional software of the first type has to be installed on the same device, while the second type of software can be installed on a separate device. In some cases, the GCS can be accessed remotely from another device, e.g. from a tablet via Internet using a web interface.

15.2 Emergency Response

There is a wide variety of emergency alerts that must be handled by the emergency services. The involved institutions and procedures regarding the response vary depending on the nature of the emergency alert [22]. In most cases, the police, firemen and medical services are involved, but other agencies might also get involved, such as the institutions in charge of the different infrastructures, i.e. gas, electricity, transportation (roads, rails), communications, water, harbours, public buildings, etc. and even the military forces.

15.2.1 Sources of Emergency Alerts

Emergency alerts may originate from several different sources, such as:

- Galileo Search And Rescue (SAR) service
- eCall service
- 112 emergency service
- Sensors: smoke detectors, temperature sensors, water level sensors, etc.
- Private security services: alarm systems, surveillance centres, private security officers.

Depending on the source of the emergency alert, among other things, the alert can contain a wide range of information about the cause and nature of the alert. A call to the 112 service might achieve a thorough context concerning the emergency, provided that the caller can give such information. In such cases, the alert is certain and must be responded to. On the other hand, an alert triggered by a smoke sensor provides information about the time, location and possible cause of the alert, a fire, but the alert must be confirmed to assess that there is a real threat [19], either by sending someone to check the area in person or by double-checking with other sensors, because it could happen that the sensor is defective and is malfunctioning, or that someone was smoking. Although it cannot be immediately assumed that there is a fire going on, emergency response regulations and procedures dictate that it must be responded to as such, and the emergency response can only be dismissed after a responsible person (i.e. a policeman, a private security officer, a fireman, etc.) has checked in person the place where the alert has originated and verified that it was a false alarm.

The type of emergency alerts that are of interest for this research are those originated by the Galileo SAR. These alerts only contain information regarding the time and location of the emergency alert. In such case, the response to the emergency alert is most challenging because, without more information, predicting the type of emergency situation is impossible, and thus it is very hard for the emergency responders to be adequately prepared for it.

15.2.2 Types of Emergency Situations

The emergency alerts can be classified into two categories depending on the current emergency situation:

1. As part of a bigger emergency situation: these alerts are caused by a major emergency situation affecting a large amount of population or with a wide impact area. For example, an alert triggered by a person trapped on a building that collapsed after an earthquake.
2. As a standalone emergency alert: emergency alerts that are not part of a major emergency situation. For example, an alert triggered by a person that is immobilised on a mountain due to a hiking accident.

In both cases, the system proposed here would be able to obtain real-time imagery of the location where the alert has originated within a fraction of the time that it would take the emergency responders to arrive to the location. The analysis of the imagery can produce more detailed information regarding the nature of the alert, making it possible for the emergency responders to prepare either before the departure or during the displacement to the alert location, thus saving valuable time.

15.3 The DroneAlert System

The proposed system aims to obtain live imagery of the location of an emergency alert using UAVs, and send that imagery to the emergency responders. The final purpose is to provide information regarding the nature of the alert to allow the emergency response team to prepare for the emergency before arriving to the location of the incident, thus saving valuable time that may be translated into saved lives and reduced damages and costs. An example of the use case is shown in Fig. 15.3.

The system is capable of providing aerial support to the emergency responders by acquiring live imagery of the location where the alert has originated. Capturing live imagery of the location of the incident allows the emergency services to analyse and prepare for the situation at an earlier stage. The information obtained from the early analysis of the situation can be used to determine the most suitable response procedure, the level of emergency and the resources needed to handle the situation (i.e. are the firemen required? And the police? Is medical assistance needed? How many ambulances are needed? Is the victim alone? Is there only one victim?, etc.). Further, the analysis of the imagery provides information about the location and its context, which can be used to determine how to access the incident location, the type of terrain, etc.

Future enhancements of the system will make use of UAVs with additional equipment to enhance the emergency support. Possible enhancements of the system are:

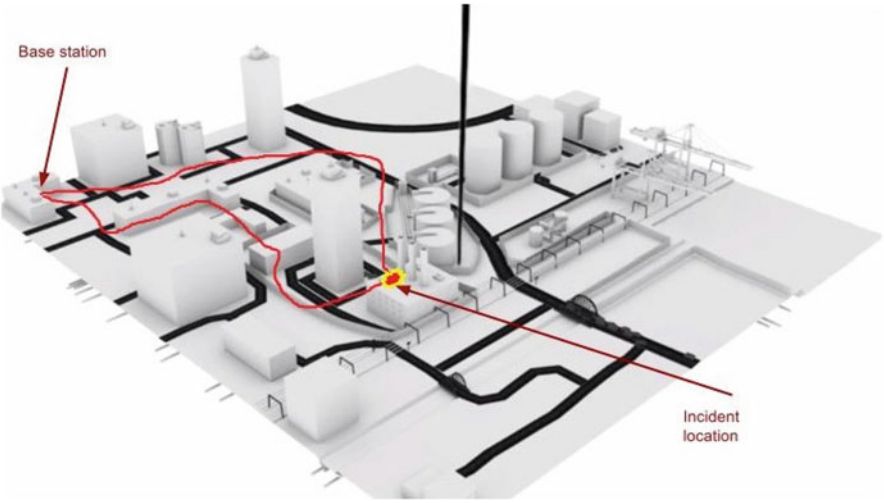

Fig. 15.3 Example of a DroneAlert autonomous drone flight operation

- Bidirectional communications: equip the drone with microphone and speakers to establish bidirectional communication with the person that originated the alert. This feature allows the emergency services to gather more details about the status of the alerter and to give him instructions until the emergency responders arrive to the location.
- First Aid kit: the drone can be equipped with a first aid kit to help the victim while the assistance arrives to the location, such as a blanket and analgesics. Using the bidirectional communications, the emergency services can give instructions to the victim.

15.3.1 Autonomous UAVs

An important aspect of the system is the autonomous [7] capabilities of the UAV. An autonomous drone is understood as a drone capable of flying without the constant direct input from a pilot. This autonomy can be achieved by two methods, or the combination of both. The first method is by loading into the UAV a flight route that the UAV can follow by itself using positioning and navigation technologies, e.g. gnss chips, motion sensor and ultrasounds. However this method has a flaw: if the flight route goes through a tree or a building, the UAV will crash into the obstacle. This is where the second method can be of use. The second method is concerned with the ability of the UAV to recognise obstacles and timely correct its flight route accordingly to the situation. While the first method relies on the accuracy of the geographical model used by the flight planner to calculate the routes, the second

method relies on the capabilities of the UAV to perform heavy computation and timely solve complex problems in certain situations. Obviously, the second method is more complex (harder to achieve) and expensive (requires more expensive UAVs). However, the autonomy of the UAVs is the key to the system. By being autonomous in their flight and takeoff, the process of deploying the aerial support is accelerated and the time to arrival is reduced considerably. Moreover, this autonomy allows one or few pilots to monitor multiple UAVs each, increasing the efficiency of the operations and the capacity of the system.

15.3.2 UAV Limitations

As an emerging technology, UAVs have several limitations that must be considered. The most important limitation is the autonomy of the battery, which in most cases is below 30 min for small UAVs powered with lithium batteries. The autonomy limitation poses a challenge because it directly affects the distance that can be covered with the drone. For example, Consider a UAV with an autonomy of 25 min. From those 25 min of autonomy, 1.5 min is reserved for the take off, 1.5 for the landing and 6 for capturing the live imagery. That leaves 16 min for the travel, 8 for the way there and 8 for the way back. If the UAV can fly at a speed of 10 m/s, then it can cover an area of maximum 4800 m radius from its base station. While this can be enough to cover a small town, in most cases it is insufficient to cover a large city. A solution to overcome this limitation is to set several base stations evenly distributed to cover the whole area. A spatial analysis can be used to determine the most suitable locations and the optimal number of base stations to cover the desired area. A similar solution was used by Pulver in [18], where the requirement was to be able to deploy the drone to arrive within 1 min.

Tightly related to the autonomy is the maximum weight of the UAV and its cargo, as the heavier it gets the higher the power consumption. This is a very limiting factor, because the heaviest component of the UAV is the power source (i.e. the battery), and a more powerful battery does not necessarily mean more autonomy due to the added cargo, hence the autonomy limiting ceiling of 30 min.

Another limitation of UAVs, mostly in small UAVs, are the capabilities to operate in different weather conditions. Due to their small size and weight, the wind can seriously affect the operation of UAVs. Also rain, snow and hail can cause impossible operating conditions. Fog should not be a problem when it comes to operating the drone, but it can be a problem for acquiring imagery. Extreme temperatures can also affect the operation of UAVs, being overheating one of the major concerns, as it may cause batteries to explode.

15.4 System Architecture

The process of providing aerial support using a UAV is summarised by the use case diagram (Fig. 15.4) and the sequence diagram (Fig. 15.5). The components of the system are detailed in Sect. 15.4.1, and the activities in the use case diagram are discussed in Sect. 15.4.2.

15.4.1 Components

The system has the following components, as shown in the sequence diagram (Fig. 15.5).

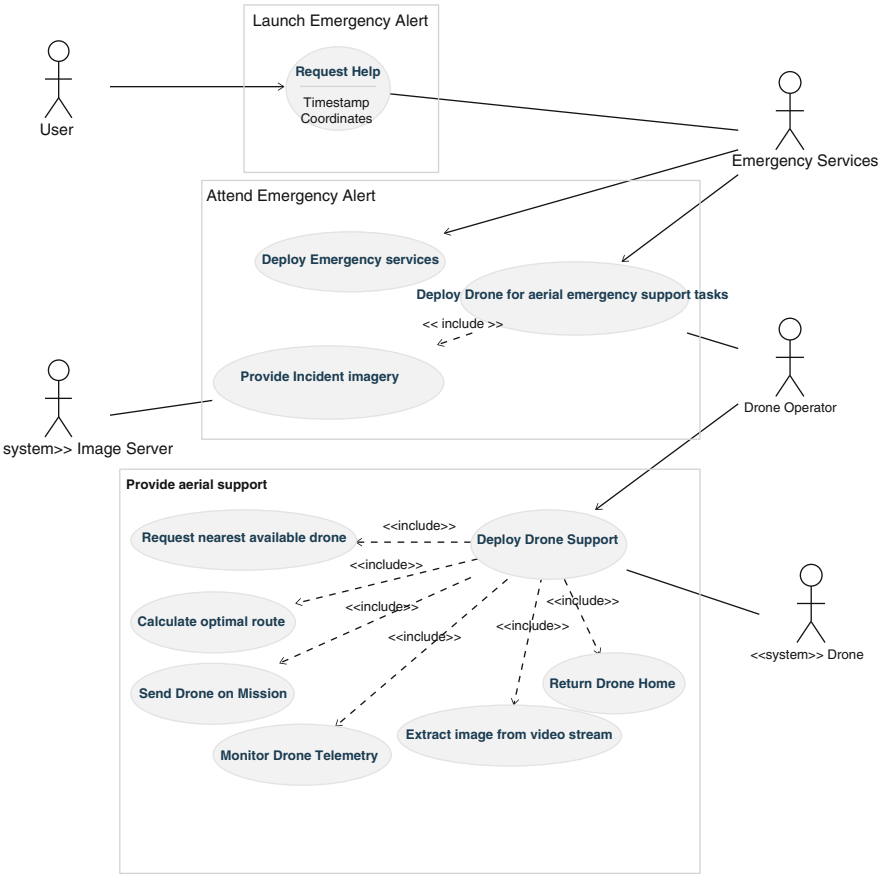

Fig. 15.4 Use Case diagram of the DroneAlert system

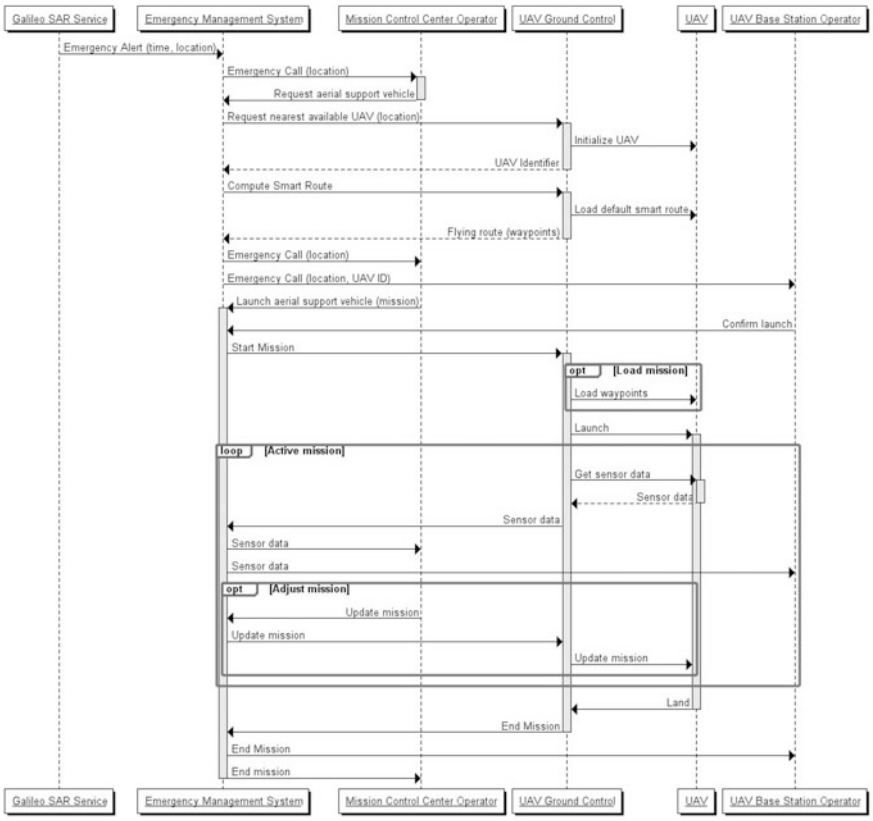

Fig. 15.5 Sequence diagram of the response to an emergency alert using the DroneAlert system

15.4.1.1 Emergency Management System

The Emergency Management System (EMS) is the system used by the emergency services, such as 112, that receives the emergency alert. The EMS is operated by an emergency services operator that is in charge of managing the emergency and deploy the adequate emergency resources and assets. If the operator decides that aerial support is needed, then the system sends the information of the alert to the UAV Mission Control Center (MCC).

15.4.1.2 UAV Ground Control System

The UAV Ground Control System (GCS) [8] is in charge of managing and controlling the UAVs, the flight routes, and acquiring the telemetry. In our case, the GCS contains three main modules:

1. The UAV Ground Control System, in charge of controlling the UAVs and acquire the telemetry.
2. The UAV management system, in charge of managing the UAVs, i.e. their location, status and capabilities.
3. The flight route planner, in charge of calculating the flight route and designing the flight plan for the missions.

15.4.1.3 UAV Base Station Operator

UAVs are located in base stations evenly distributed to cover as much area as possible. Each base station contains one or more UAVs. An operator is located in each base station to verify the proper functioning of the UAVs, set up the selected UAV for the mission and provide confirmation about the mission start and end (i.e. the UAV take off and landing).

Whenever a UAV starts a mission, the operator will provide a second confirmation that the UAV has successfully been deployed. Also, the operator will assist the UAV in the landing and give confirmation when the UAV has successfully returned to the base station after the mission has finished. The operator at the base station is also in charge of updating the status of the UAVs of his base station and of their maintenance (charging the batteries, calibrating the sensors, etc.).

15.4.1.4 UAV

The UAV is the protagonist of the system presented in this document. The UAV is initially located at a base station, whose coordinates will be the starting point of the mission. The UAV can only be selected for a mission if its status is set as available.

When it is selected for a mission, a smart route is loaded into the UAV. The UAV contains an autopilot module that uses the positioning sensors to drive the UAV through the waypoints defined by the flight route. Once the mission control center operator (Sect. 15.4.1.5) has authorised the mission, the UAV starts the mission. During the mission, the UAV will be constantly monitored by streaming its telemetry to the GCS. The operator can update the mission at any time by sending new commands to the UAV or by updating the flight route. When the UAV has fulfilled its goal, i.e. provided the requested aerial imagery, a new flight route is loaded into the UAV containing the route to return it to its corresponding base station. The Drone management system, part of the GCS, contains information about the status, location and capabilities of each UAV.

15.4.1.5 Mission Control Center Operator

The operator at the mission control center [10] receives and processes the request for aerial support and monitors the mission. This operator is the human link between the DroneAlert system and the emergency services operators that has received the emergency alert and has performed the request for aerial support.

15.4.2 Activities

To provide the desired functionality, the DroneAlert system carries out the activities presented in the use case diagram (Fig. 15.4). This section describes these activities, listed in the order in which they are performed.

15.4.2.1 Request Nearest Available UAV

The UAV management system, part of the GCS, contains the location, status and capabilities of all the UAVs. When a UAV is requested, the system will select the most suitable UAV from the list. The UAV selection criteria is based on three aspects:

1. Availability: only available UAVs can be selected for a mission.
2. Location: the system will give priority to those UAVs located in the nearest base station(s).
3. Capabilities: the UAV is required to have sufficient flight autonomy to reach the incident location from its base station. Also, special capabilities might be used to filter out UAVs, such as night vision (for poor light conditions) or waterproof operation (for rainy weather conditions).

The availability of UAVs is determined by its status. The status of UAVs can be one of the following:

1. Available
2. Unavailable
   (a) Reserved or waiting for orders
   (b) Charging battery
   (c) Broken, repairing or under maintenance
   (d) Battery discharged
3. Flying (on mission)
   (a) Flying to target
   (b) At target
   (c) Flying back to base station.

15.4.2.2 Calculate Optimal Flight Route

In order to calculate the flight route, the system takes into consideration the obstacles and the no-fly zones (Fig. 15.2). The obstacles must be avoided by considering their shape and height. Depending on the obstacle, the flight route will surround it or fly over it. An obstacle can be anything that is not considered ground, i.e. buildings, trees, streetlamps, bridges, statues, etc. No-fly zones must be avoided at all costs,

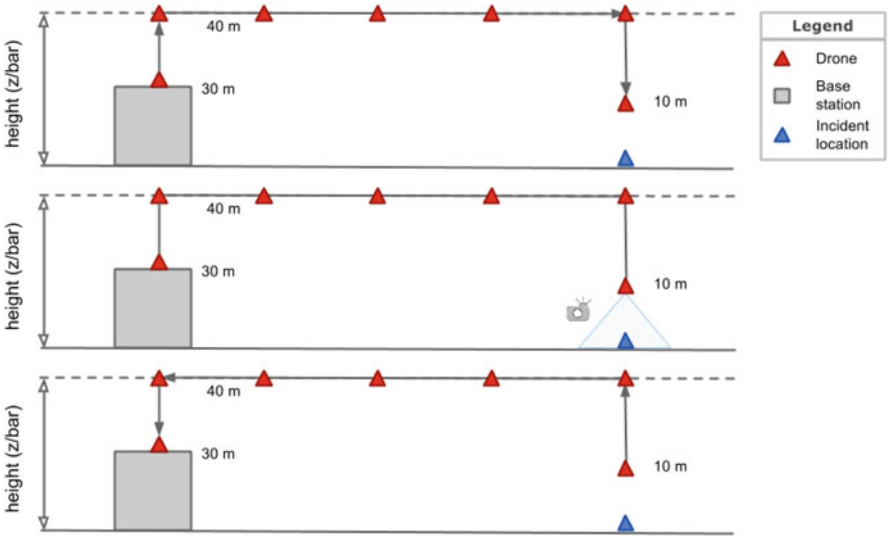

Fig. 15.6 Schematic example of a simple UAV mission

as it is forbidden to fly in these zones, with the exception of high restriction zones, where it is possible to fly under certain conditions, e.g. only flight below 100 m is allowed, or by obtaining a special permit. Both the obstacles and the no-fly zones are stored in geographical databases.

The flight route is defined in the flight plan. To create the flight plan, this project uses the standard defined by the PaparazziUAV autopilot system [17]. The flight plan is composed of two main elements: waypoints and blocks. Waypoints represent geographic locations, and are used to define the trajectories of the UAV. Blocks are sets of instructions given to the UAV. When one block has been accomplished, the UAV autopilot will step to the next block.

The default parameters are defined in the first line of the flight plan: default altitude, base station coordinates and altitude, name of the flight plan, maximum distance from base station, etc. Then, the waypoints are defined, and finally the blocks. The code below represents an example of a short flight plan for the mission shown in Fig. 15.6.

```
<!DOCTYPE flight_plan SYSTEM "flight_plan.dtd">
<flight_plan name="Example Mission" alt="45"
  ground_alt="15" security_height="50" lat0="4.56789"
  lon0="1.23456" max_dist_from_home="3000" qfu="180">
  <waypoints>
    <waypoint name="BASESTATION"
      lat="4.56789" lon="1.23456"/>
    <waypoint name="INCIDENTLOCATION"
      lat="4.111112" lon="1.000111"
```

```
        alt="ground_alt+10."/>
      <waypoint name="1" lat="4.123412" lon="1.123444"
        alt="ground_alt+50"/>
      <waypoint name="2" lat="4.123321" lon="1.123212"
        alt="ground_alt+60"/>
      <waypoint name="3" lat="4.122111" lon="1.122665"
        alt="ground_alt+60"/>
      <waypoint name="4" lat="4.121212" lon="1.111122"
        alt="ground_alt+50"/>
      <waypoint name="5" lat="4.120123" lon="1.111221"
        alt="ground_alt+40"/>
    </waypoints>
    <blocks>
      <block name="toIncident">
        <path wpts="1,2,3,4,5,INCIDENTLOCATION"/>
      </block>
      <block name="toIncident">
        <circle radius="10" wp="INCIDENTLOCATION"
          alt="ground_alt+12" until="90"/>
        <stay wp="INCIDENTLOCATION" alt="ground_alt+10"
          until="90"/>
      </block>
      <block name="returnToBaseStation">
        <path wpts="5,4,3,2,1,BASESTATION"/>
        <stay wp="BASESTATION" alt="44"/>
      </block>
    </blocks>
  </flight_plan>
```

In this flight plan, the default altitude is 45 m, which is the altitude of the ground from sea level plus the altitude of the base station. The security height is 50 m above the default altitude, and the UAV will initially head south (180), towards the incident location. Seven waypoints have been defined: the base station, the incident location and five midway points. There are three blocks containing the three steps of the mission. The first block contains the route from the base station to the incident location. The second block gives the UAV instructions to circle around the incident location for 1.5 min from an altitude of 15 m above ground, and then to stay still at 10 m over the incident location for another 1.5 min, time during which the UAV will capture the imagery of the incident. Finally, the third block will instruct the UAV to return to the base station and hover over it until the base station operator is ready to assist in the landing and gives it the command to land.

After the flight plan has been created and revised, it can be loaded into the UAV, and then it will be ready to start the mission.

15.4.2.3 Send UAV on Mission

Once the flight path has been loaded into the UAV and the operator at the mission control center has authorised the mission, the UAV will begin its mission. A schematic of the mission's flight path is shown in Fig. 15.6. This figure presents a simple mission in three steps:

1. The UAV takes off from base station and flies to the incident location.
2. The UAV stops 10 m above incident location and takes a snapshot.
3. The UAV returns to its base station.

15.4.2.4 Acquire UAV Telemetry

The UAV is constantly streaming its telemetry to the GCS, and the mission control center operator is monitoring its progress, ready to take action when required by the mission. The telemetry is streamed directly to the GCS using a standard protocol. If the drone is connected via Wi-Fi, the standard protocol is MavLink. The main telemetry data provided by the UAV is: location (coordinates), altitude, speed, orientation, inclination, remaining battery and video footage. During the travel to the incident location, the location of the UAV is sent to the emergency services, along with an estimation of the remaining arrival time.

15.4.2.5 Send UAV Imagery to Emergency Services

For privacy reasons, the video footage obtained from the UAV during the travel, during the flight to the incident location and on the way back to the base station is neither recorded nor sent to the emergency services.

Once the UAV has arrived to the incident location, the acquired imagery is sent to the emergency services. During this time, if the emergency services need it, it is possible to give extra instructions to the UAV, such as obtaining a closer look of the incident or the surroundings. Once the aerial support is no longer needed, the connection with the emergency services is terminated. There are multiple methods to provide imagery to the emergency services. If the emergency services only require pictures, then an image sharing service is used. However, if the emergency services require video footage, then a video streaming service is used. In either case, dedicated services must be used to ensure privacy. Making use of public media sharing platforms can be a violation of privacy rights because the UAV is acquiring imagery of emergency situations in public spaces.

15.4.2.6 Return UAV from Mission

When the UAV has accomplished its goal, the connection with the emergency services is terminated. At this moment, the UAV mission control center operator will give confirmation to end the mission and the UAV is sent back to the base station. Once the UAV has arrived to the base station, the operator at the base station will confirm the end of the mission. When the UAV arrives to the base station, the operator executes the command to land the UAV and assists on its landing. After it has landed, the operator is responsible for updating the status of the UAV in the UAV management system, charge its battery and check it for possible damages.

15.5 Conclusions

The system presented aims to provide aerial support to the emergency services by using autonomous drones to acquire live imagery of the location where emergency alerts have originated. Although the first aim of this system is the simplest case of providing aerial imagery, there are more complex disaster management use cases that can be developed on top of this system, such as providing first aid kits, medicines, Internet access points, mapping of affected areas and evacuation monitoring. The possibilities are endless, but currently limited by two main factors: the absence of proper regulations and the early stage of the technology.

Although many countries have already started to develop comprehensive laws and regulations for the use of UAVs, these regulations are currently being used to control and limit its usage and still have a long way ahead until the specific applications of this technology are regulated. However, a good step has been taken by differentiating between the usage of UAVs by governments, by hobbyists and for commercial purposes.

The benefits of autonomous drones versus manually piloted drones rely on the fact that the support does not require the constant dedication and availability of an experienced pilot for each operation. Although it is recommended that there is an experienced pilot always present, the reason is only that someone will be able to pilot a drone if a contingency is presented, but the main work is performed by the UAV Mission Control Center operators that are in charge of constantly monitoring the progress of the mission.

Future work in this field will include the creation of a working prototype of the system, performing a spatial analysis to optimise the locations of the Base Stations, and the design of specialised UAVs to suit the specific needs of the Emergency Services.

Acknowledgements The author wishes to acknowledge Erik van der Zee, Steven Bos and the Geodan Research team for their support and assistance on the analysis of the system.

This work was financially supported by EU FP7 Marie Curie Initial Training Network MULTI-POS (Multi-technology Positioning Professionals) under grant nr. 316528.

References

1. 3DR, Mapping drones (2016). https://3dr.com/mapping-drones/ (visited on 01/28/2016)
2. Airvid, Hire drone pilot (2016). http://air-vid.com/wp/ (visited on 01/28/2016)
3. DJI, No fly zones (2016). http://www.dji.com/fly-safe/category-mc?www=v1 (visited on 01/28/2016)
4. H. Eisenbeiss, A mini unmanned aerial vehicle (UAV): system overview and image acquisition, in *International Archives of Photogrammetry. Remote Sensing and Spatial Information Sciences* 36.5/W1 (2004)
5. FAA, No drone zone (2016). https://www.faa.gov/uas/no_drone_zone/ (visited on 01/28/2016)
6. FAA, Unmanned Aircraft Systems (UAS) Regulations and Policies (2016). https://www.faa.gov/uas/regulations_policies/ (visited on 01/28/2016)
7. P. Fabiani et al., Autonomous flight and navigation of VTOL UAVs: from autonomy demonstrations to out-of-sight flights. Aerosp. Sci. Technol. **11**(2), 183–193 (2007)
8. G. Hattenberger, M. Bronz, M. Gorraz, Using the Paparazzi UAV system for scientific research, in *IMAV 2014, International Micro Air Vehicle Conference and Competition 2014*, Delft. HAL. Hal-id: hal-01059642, Aug (2014), pp. 247–252. doi:10.4233/uuid:b38fbdb7-e6bd-440d-93be-f7dd1457be60
9. S. Hickey, Humanitarian drones to deliver medical supplies to roadless areas (2014). http://www.theguardian.com/world/2014/mar/30/humanitarian-drones-medical-supplies-no-roads-technology (visited on 03/21/2016)
10. J. How, E. King, Y. Kuwata, Flight demonstrations of cooperative control for UAV teams, in *AIAA 3rd Unmanned Unlimited Technical Conference, Workshop and Exhibit*, Chicago, IL. American Institute of Aeronautics and Astronautics (AIAA), pp. 20–23, Sept (2004)
11. H. Kelly, Drones: the future of disaster response (2013). http://edition.cnn.com/2013/05/23/tech/dronesthefuture%20ofdisasterresponse/index.html (visited on 11/09/2015)
12. P. Kemao et al., Design and implementation of a fully autonomous flight control system for a UAV helicopter, in *2007 Chinese Control Conference* (IEEE, New York, 2007), pp. 662–667
13. G. Khaselev, J. Singleton, UAV: Autonomous flight (2014)
14. Know Before You Fly, Know before you fly (2016). http://knowbeforeyoufly.org/ (visited on 02/12/2016)
15. M. Maciag, Law enforcement agencies using drones list, map (2013). http://wwwg.overning.com/gov-data/safety-justice/drones-state-local-law-enforcement-agencies-license-list.html (visited on 03/21/2016)
16. K. Nonami, Prospect and recent research & development for civil use autonomous unmanned aircraft as UAV and MAV. J. Syst. Design Dyn. **1**(2), 120–128 (2007)
17. Paparazzi UAV, Paparazzi UAV, The free autopilot: overview (2016). https://wiki.paparazziuav.org/wiki/Overview (visited on 01/25/2016)
18. A. Pulver, R. Wei, C. Mann, Locating AED enabled medical drones to enhance cardiac arrest response times. Prehosp. Emerg. Care **20**(3), 378–389 (2016)
19. J. San-Miguel-Ayanz, N. Ravail, Active fire detection for fire emergency management: potential and limitations for the operational use of remote sensing. Nat. Hazards **35**(3), 361–376 (2005)
20. Sensefly, Mapping drones applications (2016). https://www.sensefly.com/applications/overview.html (visited on 03/21/2016)
21. T. Tomic et al., Toward a fully autonomous UAV: research platform for indoor and outdoor urban search and rescue IEEE Robot. Autom. Mag. **19**(3), 46–56 (2012). ISSN:1070–9932. doi:10.1109/MRA.2012.2206473
22. B. Van de Walle, M. Turoff, Decision support for emergency situations. IseB **6**(3), 295–316 (2008)
23. Webredactie Communication, TU Delft's ambulance drone drastically increases chances of survival of cardiac arrest patients (2014). http://www.tudelft.nl/en/current/latest-news/article/detail/ambulance-drone-tu-delft-vergroot-overleving%20skans-bij-hartstilstand-drastisch/ (visited on 01/28/2016)

Chapter 16
MULTI-POS: Lessons Learnt from Fellows and Supervisors

Elena-Simona Lohan, Jari Nurmi, Gonzalo Seco-Granados, Henk Wymeersch, and Ossi Nykänen

16.1 Introduction

A training network, such as the Marie Curie networks, enables wide international interconnections and can offer multiple benefits to its participants, such as open access to different units and different research facilities and equipments, easy contacts with people from various scientific areas, and favorable conditions for research and mobility funding for the selected fellows. One of the particularities of these networks, that is, the requirement for hiring only people who have not been living more than 12 months in the 3 years previous to their contract start date in the country of employment, can be both a blessing and a curse.

On one hand, creating a large network with a vast cultural background can foster great learning opportunities in terms of social, inter-personal, and working style skills and can enable the participants to become more open-minded, more flexible, and more tolerant.

On the other hand, adaptation to a new country and a new culture and interacting with the other fellows and supervisors in the network is not always an easy task, and

E.-S. Lohan (✉)
Tampere University of Technology, Korkeakoulunkatu 10, 33720 Tampere, Finland
e-mail: elena-simona.lohan@tut.fi

J. Nurmi • O. Nykänen
Tampere University of Technology, Tampere, Finland
e-mail: ossi.nykanen@tut.fi; jari.nurmi@tut.fi

G. Seco-Granados
Universitat Autonoma de Barcelona, Barcelona, Barcelona, Spain
e-mail: gonzalo.seco@uab.es

H. Wymeersch
Chalmers University of Technology, Gothenburg, Sweden
e-mail: henkw@chalmers.se

J. Nurmi et al. (eds.), *Multi-Technology Positioning*,
DOI 10.1007/978-3-319-50427-8_16

sometimes the administrative and language barriers can make the experience quite difficult.

Nevertheless, in MULTI-POS the positive points triggered by the mobility surpassed the negative points and the overall experience of the fellows has been valued with a 92% of positiveness, according to an anonymous web-based survey conducted in May 2016 among the fellows. Moreover, 58% of the fellows, if given the chance to choose again, would repeat the MULTI-POS experience without any qualms and the other 42% of the fellows would be willing to repeat the experience by paying more attention to choosing the host unit, the host country (e.g., by choosing among countries where they have basic language proficiency), and by being more pro-active in the self-management of time and research tasks.

16.2 Administrative Issues

The administrative issues are mainly seen from the coordinator and supervisors' points of view.

The first administrative lesson to learn was that although the timing of the project seems very relaxed, as the Early Stage Researchers have three-year contracts in a four-year project and thus one year of margin in the recruitment stage, it is, however, likely to have longer recruitment delays. Such long recruitment delays would cause at worst the reduction of the length of the fellow's duration of employment. The fact that the contracts are not reimbursed beyond the end of the project made the contract duration flexibility very low. In MULTI-POS case such delays were caused by:

- Partner withdrawal, causing a chain of management decisions on the withdrawal, search of a partner to take over the position, another decision on moving the position and related funding, restart of the recruitment process, and finally agreeing on the starting date with the selected fellow.
- Difficulties in obtaining a visa and work permit for fellows recruited from outside Europe.
- Resignation of a recruited fellow, leading to management decisions, restart of the recruitment, and matching the starting date to the incoming fellow's situation.
- Delay in obtaining the required degree diplomas (MSc or PhD) by some selected fellows, due, for example, to various administrative delays in their home country.

The main lesson learnt in the recruitment process was that it may be good to start the advertising of the possible open positions even before the final signatures are in place. This would increase the chance of finding excellent candidates.

Another lesson related to the length of the fellow contracts was that in most of the countries, the PhD studies take more than three years. In this kind of a training network, the timing is even more challenging, as the fellow will spend a long time in secondment which will practically spread the focus a bit and cause a new learning curve to be followed by the fellow. Additionally, the training in MULTI-POS was more extensive than in most of the PhD programmes in individual universities. So

much to do, so little time. As a consequence, most of the universities in MULTI-POS will have to find additional funding to support the PhD completion by the fellows.

From the point of view of the local management and administrative issues at each partner in the network, the implementation of a Marie Curie network requires some learning curve for partners not used to this type of funding. In a Marie Curie network, the funding follows the fellow, and in case there is a need to make a decision to transfer the fellow from one employer to another, the budget also has to be transferred. In MULTI-POS, unfortunately a couple of such transfers had to be made in order to ensure that the planned training will take place. Based on MULTI-POS experience, the standard consortium agreements introduce too time-consuming procedures to deal with acute situations where the fellow needs to be "rescued" to enable the training to continue. The management of regular EU projects does not face similar situations as in the fellow-centric Marie Curie networks.

Some learning was needed also at the financial and/or HR departments of the companies and universities. The requirements to use all the living and mobility allowances to the salary of the fellow and related compulsory employer costs, as well as the freely usable lump sum training and overhead allowances were at first interpreted in different ways and have required some advising every now and then. A financial learning was also that the interim payments have been severely delayed due to a number of clarifications needed from one or few of the partners, which affected all partners in the network.

From above, it starts to be obvious that a dedicated project manager is essential in Marie Curie projects. As several changes in the secondment plans had to be made, and new associate partners to host some secondments had to be found and attracted to the network, there has been a lot of effort to follow all the changes in the secondment and associate partner contracts. It proved to be very difficult also to get new associate partners to join the network after its starting date, mostly related to IPR constraints. By default, the IPRs are owned by the employer unless they are explicitly transferred or licensed to the secondment host. A lesson learnt is that the IPR rules and the possibility to transfer or license the results should be very clear and specific from the beginning for each partner in the network, and the secondments plans and possible changes in secondments should be considered with care as soon as possible (i.e., at the network creation, or for the changes, as soon as the need appears).

Figure 16.1 shows the self-evaluation of supervisors about the percentage of time (from their full working time) spent on supervisory issues related to MULTI-POS, such as directly supervising the fellows, meetings, and administrative tasks. Most of the supervisors estimated that they needed between 5% and 10% of their working time to dedicate to these matters. Also, a vast majority of the supervisors (90%) rated their overall experience with MULTI-POS as good or very good, 80% wrote that they perceived their work network to have benefited from MULTI-POS, 100% agreed that they would participate in such a network also in the future if given the chance.

In terms of recruiting a fellow and according to the anonymous survey, the top four characteristics to be looked after in a candidate have been a good technical

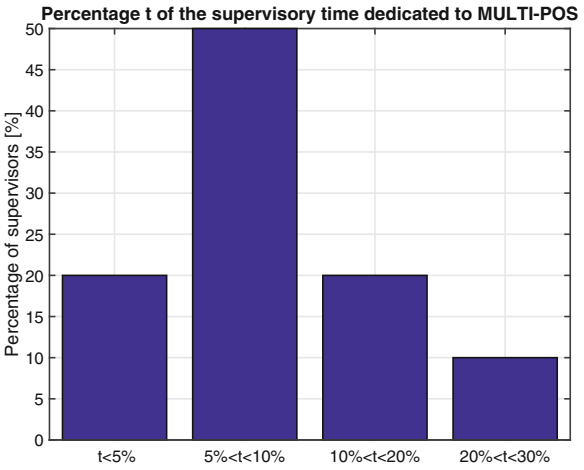

Fig. 16.1 Statistics on the time used in supervising matters in MULTI-POS

background in fields supporting the chosen research area (70% of the supervisors agreed that this is very important), a good technical knowledge in the chosen field (60% of the supervisors), an ability to work hard (50% of the supervisors), and good social skills (50% of the supervisors). Surprisingly, no supervisor considered that the presentation skills and an easy-going personality are among the top characteristics to be searched for in a candidate.

16.3 Scientific Issues

MULTI-POS covers a wide and varied range of research topics. This meant that during the workshops, some fellows had difficulties to follow the research issues presented by other fellows working in a distinct area. It was noticed that the best way to address this challenge has been to allocate the fellows into smaller groups of four or five people (e.g., by work package) and to let them present their work to a reduced audience. This proved to encourage networking, brainstorming, and being pro-active. For instance, this kind of thinking was developed in the workshop in Gothenburg, in 2015. One of the outputs of that workshop was a joint paper submission by five fellows in October 2015. Joint work has also been encouraged during local meetings, for example, when several fellows coincided during their secondments and during conference and workshop attendance. As a result, MULTI-POS has now 13 joint publications where at least two MULTI-POS fellows worked together, plus other 24 publications where at least two MULTI-POS fellows or supervisors coming from different units worked with each other.

MULTI-POS greatly helped the fellows getting access to top researchers and sharing ideas and thoughts with them. Some fellows confessed to have experienced "personal paradigm shifts" when tackling jointly certain research problems, by get-

ting inspiration from the other network members and their collaborators. However, the full potential of the network was not used since the technical skills of each one were used solely towards their own work and not for a common goal. This has been indeed one of the main criticisms of the fellows, when asked about their opinion on MULTI-POS, that the inter-project communication could be improved, and working on common goals with higher amount of scientific overlap could foster better the collaboration. It has also been pointed out by the fellows that the sharing of knowledge during workshops could be easier by hands-on experiments instead of sharing the knowledge only through slide presentations.

The fact that most of the network topics were very focused towards specific and novel issues made that the majority of the hired researchers to have a rather low knowledge in the selected research area. This, combined with the relatively small duration of a maximum of 3 years of the network made some of the fellows to feel that they did not have enough time to absorb all the new knowledge needed or to benefit from a deep learning of the addresses topics.

Another challenge pointed out by the fellows has been the fact that it is not easy to match the academy requirements with companies' interests. In the private sector, which was heavily present both as main hosts and secondment places in MULTI-POS, publications are more restricted, the data sharing is more controlled, and the instructions between the academy and industry supervisors may often be mismatching or even incompatible. The possible solutions to tackle such challenges are to discuss the academic requirements with the industrial units from the very beginning and to agree upon the common interests and to initiate more joint meetings with both academic and industrial supervisors so as to reduce the risk of misunderstandings.

16.4 Personal and Cultural Issues

One of the very pleasant lessons was to find out, in a multi-cultural workshop organized as part of one of the training events, that the fellows and supervisors are representing a broad variety of personalities. That depends on the nationality, but also on individual properties. With this take-away from the workshop we have been able to understand and interpret each other much better, and the team of fellows fits very well together at work and in leisure. There have been very few personal adaptability issues even with "extreme personalities."

What was also learned was that the selective process of recruiting the fellows yielded, in general, very good fellows. However, the remote selection of fellows, based only on phone or video interviews is not always an easy task, and, in some cases, the results can be unsatisfactory for both the fellow and the supervisor. In MULTI-POS, following an anonymous survey, 60% of supervisors said that they have been very happy with their choice of fellows, 30% were moderately happy, and 10% were not satisfied with the outcome. From the fellows' point of view, 40% have been very happy with their supervisors, 40% have been moderately

happy and 20% have been unhappy with the amount of guidance and supervision and the interactions with their supervisors. This shows, on one hand, that there is a non-negligible percent of mismatches between supervisors and fellows, maybe due to different expectations, different work goals and work practices, and different personalities, and, on other hand, that fellows tend to have a more critical view on what they expect in terms of supervision in such kind of a program. Nevertheless, it is not known if these percentages differ from those in conventional PhD programmes and supervision.

A good experience with positive outcome in MULTI-POS has been to pair and group the fellows to work towards common goals, such as joint paper writing, team work during workshops, or joint book chapter writing. That has triggered more interaction and collaboration between the fellows.

As expected, the recruited fellows have different personalities and may have very different motivations: some more technical, others more scientific, other more "qualitative" oriented. The lesson learnt here is that the network goals and the supervisory support should be flexible enough in order to accommodate all the different needs and to try to make all of the participants happy, or at least, to feel that there is a mutual gain from the mutual effort. Another important lesson has also been that the social skills are as important as the technical skills in such a network.

Some fellows also found challenging the economic conditions in the new country, such as expensive accommodation (or poor quality/price ratio for the available accommodations) and long delays in getting receipts reimbursed after a work travel.

Some fellows pointed out that the language barrier can be an important challenge in such a network, especially when the fellows have to work in countries with languages very different from their own native language. A fraction equal to 42% of the fellows confessed to have acquired at least some basic skills of the language of the country of their host or secondment unit, 8% acquired some moderate knowledge of a new language, 17% acquired advanced knowledge of a new language, and 33% of the fellows did not put any effort or did not have any opportunity to learn a new language in the country where they were staying.

In terms of perceived future employability after MULTI-POS, 80% of the fellows are positive that this training has improved significantly their success in any future career path they might choose. The rest of 20% have no opinion or are not sure whether this network training benefited their career or not.

The vast majority of fellows also declared (in the anonymous survey) that their overall experience within MULTI-POS has been good (67% of the answers) or very good (25% of the answers). The remaining 8% opted for an "indifferent" answer, meaning that they neither disliked, nor they liked their MULTI-POS experience.

16.5 Summary

A summary of the main challenges and recommendations of how to address them is shown in Table 16.1.

Table 16.1 Summary of the main challenges and lessons learnt

Challenge	Possible solution(s)
Recruitment delays	Start to advertise the positions as soon as possible, even before the official contract signatures
Choosing an "unsuitable" fellow	Pay attention to both the scientific and social skills during the interviews; try to find "fitting personalities" to the group and adapt supervision strategies
Cooperation level	Inter-fellow cooperation can be increased through joint team work, pairing and grouping; inter-unit collaboration can be increased by regular participation of all supervisors at the network events
Choice of the secondment places	It is good to leave a certain amount of flexibility in choosing the secondment place; IPR issues should be addressed as soon as possible
Possible heavy administrative tasks	A dedicated project manager is a must
	Improve the support structures in the host units
Fellows' motivation	In order to maintain a constant and good self-motivation, setting own goals and following them all through the time in the network can greatly help; also task division for learning and creation can help moving forward
Supervisors' motivation	Some incentives for the supervisors to participate regularly to the networking events should be created
Adaptation to a new culture	Learn at least the basics of the language of the host country
	Promote selflessness and sociability
Network follow-up	Find a hobby to balance the personal/work time division
	Finding sufficient funding for at least one extra year after the project ends for each fellow can increase the fellows' motivation and the project follow-up impact

Chapter 17
Conclusions

Elena-Simona Lohan, Gonzalo Seco-Granados, Henk Wymeersch, Ossi Nykänen, and Jari Nurmi

Navigation and localization are nowadays an intrinsic part of the majority of wireless communication devices. They are also likely to play crucial roles in the new generations of communication devices, such as Device-to-Device (D2D), Internet of Things (IoT) and 5G.

Nowadays it is widely recognized that seamless and ubiquitous navigation and positioning can be only obtained via hybrid solutions, such as GNSS combined with WiFi, Ultra Wide Band (UWB), inertial sensors or other types of sensors and technologies. While there is no unique technology to offer low-cost full-coverage and continuous availability positioning solutions, there are already many commercial solutions combining various technologies and signals and offering accurate and continuous location information under certain target scenarios. At the same time, it is likely that new services and applications will keep emerging thanks to the availability of more power-efficient and lower-cost localization technologies. Nowadays these technologies are under intense development in order to use them in more challenging conditions and, in particular, in indoor scenarios.

E.-S. Lohan
Tampere University of Technology, Korkeakoulunkatu 10, 33720 Tampere, Finland
e-mail: elena-simona.lohan@tut.fi

H. Wymeersch
Chalmers University of Technology, Gothenburg, Sweden
e-mail: henkw@chalmers.se

O. Nykänen • J. Nurmi
Tampere University of Technology, Korkeakoulunkatu 10, 33720 Tampere, Finland
e-mail: ossi.nykanen@tut.fi; jari.nurmi@tut.fi

G. Seco-Granados
Universitat Autonoma de Barcelona, Barcelona, Barcelona, Spain
e-mail: gonzalo.seco@uab.es

J. Nurmi et al. (eds.), *Multi-Technology Positioning*,
DOI 10.1007/978-3-319-50427-8_17

Our book has addressed the different positioning technologies existing nowadays on the market and in the research community, ranging from the four Global Navigation Satellite Systems to local or global solutions employing WiFi, 5G signals, Bluetooth Low Energy, vision navigation or any other signals of opportunity. We have also addressed different application areas where wireless localization is important, ranging from environmental applications to drones-based applications and e-health areas.

The book topics have been addressed in three levels: first the aspects related to satellite navigation at physical layer design were discussed, second the terrestrial non-GNSS solutions were presented, also with the main focus placed on the physical layer, and last but not least, the application areas of wireless positioning were discussed. In addition, a final chapter was dedicated to the lessons learnt during MULTI-POS network, lessons that are hopefully useful not only for people involved in future international training networks, but also at individual level, for people undergoing international mobility for their jobs or education.

About the Editors

Jari Nurmi works as a Professor at Tampere University of Technology, Finland since 1999, in the Faculty of Computing and Electrical Engineering. He is working on embedded computing systems, wireless localization, positioning receiver prototyping, and software-defined radio. He held various research, education, and management positions at TUT since 1987 and was the Vice President of the SME VLSI Solution Oy 1995–1998. Since 2013, he is also a partner and co-founder of Ekin Labs Oy, a research spin-off company commercializing technology for human presence detection, now headquartered in Silicon Valley as Radiomaze, Inc. He has supervised 19 PhD and over 130 MSc theses at TUT, and been the opponent or reviewer of 29 PhD theses for other universities worldwide. He is a senior member of IEEE, member of the technical committee on VLSI Systems and Applications at IEEE CAS, and board member of Tampere Convention Bureau. In 2011, he received IIDA Innovation Award, and in 2013 the Scientific Congress Award and HiPEAC Technology Transfer Award. He is a steering committee member of four international conferences. He has edited five Springer books, and has published over 300 international conference and journal articles and book chapters. He has participated in the Marie Curie ITN network MULTI-POS as the network coordinator and scientist in charge.

Elena Simona Lohan is an Associate Professor at the Department of Electronics and Communication Engineering at Tampere University of Technology (TUT). She is also a Visiting Professor at Universitat Autònoma de Barcelona, Spain. She received an M.Sc. degree in Electrical Engineering from Polytechnics University of Bucharest, Romania, in 1997, a D.E.A. degree (French equivalent of master) in Econometrics, at École Polytechnique, Paris, France, in 1998, and a PhD degree in Telecommunications from TUT, in 2003. She has more than 170 international publications, 3 patents, and 2 patent applications. She is serving as Associate Editor to IET Radar, Sonar and Navigation journal and to RIN Cambridge Journal of Navigation since 2013. Her current research interests include wireless location

J. Nurmi et al. (eds.), *Multi-Technology Positioning*,
DOI 10.1007/978-3-319-50427-8

techniques based on signals of opportunity and cognitive spectrum sensing for positioning purposes. She has participated in the Marie Curie ITN network MULTI-POS as scientist in charge and equality officer.

Henk Wymeersch is a Professor in Communication Systems with the Department of Signals and Systems at Chalmers University of Technology, Sweden. Prior to joining Chalmers, he was a postdoctoral researcher from 2005 until 2009 with the Laboratory for Information and Decision Systems at the Massachusetts Institute of Technology. Henk Wymeersch obtained the PhD degree in Electrical Engineering/Applied sciences in 2005 from Ghent University, Belgium. He served as Associate Editor for IEEE Communication Letters (2009–2013), IEEE Transactions on Wireless Communications (since 2013), and IEEE Transactions on Communications (since 2016). His current research interests include cooperative systems and intelligent transportation. He has participated in the Marie Curie ITN network MULTI-POS as scientist in charge.

Gonzalo Seco-Granados received the PhD degree on Telecommunications Engineering from Universitat Politècnica de Catalunya in 2000 and an MBA from IESE in 2002. From 2002 to 2005, he was member of the technical staff of the European Space Agency, involved in the design of the Galileo system. Since 2006, he is Associate Professor at the Department of Telecommunications, Universitat Autònoma de Barcelona. He has been principal investigator of over 25 research projects. In 2014, he received an ICREA Acadèmia fellowship. In 2015, he was Fulbright Visiting Scholar at University of California, Irvine. His research interests include the design of signals and reception techniques for satellite-based and terrestrial positioning systems, multi-antenna receivers, and signal-level integrity. He has participated in the Marie Curie ITN network MULTI-POS as scientist in charge.

Ossi Nykänen works for M-Files, helping enterprises find, share, and secure documents and information. Dr.O. Nykänen also serves as an Adjunct Professor at Tampere University of Technology, Department of Mathematics. His long-term research interests include semantic computing, machine learning, information modeling and visualization, (computer-supported) mathematics, and the related applications. He is affiliated with many industrial and academic research networks, and is an advocate of international web standards. He has participated in the Marie Curie ITN network MULTI-POS as scientist in charge.

Acronyms

A-GPS Assisted Global Positioning System. 292, 295
ADC Analog to Digital Converter. 71, 80, 180, 181
AESA Spanish Air Safety Agency. 305
AGC Automatic Gain Control. 71
AHP Analytical Hierarchical Process. 3, 286, 288, 290, 294, 297, 298
ALL-SIDES Advanced Large-Scale Language Analysis for Social Intelligence Deliberation Support. 265
AltBOC Alternate Binary Offset Carrier. 117
AM Amplitude Modulation. 67
ANN Artificial Neural Network. 60
AOA angle-of-arrival. 167
AOD angle-of-departure. 167
AWGN Additive White Gaussian Noise. 16, 84

BDT BeiDou Time. 120
BLE Bluetooth Low Energy. 129, 153, 162, 292, 293, 295
BOC Binary Offset Carrier. 19–22, 64, 111, 117
BPSK Binary Phase-Shift Keying. 19–22, 49, 64, 109, 110, 113, 119
BS Base Station. 169–171, 182, 320

C/A Coarse acquisition. 64
C/N_0 carrier-to-noise density. 63, 95–97, 100, 103
CAS Context-aware Systems. 251, 264, 265
CBOC Composite Binary Offset Carrier. 84, 116
CDMA Code Division Multiple Access. 15, 20, 79, 109, 113, 115, 118
CE Context Engine. 251–254, 256, 259, 265
CGCS2000 China Geodetic Coordinate System 2000. 120

J. Nurmi et al. (eds.), *Multi-Technology Positioning*,
DOI 10.1007/978-3-319-50427-8

CM Crisis Management. 303
COA Certificate Of Authorization. 307
COTS Commercial off-the-shelf. 98
CRLB Cramér-Rao Lower bound. 21
CS Commercial Service. 115–117
CW Continuous Wave. 71, 72

DAC digital-to-analog-converter. 180, 181
dBc Decibels relative to the carrier. 68
DHM Deterministic Homophily Method. 263
DLL Delay Lock Loop. 62, 80, 82–85, 90, 91, 95
DME Distance Measuring Equipment. 69, 74
DOA Direction Of Arrival. 73
Doppler Doppler frequency shift. 98
DP Dot Product. 85
DTV Digital Television. 150
DVB-T Digital Video Broadcasting -Terrestrial. 87

E Early Correlator. 84
EASA European Aviation Safety Agency. 305
ECEF Earth Centered Earth Fixed. 92, 93, 96, 97
EKF Extended Kalman Filter. 90, 91, 93–95, 103
EML Early-Minus-Late. 64, 84
EMS Emergency Management System. 314
ENAC Ecole Nationale de l'Aviation Civile. 95
EPN EUREF Permanent Network. 60
ES Emergency Services. 320
ESA European Space Agency. 6

FAA Federal Aviation Agency. 305–307
FCC Federal Communications Commission. 149, 163
FDMA Frequency Division Multiple Access. 14, 15, 20, 112, 113
FEC Forward Error Correction. 111
FLL Frequency Lock Loop. 80, 82, 85, 86, 90, 91
FM Frequency Modulation. 67, 68
FOC Fully Operational Capability. 115
FSL Free Space Loss. 152

Galileo Galileo, the European Global Navigation Satellite System. 82, 84, 92, 94, 95, 103
GCS UAV Ground Control System. 308, 314–316, 319

GDOP Geometric Dilution of Precision. 121
GEO Geostationary Earth Orbit. 118, 119
GEORED Geodesia: Red de Estudios de Deformación. 60
GGTO Galileo to GPS Time Offset. 117
GIM Global Ionospheric Map. 59, 60
GLONASS Globalnaya Navigazionnaya Sputnikovaya Sistema. 14, 15, 20, 45, 47, 149
GMS Galileo Mission Segment. 117
GNSS Global Navigation Satellite System. 1–3, 6, 9, 11, 13–17, 19–21, 23–27, 29, 31, 33, 35, 37, 39, 41, 43, 45–47, 49, 51, 53, 55–57, 59–61, 63–73, 79–83, 85–90, 95, 98, 103, 107, 108, 110–113, 118–123, 125, 135, 136, 149, 150, 156, 163, 269, 270, 272, 273, 275
GPS Global Positioning System. 14, 19, 25, 26, 45, 47, 51, 59, 62, 64, 80, 84, 85, 90, 92, 94, 95, 103, 149, 166, 292
GPST GPS Time. 111, 114
GST Galileo System Time. 117
GTRF Galileo Terrestrial reference Frame. 117

HRZ UAV High Restriction Fly Zone. 305, 306

I In Phase. 83
IAR Integer Ambiguity Resolution. 98, 99
ICD Interface Control Document. 119
IF Intermediate Frequency. 80, 82
IGS International GNSS Service. 60, 82
IGSO Inclined Geosynchronous Orbit. 118, 119
ILS Instrument Landing System. 67, 68
IMU Inertial Measurement Units. 135, 156
INS Inertial Navigation System. 88, 101, 135
IR Infrared. 308
IRI International Reference Ionosphere. 59
ITL Human Environment and Transport Inspectorate. 305
ITN Initial Training Network. 4, 6
ITRF International Terrestrial Reference Frame. 112, 117, 122
ITU International Telecommunication Union. 67, 68, 152

KF Kalman Filter. 81, 97, 103

L Late Correlator. 84
LBS Location-Based Services. 2, 6, 8, 11, 252, 264, 265, 270, 283–287
LED Light Emitting Diodes. 292

LiDAR Light Detection and Ranging. 307, 308
LMS Least Mean Squares. 72
LNA Low Noise Amplifier. 80
LOS line-of-sight. 60, 61, 87, 89, 90, 93, 99, 166, 167, 169, 170

MAC Medium Access Control Address. 158, 159, 161
MARS Multivariate Adaptive Regression Splines. 60
MAV Micro unmanned Air Vehicle. 304
MBOC Multiplexed Binary Offset Carrier. 116, 119
MCC UAV Mission Control Center. 314, 320
MCDM Multi Criteria Decision Making. 286, 298
MDC Mobile Data Challenge. 257, 259
MEDLL Multipath Estimating Delay Lock Loop. 65
MEMS Micro-Machined Electro-Mechanical sensors. 87, 88
MEO Medium Earth Orbit. 109, 119
MEOLUT Medium Earth Orbit Local User Terminals. 115
MIMO Multiple Input, Multiple Output. 156, 165, 167, 175
mm-wave millimeter wave. 165–170, 173–177, 179–183, 187, 188, 191, 192
MS mobile station. 182, 183

NCO Numerically Controlled Oscillator. 82, 83, 90–92, 94
NFZ UAV No-Fly Zone. 305, 306
NLOS Non Line-of-Sight. 11, 87–89, 99, 103, 169

OLOS obstructed-line-of-sight. 177
OS Open Service. 115, 116, 119
OWL Web Ontology Language. 255

PLL Phase Lock Loop. 80, 82, 85, 86, 90, 95
PMR Private Mobile Radio. 68
PPP Precise Point Positioning. 101–104
PRN pseudo-random noise. 80, 107, 109–111, 113, 116
PRS Public Regulated Service. 115–117
PSD Power Spectral Density. 16, 23–28, 32, 33, 35, 36, 93
PVT Position, Velocity, and Time. 66, 69–71, 73, 79, 91, 95, 103, 107, 108, 120–122
PZ-90 Parametry Zemli 1990 (Parameters of the Earth 1990). 114, 115, 122

Q Quadrature. 83
QPSK Quadrature Phase Shift Keying. 111, 113, 117, 119

REGME Red GNSS de Monitoreo Continuo del Ecuador. 60
RF Radio Frequency. 15, 17, 24, 26, 27, 30–33, 57, 66–73, 80, 95, 107, 108, 136, 137
RFID Radio Frequency Identification. 1, 137, 291–293, 295
RHCP Right Hand Circularly Polarized. 65, 118
RM Random Method. 263
RMSE root mean square errors. 97
RSS Received Signal Strength. 151, 155, 156, 158, 161, 162
RTK Real Time Kinematics. 101, 102, 104
RTLS Real Time Location Service. 5
RX receiver. 39

SAR Search And Rescue. 309
SHRM Structural Homophily Randomized Method. 263
SIS Signal-in-Space. 80
SME Small or Medium-size Enterprise. 6, 7
SNA Social Network Analysis. 259, 261
SNR Signal-to-Noise Ratio. 22, 24–26, 31, 32, 63, 69, 72
SOCAM Service-Oriented Context-Aware Middleware. 255
SSN W3C Semantic Sensor Network. 254
ST Scalar Tracking. 89
SV Space Vehicle. 109, 111, 113

T-DAB Terrestrial Digital Audio Broadcasting. 68
TACAN Tactical Air Navigation System. 69
TEC Total Electron Content. 58–60, 82
TMBOC Time-Multiplexed Binary Offset Carrier. 111
TOA time-of-arrival. 165–167
TOW Time of Week. 80
Trafi Finnish Transport Safety Agency. 305
TX transmitter. 38

UAS Unmanned Aerial System. 305
UAV Unmanned Air Vehicle. 303, 304, 307, 308, 311–313, 315–320
UAVs Unmanned Air Vehicles. 303–308, 310–312, 314–316, 320
UERE User Equivalent Range Error. 81, 82
UHF Ultra High Frequency. 67, 68, 87
UTC Coordinated Universal Time. 111, 114, 117, 120, 122
UWB Ultra Wide Band. 1, 137, 149, 166, 167, 174, 292, 295

VDFLL Vector Delay Frequency Lock Loop. 90–95, 97, 103
VDLL Vector Delay Lock Loop. 90
VFLL Vector Frequency Lock Loop. 90
VHF Very High Frequency. 67, 68, 87
VOR VHF Omni-directional Range. 67, 68
VT Vector Tracking. 89, 90

WARTK Wide Area RTK. 101
WGS-84 World Geodetic System 1984. 112, 122, 158
WiFi Wireless LAN network or, arguably, Wireless Fidelity. 149–151, 155, 156, 158, 159, 163, 292–295
WLAN Wireless Local Area Networks. 1, 292
WLS Weighted Least Square. 81

Index

J. Nurmi et al. (eds.), *Multi-Technology Positioning*,
DOI 10.1007/978-3-319-50427-8