POSITION, NAVIGATION, AND TIMING TECHNOLOGIES IN THE 21ST CENTURY

INTEGRATED SATELLITE NAVIGATION, SENSOR SYSTEMS, AND CIVIL APPLICATIONS

VOLUME 2

EDITED BY

Y. JADE MORTON • FRANK VAN DIGGELEN

JAMES J. SPILKER, JR. • BRADFORD W. PARKINSON

ASSOCIATE EDITORS SHERMAN LO • GRACE GAO

IEEE PRESS

WILEY

Position, Navigation, and Timing Technologies in the 21st Century

IEEE Press
445 Hoes Lane
Piscataway, NJ 08854

Position, Navigation, and Timing Technologies in the 21st Century

Integrated Satellite Navigation, Sensor Systems, and Civil Applications

Volume 2

Edited by

Y. T. Jade Morton, University of Colorado Boulder
Frank van Diggelen, Google
James J. Spilker, Jr., Stanford University
Bradford W. Parkinson, Stanford University

Associate Editors: Sherman Lo, Stanford University
Grace Gao, Stanford University

IEEE PRESS

WILEY

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

Published simultaneously in Canada.

Library of Congress Cataloging-in-Publication Data

Names: Morton, Y. Jade, editor. | Van Diggelen, Frank Stephen Tromp, editor. | Spilker, James J., Jr., 1933- editor. | Parkinson, Bradford W., editor.
Title: Position, navigation, and timing technologies in the 21st century : integrated satellite navigation, sensor systems, and civil applications / editors, Y. Jade Morton, University of Colorado Boulder, Frank van Diggelen, Google, James J. Spilker, Jr., Stanford University, Bradford W. Parkinson, Stanford University.
Description: First edition. | Hoboken : Wiley/IEEE Press, [2020] | Includes index. | Contents: Introduction, early history, and assuring PNT (PTA) / Bradford W. Parkinson, Stanford University, USY, T. Jade Morton, University of Colorado Boulder, US, Frank van Diggelen, Google, US, James J. Spilker Jr., Stanford University, US – Fundamentals of satellite-based navigation and timing / John W. Betz, The Mitre Corporation, US – The Navstar Global Positioning System / John W. Betz, The Mitre Corporation, US.
Identifiers: LCCN 2020044255 (print) | LCCN 2020044256 (ebook) | ISBN 9781119458340 (set ; hardback) | ISBN 9781119458418 (volume 1 ; hardback) | ISBN 9781119458494 (volume 2 ; hardback) | ISBN 9781119458463 (volume 1 ; adobe pdf) | ISBN 9781119458456 (volume 1 ; epub) | ISBN 9781119458425 (volume 2 ; adobe pdf) | ISBN 9781119458401 (volume 2 ; epub)
Subjects: LCSH: Mobile geographic information systems.
Classification: LCC G109.4 .P67 2020 (print) | LCC G109.4 (ebook) | DDC 910.285–dc23
LC record available at https://lccn.loc.gov/2020044255
LC ebook record available at https://lccn.loc.gov/2020044256

Cover Design: Wiley
Cover Images: Global telecommunication network
© NicoElNino/Getty Images, GPS Satellite
© BlackJack3D/iStockphoto

Set in 9.5/12.5pt STIXTwoText by SPi Global, Pondicherry, India

10 9 8 7 6 5 4 3 2 1

In Memory of:
Ronald L. Beard
Per Enge
Ronald Hatch
David Last
James J. Spilker, Jr.
James B. Y. Tsui

Contents

Preface

The ability to navigate has been an essential skill for survival throughout human history. As navigation has advanced, it has become almost inseparable from the ability to tell time. Today, position, navigation, and timing (PNT) technologies play an essential role in our modern society. Much of our reliance on PNT is the result of the availability of the Global Positioning System (GPS) and the growing family of Global Navigation Satellite Systems (GNSSs). Satellite-based navigation and other PNT technologies are being used in the many fast-growing, widespread, civilian applications worldwide. A report sponsored by the US National Institute of Standards and Technology (NIST) on the economic benefits of GPS indicated that GPS alone has generated a $1.4 trillion economic benefit in the private sector by 2019, and that the loss of GPS service would have a $1 billion per-day negative impact.[1] PNT has become a pillar of our modern society. Knowledge and education are essential for the continued advancement of PNT technologies to meet the increasing demand from society. That is the rationale that led to the creation of this book.

While there are many publications and several outstanding books on satellite navigation technologies and related subjects, this two-volume set offers a uniquely comprehensive coverage of the latest developments in the broad field of PNT and has been written by world-renowned experts in each chapter's subject area. It is written for researchers, engineers, scientists, and students who are interested in learning about the latest developments in satellite-based PNT technologies and civilian applications. It also examines alternative navigation technologies based on other signals and sensors and offers a comprehensive treatment of integrated PNT systems for consumer and commercial applications.

The two-volume set contains 64 chapters organized into six parts. Each volume contains three parts. Volume 1 focuses on satellite navigation systems, technologies, and applications. It starts with a historical perspective of GPS and other related PNT developments. Part A consists of 12 chapters that describe the fundamental principles and latest developments of all global and regional navigation satellite systems (GNSSs and RNSSs), design strategies that enable their coexistence and mutual benefits, their signal quality monitoring, satellite orbit and time synchronization, and satellite- and ground-based systems that provide augmentation information to improve the accuracy of navigation solutions. Part B contains 13 chapters. These provide a comprehensive review of recent progress in satellite navigation receiver technologies such as receiver architecture, signal tracking, vector processing, assisted and high-sensitivity GNSS, precise point positioning and real-time kinematic (RTK) systems, direct position estimation techniques, and GNSS antennas and array signal processing. Also covered are topics on the challenges of multipath-rich urban environments, in handling spoofing and interference, and in ensuring PNT integrity. Part C finishes the volume with 8 chapters on satellite navigation for engineering and scientific applications. A review of global geodesy and reference frames sets the stage for discussions on the broad field of geodetic sciences, followed by a chapter on the important subject of GNSS-based time and frequency distribution. GNSS signals have provided a popular passive sensing tool for troposphere, ionosphere, and Earth surface monitoring. Three chapters are dedicated to severe weather, ionospheric effects, and hazardous event monitoring. Finally, a comprehensive treatment of GNSS radio occultation and reflectometry is provided.

The three parts in Volume 2 address PNT using alternative signals and sensors and integrated PNT technologies for consumer and commercial applications. An overview chapter provides the motivation and organization of the volume, followed by a chapter on nonlinear estimation methods which are often employed in navigation system modeling and sensor integration. Part D devotes 7 chapters to using various radio signals transmitted from sources on the ground, from aircraft, or from low Earth orbit (LEO) satellites for PNT purposes. Many of these signals were

1 RTI International Final Report, Sponsored by the US National Institute of Standards and Technology, "Economic Benefits of the Global Positioning System (GPS)," June 2019.

intended for other functions, such as broadcasting, networking, and imaging and surveillance. In Part E, there are 8 chapters covering a broad range of non-radio frequency sensors operating in both passive and active modes to produce navigation solutions, including MEMS inertial sensors, advances in clock technologies, magnetometers, imaging, LiDAR, digital photogrammetry, and signals received from celestial bodies. A tutorial-style chapter on multiple approaches to GNSS/INS integration methods is included in Part E. Also included in Part E are chapters on the neuroscience of navigation and animal navigation. Finally, Part F presents a collection of work on contemporary PNT applications such as surveying and mobile mapping, precision agriculture, wearable systems, automated driving, train control, commercial unmanned aircraft systems, aviation, satellite orbit determination and formation flying, and navigation in the unique Arctic environment.

The chapters in this book were written by 131 authors from 18 countries over a period of 5 years. Because of the diverse nature of the authorship and the topics covered in the two volumes, the chapters were written in a variety of styles. Some are presented as high-level reviews of progress in specific subject areas, while others are tutorials with detailed quantitative analysis. A few chapters include links to MATLAB or Python example code as well as test data for those readers who desire to have hands-on practice. The collective goal is to appeal to industry professionals, researchers, and academics involved with the science, engineering, and application of PNT technologies. A website, pnt21book.com, provides chapter summaries; downloadable code examples, data, worked homework examples, select high-resolution figures, errata, and a way for readers to provide feedback.

A comprehensive project of this scale would not be possible without the collective efforts of the GNSS and PNT community. We appreciate the leading experts in the field taking time from their busy schedules to answer the call in contributing to this book. Some of the authors also provided valuable input and comments to other chapters in the book. We also sought input from graduate students and postdocs in the field as they will be the primary users and represent the future of the field. We want to acknowledge the following individuals who have supported or encouraged the effort and/or helped to improve the contents of the set: Michael Armatys, Penina Axelrad, John Betz, Rebecca Bishop, Michael Brassch, Brian Breitsch, Phil Brunner, Russell Carpenter, Charles Carrano, Ian Collett, Anthea Coster, Mark Crews, Patricia Doherty, Chip Eschenfelder, Hugo Fruehauf, Gaylord Green, Richard Greenspan, Yu Jiao, Kyle Kauffman, Tom Langenstein, Gerard Lachapelle, Richard Langley, Robert Lutwak, Jake Mashburn, James J. Miller, Mikel Miller, Pratap Misra, Oliver Montenbruck, Sam Pullen, Stuart Riley, Chuck Schue, Logan Scott, Steve Taylor, Peter Teunissen, Jim Torley, A. J. van Dierendonck, Eric Vinande, Jun Wang, Pai Wang, Yang Wang, Phil Ward, Dongyang Xu, Rong Yang, and Zhe Yang. The Wiley-IEEE Press team has demonstrated great patience and flexibility throughout the five-year gestation period of this project. And our families have shown great understanding, generously allowing us to spend a seemingly endless amount of time to complete the set.

This project was the brainchild of Dr. James Spilker, Jr. He remained a fervent supporter until his passing in October 2019. A pioneer of GPS civil signal structure and receiver technologies, Dr. Spilker was truly the inspiration behind this effort. During the writing of this book set, several pioneers in the field of GNSS and PNT, including Ronald Beard, Per Enge, Ronald Hatch, David Last, and James Tsui also passed away. This set is dedicated to these heroes and all those who laid the foundation for the field of PNT.

Jade Morton
Frank van Diggelen
Bradford Parkinson
Sherman Lo
Grace Gao

Contributors

Zuheir Altamimi
Institut National de l'Information Géographique et Forestière, France

Chi O. Ao
Jet Propulsion Laboratory, United States

Benjamin W. Ashman
National Aeronautics and Space Administration, United States

Yoaz Bar-Sever
Jet Propulsion Lab, United States

Miguel Manteiga Bautista
European Space Agency, the Netherlands

John W. Betz
The MITRE Corporation, United States

David Bevly
Auburn University, United States

Sunil Bisnath
York University, Canada

Yehuda Bock
Scripps Institution of Oceanography, United States

Alexei Bolkunov
PNT Center, Russia

Harrison Bourne
University of Colorado Boulder, United States

Brian Breitsch
University of Colorado Boulder, United States

Hugues Brenot
Royal Belgian Institute for Space Aeronomy, Belgium

Jacob Campbell
Air Force Research Laboratory, United States

Aaron Canciani
Air Force Institute of Technology, United States

J. Russell Carpenter
National Aeronautics and Space Administration, United States

Eric Châtre
European Commission, Belgium

Chi-Chih Chen
The Ohio State University, United States

Pau Closas
Northeastern University, United States

H. Stewart Cobb
Satelles, United States

Simone D'Amico
Stanford University, United States

Evan Dill
National Aeronautics and Space Administration, United States

Gillian Durieux
Max Plank Institute for Evolutionary Biology, Germany

Alejandro Egido
Starlab, Spain

Naser El-Sheimy
University of Calgary, Canada

Per K. Enge
Stanford University, United States

James Farrell
Vigil Inc., United States

Alissa M. Fitzgerald
A.M. Fitzgerald & Associates, LLC, United States

Grace Gao
Stanford University, United States

James Garrison
Purdue University, United States

Christine Gommenginger
National Oceanography Centre, United Kingdom

Mark Gretton
TomTom, United Kingdom

Paul D. Groves
University College London, United Kingdom

Robert Guinness
Finnish Geospatial Research Institute, Finland

Sanjeev Gunawardena
Air Force Institute of Technology, United States

Inder J. Gupta
The Ohio State University, United States

Greg Gutt
Satelles, United States

Maarten Uijt de Haag
Technische Universität Berlin, Germany

Jörg Hahn
European Space Agency, the Netherlands

Leo Hollberg
Stanford University, United States

Todd Humphreys
University of Texas–Austin, United States

Mathieu Joerger
Virginia Tech, United States

Gary Johnson
Geoscience Australia, Australia

Sergey Karutin
PNT Center, Russia

Zaher (Zak) M. Kassas
University of California Irvine, United States

Kyle Kauffman
Integrated Solutions for Systems, United States

Yasuhiko Kawazu
National Space Policy Secretariat, Japan

Norvald Kjerstad
Norwegian University of Science and Technology, Norway

Satoshi Kogure
National Space Policy Secretariat, Japan

Attila Komjathy
Jet Propulsion Laboratory, United States

Zoltan Koppanyi
The Ohio State University, United States

Heidi Kuusniemi
Finnish Geospatial Research Institute, Finland

Arthur F. Lange
Trimble Navigation, United States

Zahra Lari
Leica Geosystems Inc., Canada

Kristine M. Larson
University of Colorado Boulder, United States

Matthew V. Lashley
Georgia Tech Research Institute, United States

David Lawrence
Satelles, United States

Judah Levine
National Institute of Standard and Technology, United States

Miriam Liedvogel
Max Plank Institute for Evolutionary Biology, Germany

Sherman Lo
Stanford University, United States

Mingquan Lu
Tsinghua University, China

Anthony J. Mannucci
Jet Propulsion Laboratory, United States

Francisco Martin
Starlab, Spain

Scott Martin
Auburn University, United States

Gary A. McGraw
Collins Aerospace, United States

Subbu Meiyappan
NextNav LLC, United States

Xing Meng
Jet Propulsion Laboratory, United States

Meredith E. Minear
University of Wyoming, United States

Antonio Mollfulleda
Starlab, Spain

Oliver Montenbruck
German Aerospace Center, Germany

Y.T. Jade Morton
University of Colorado Boulder, United States

Alessandro Neri
University of Roma TRE, Italy

Felipe Nievinski
UFRGS, Brazil

Andrew O'Brien
The Ohio State University, United States

Michael O'Connor
Satelles, United States

Bradford W. Parkinson
Stanford University, United States

Sudeep Pasricha
Colorado State University, United States

Ganesh Pattabiraman
NextNav LLC, United States

Peter Frans Pauwels
TomTom, the Netherlands

John Peake
Trimble Navigation, United States

Wouter Pelgrum
Blue Origin LLC, United States

Boris Pervan
Illinois Institute of Technology, United States

Mark Psiaki
Virginia Tech, United States

Sam Pullen
Stanford University, United States

Arun Raghupathy
NextNav LLC, United States

Vyasaraj Rao
Accord Software and Systems, India

John F. Raquet
Integrated Solutions for Systems, United States

Tyler G. R. Reid
Stanford University, United States

Charles Rino
University of Colorado Boulder, United States

Chris Rizos
University of New South Wales, Australia

José Ángel Ávila Rodríguez
European Space Agency, the Netherlands

Giulio Ruffini
Starlab, Spain

Takeyasu Sakai
National Institute of Maritime, Port, and Aviation Technology, Japan

Charles Schue, III
UrsaNav, Inc., United States

Logan Scott
LS Consulting, United States

James Sennott
Tracking and Imaging Systems, United States

Tesalee K. Sensibaugh
University of Wyoming, United States

Suneel Sheikh
ASTER Labs, Inc., United States

Stephen P. Smith
The Charles Stark Draper Laboratory Inc., United States

Andrey Soloviev
QuNav, United States

James J. Spilker Jr.
Stanford University, United States

Thomas A. Stansell, Jr.
Stansell Consulting, United States

Peter Steigenberger
German Aerospace Center, Germany

Nikolai Testoedov
PNT Center, Russia

Peter J. G. Teunissen
Curtin University, Australia and Delft University of Technology, The Netherlands

Sarang Thombre
Finnish Geospatial Research Institute, Finland

Charles Toth
The Ohio State University, United States

Andrei Tyulin
PNT Center, Russia

Sabrina Ugazio
Ohio University, United States

Frank van Diggelen
Google, United States

Frank van Graas
Ohio University, United States

Panagiotis Vergados
Jet Propulsion Laboratory, United States

Michael J. Veth
Veth Research Associates, United States

Todd Walter
Stanford University, United States

Shimon Wdowinski
Florida International University, United States

David Whelan
University of California San Diego, United States

Walton Williamson
Jet Propulsion Laboratory, United States

Chun Yang
Sigtem Technology Inc., United States

Rong Yang
University of Colorado Boulder, United States

Zhe Yang
University of Colorado Boulder,
United States

Zheng Yao
Tsinghua University, China

Steven Young
National Aeronautics and Space Administration,
United States

Valery U. Zavorotny
National Oceanic and Atmospheric Administration,
United States;
University of Colorado Boulder, United States

Zhen Zhu
East Carolina University,
United States

Part D

Position, Navigation, and Timing Using Radio Signals-of-Opportunity

35

Overview of Volume 2: Integrated PNT Technologies and Applications

John F. Raquet

Integrated Solutions for Systems, United States

There is little doubt that global navigation satellite systems (GNSS) have changed the way that we think about and use navigation systems. Prior to GPS and other GNSSs, the use of systems which could automatically (without human intervention) determine their own position was generally limited to large, expensive platforms such as aircraft or ships, and even these types of vehicles often required human navigators to assist in the task of navigation. This has all changed with the advent of GNSS, however.

Thanks to GNSS, most people have now become accustomed to their smartphone or vehicle knowing exactly where it is as a part of their everyday lives, and this capability has been built into our expectations. Just as we expect the lights to come on when we turn on a light switch, we also expect a GNSS position fix whenever we turn on a smartphone or other navigation device. This reliance on GNSS goes well beyond obvious navigation devices – we very much depend on many systems which heavily use GNSS for timing purposes, such as banking, communications, and our power grid.

Some have said that navigation is addictive – no matter how much accuracy or availability you have, you always want more. The extreme success of GNSS has, ironically, led to a desire to complement GNSS with other types of sensors for situations in which GNSS is not available, in order to guarantee (as much as is possible) the ability to determine time or position.

Volume 2 focuses in on many of these complementary navigation systems and methods and how they are integrated together to obtain the desired performance. Before diving into the details, it can be helpful to step back and look at the big picture of what is really happening within navigation systems, in order to better understand how the various approaches relate to each other. To do this, it is helpful to develop a "navigation framework."

35.1 Generalized Navigation Framework

Fundamentally, virtually every navigation system operates the same way. This can be expressed as a predict–observe–compare cycle, as shown in Figure 35.1. The "Navigation State" at the lower right represents the user's current navigation state, or all of the information about the user's position, velocity, and so on, as well as estimates of that information's quality. This can be thought of as the system's best guess of the user's position as well as how accurate the system thinks the guess is. As depicted in the "Sensor" box on the left, the system takes a measurement or makes an observation which gives some insight into the user's navigation state. For GPS, perhaps the system observes the range to a satellite. The system also uses a model of the real world, depicted with the "World Model" box in the upper right. In the case of GPS, the world model might consist of the locations (orbits) of the GPS satellites.

During the predict phase, the prediction algorithm determines what the system expects to observe based upon the world model and the current navigation state, annotated as the "Prediction Algorithm" box in Figure 35.1. During the observe phase, the system receives a noise-corrupted measurement from the real world. During the compare phase, the predicted measurement is compared to the actual measurement. Any discrepancies are used to improve the navigation state and possibly the model of the world.

Consider a simplified example in which a user attempts to determine the distance to a wall. Perhaps the user predicts the distance to the wall is about 30 feet based upon mere eyesight to judge the distance. (The navigation state is 30 feet with much uncertainty.) Then, suppose a precise

Position, Navigation, and Timing Technologies in the 21st Century: Integrated Satellite Navigation, Sensor Systems, and Civil Applications, Volume 2, First Edition. Edited by Y. T. Jade Morton, Frank van Diggelen, James J. Spilker Jr., and Bradford W. Parkinson.

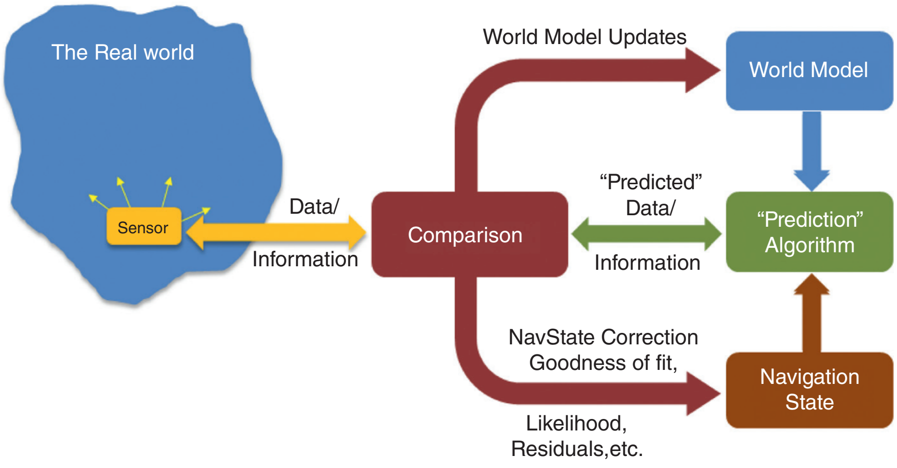

Figure 35.1 General navigation framework.

laser range finder is used to measure, or observe, the distance as 31.2 feet. Next, the prediction is compared to the observation. The user quickly dismisses the prediction and trusts the observation, because the user observation was viewed as being a more reliable estimate of distance than the prediction. Likewise, examples could be drawn which highlight the prediction heavily outweighing an observation.

The most interesting applications involve a blending of the prediction with the observation. Typical GPS applications use a Kalman filter to perform the predict–observe–compare cycle. The world model consists of GPS satellite locations. Based upon some prior information, the receiver predicts the user's location. The observations might consist of ranges to each satellite in view. These observations are compared to a prediction of what the ranges should be based upon the receiver's estimate of position (and assumed knowledge of the world). The system conducts a blended comparison based upon the relative quality of the predicted navigation state and the observations.

In Figure 35.1, the arrow labeled "world model updates" indicates that the world model can be changed based upon the measurements that have been taken. Some navigation systems, particularly those which are designed and deployed specifically for navigation, do not require the end user of the system to be involved in this part of the process. For example, in GPS, the world model consists of information about the satellite orbits (ephemeris), the satellite clock errors, and details that are given in the signal specification (frequency, chipping rate, etc.). The GPS system uses its own receiver network on the ground to estimate satellite orbits and clock errors and to monitor the signals coming from space, and measurements from this network are used to continually update the GPS world model. As a result, the user simply obtains the most recent ephemeris and satellite clock terms and uses them for positioning. In this way, the user is completely uninvolved in the updating of the world model, which is helpful, because it greatly reduces the complexity of the system for the user.

Unlike man-made signals, natural signals do not generally have a dedicated part of the system that is continually updating a concise world model which describes how sensed measurements relate to the real world. As a result, the challenge with such systems is often to determine a usable world model. For example, it is very easy to obtain images of the nearby environment using a camera. However, in order to determine position and/or attitude from this kind of measurement, the user must have knowledge of what the world looks like as a function of position and attitude (the world model).

35.1.1 What Is a Navigation Sensor?

The physical sensor, depicted as the yellow block in Figure 35.1, is a critical part of any navigation system, and selection of the right sensor or combination of sensors is one of the most important decisions a navigation system designer can make. What comprises a navigation sensor?

At a basic level, any physical sensor that measures something which changes when the sensor is moved is a potential navigation sensor. Additionally, since clocks are an integral part of many navigation systems, we also consider clocks in this section as well. In contrast to a navigation sensor, which measures something that changes when the sensor is moved in some way, a clock is a sensor that measures how time "moves." A summary of the major sensors covered in Volume 2 is given in Table 35.1.

Table 35.1 Sensors covered in Volume 2

Sensor	Sensed phenomenon	World model required	Other considerations
Cellular RF receiver	Cellular phone RF signals	Positions of cell towers, signal timing	Example of signal of opportunity (SoOP), reference receiver sometimes required
Terrestrial beacon receiver	Navigation signals from terrestrial beacons	Beacon locations, signal structure, signal timing	Requires dedicated infrastructure, more design flexibility than SoOP
Digital TV receiver	Digital TV signals	Transmitter locations, signal timing	Example of SoOP, reference receiver sometimes required
Low-frequency receiver	Low-frequency RF signals	Transmitter location or direction of arrival, local distortion effects	Susceptible to local distortions, generally less accurate than higher frequency/wider bandwidth signals
Radar	RF signals	Locations of identifiable RF reflectors for absolute positioning	Generally larger/higher power than receiver-based systems
Low-Earth orbit (LEO) satellite receiver	Signals from LEO satellites	LEO satellite position/velocity, signal timing (in some cases), atmospheric models	Greater geometric/signal diversity and higher received power than GNSS
Inertial	Rotation and specific force	Gravitational field	Dead-reckoning only – drift normally requires update
GNSS	RF signals from satellites	Satellite ephemeris and clock errors, atmospheric models	Ideal for updating inertial
Magnetometer	Magnetic field (including variations)	Magnetic field map	Local (vehicle) effects calibration may be required
LiDAR	Range and intensity of laser returns	Shape/location of objects being sensed	Can be used in dead-reckoning or absolute modes
Cameras	Intensity of light as a function of direction	Map of image features or three-dimensional image model for absolute positioning	Can be used in dead-reckoning or absolute modes
X-ray detector	X-ray signals coming from pulsars	Knowledge of pulsar directions and signal characteristics (including timing)	Positioning is based on signal time of arrival
Clock	Varies by clock type	Perhaps calibration parameters	Measures rate of time passage (frequency), and if initialized and integrated, absolute time

35.2 Summary of Content of Volume 2

Volume 2 begins with an overview of nonlinear estimation techniques (Chapter 36), which are often required when integrating complementary navigation sensors. This chapter also lays the groundwork for the estimation strategies that are described in subsequent chapters.

The next group of chapters covers a variety of RF-based complementary navigation techniques. Many of the principles and algorithmic approaches for indoor navigation are summarized in Chapter 37, as well as a survey of different types of indoor navigation sensors and phenomenologies. This is followed by several chapters which describe in detail a variety of RF signals, including cellular (Chapter 38), terrestrial navigation beacons (Chapter 39), digital television (Chapter 40), low-frequency systems (Chapter 41), radar (Chapter 42), and RF signals from low-Earth orbiting (LEO) satellites (Chapter 43).

There are two chapters that describe *inertial technology*: a general introduction to INS (Chapter 44) and MEMS inertial systems (Chapter 45). The introduction chapter provides an overview of inertial systems. It describes the fundamental mechanisms of various accelerometers and gyroscopes that are the building blocks of INS, their error characteristics and performances, and outlook of technology advancement. The focus of MEMS inertial sensors is to reduce the cost, size, weight, and power when compared to existing inertial sensors. Doing so would expand the applications in which it is feasible to leverage inertial technology.

It is important to recognize that inertial systems cannot operate without aiding from additional sensors, other than for short time periods. The primary reason for this is that inertial systems are unstable in the vertical channel, so at a minimum they need some sort of aiding of the vertical channel (such as a barometric altimeter or terrain height aiding). Even if the vertical channel is aided, the horizontal directions will drift in an inertial system, with the rate of drift determined by the quality of the system and the accuracy of the initialization of the attitude and position of the system. (Even if an INS had perfect gyroscopes and accelerometers, there would still be growing error due to imperfections in our knowledge of gravity).

Probably the most common sensor used to aid an inertial is a *GNSS receiver*. Chapter 46 describes classic approaches for integrating GPS with INS, including loose and tight integration. It also describes a different way of thinking about the GPS/INS integration problem, in which there is more emphasis on using carrier-phase measurements to provide velocity-like updates to the INS, with additional correction from the pseudorange measurements.

Clock has been an essential sensor for navigation since ancient times. The accuracy and stability of clocks continue the improve in recent decades. Chapter 47 provides an overview of recent technology development in atomic clocks for GNSS.

An approach for using knowledge of the variation in Earth's magnetic field for absolute positioning using a *magnetometer* is described in Chapter 48. This method works indoors, on a ground vehicle, and in an aircraft, and this chapter describes the differences between these different environments and shows examples of working systems in each case.

Next, the use of *LiDAR* for navigation is described in Chapter 49. Various types of LiDARs are considered, as well as different ways in which LiDAR data can be leveraged for navigation purposes. This chapter also describes features that can be identified using LiDAR data, and how those features can be incorporated into an integrated navigation system. Both dead-reckoning and absolute positioning/attitude approaches are considered.

Chapter 50 describes the many ways in which cameras can be used for navigation. Initially, a mathematical model of a camera is provided, as well as methods for camera calibration. Image features are described as well as algorithms for using these features to relate camera images to position and rotation of the camera. Several methods for image navigation are described, and as with LiDAR, both dead-reckoning and absolute positioning/attitude approaches are considered. Another chapter (51) is dedicated to the topic of photogrammetry, which also uses a camera, but lays more emphasis on using the camera in order to develop knowledge about the scene that is viewed by one or more cameras. The vision navigation and photogrammetry chapters can be thought of as opposite sides of the same coin. With vision navigation, the desire is to figure out where the camera is, based on some knowledge of the scene. With photogrammetry, the desire is to figure out information about the scene, based on some knowledge of the camera position (and perhaps orientation).

As mentioned earlier in this introductory chapter, any measurement that changes when the sensor position changes can potentially be used as a navigation source. A good example of this is X-ray pulsar-based navigation which is described in Chapter 52, along with other variable celestial sources for navigation. The fundamental premise here is that if we can accurately measure the time of arrive of the periodic signal coming from several X-ray-emitting pulsars, we can use this information to determine our location. Additionally, methods for performing X-ray pulsar-based attitude determination are given.

In contrast to all of the technology-based approaches describe thus far, Chapter 53 focuses on brain neural processing in order to perform various navigation tasks. While these neurological approaches are quite difference from the approaches that engineers normally take to develop navigation systems, the way in which navigation is done by the brain suggest possibilities that we can attempt to implement with various forms of computing. Chapter 54 further describes various ways in which animals are able to navigate and orient without the use of the modern sensors described elsewhere in this volume.

Volume 2 then concludes with several chapters that describe specific applications that make heavy use of navigation systems. Many of these applications did not exist prior to the arrival of GNSS, and those that did exist have seen large increases in capabilities by leveraging both GNSS and complementary navigation approaches.

The applications covered include survey and mobile mapping (Chapter 55), precision agriculture (Chapter 56), wearable navigation technology (Chapter 57), driverless vehicles (Chapter 58), train control (Chapter 59), unmanned aerial systems (Chapter 60), aviation (Chapter 61), spacecraft navigation and orbit determination (Chapter 62), spacecraft formation flying and rendezvous (Chapter 63), and finally Arctic navigation (Chapter 64).

Taken together, Volume 2 shows the incredible value of navigation systems and the variety of approaches that are available in cases where GNSS is not sufficient. Whether we realize it or not, our day-to-day lives are heavily dependent on the ability of many systems that interact with (or that are behind the scenes) to determine time and position, and there is an increasing number of creative options and opportunities for precise navigation and time that can meet the needs of current and future applications.

36

Nonlinear Recursive Estimation for Integrated Navigation Systems

Michael J. Veth

Veth Research Associates, United States

36.1 Introduction

Almost immediately following its introduction in 1960, the Kalman filter and the extended Kalman filter have served as the primary algorithms used to solve navigation problems [1–3]. The optimal, recursive, and online characteristics of the algorithm are perfectly suited to serve a wide range of applications requiring real-time navigation solutions.

The traditional Kalman filter and extended Kalman filter are based on the following assumptions:

- Linear (or nearly linear) system dynamics and observations.
- All noise and error sources are Gaussian.

While these assumptions are valid in many cases, there is increasing interest in incorporating sensors and systems that are non-Gaussian, nonlinear, or both. Because these characteristics inherently violate the fundamental assumptions of the Kalman filter, when Kalman filters are used, performance suffers. More specifically, this can result in filter estimates that are inaccurate, inconsistent, or unstable. To address this limitation, researchers have developed a number of algorithms designed to provide improved performance for nonlinear and non-Gaussian problems [4–6].

In this chapter, we provide an overview of some of the most common and useful classes of nonlinear recursive estimators. The goal is to introduce the fundamental theories supporting the algorithms, identify their associated performance characteristics, and finally present their respective applicability from a navigation perspective.

The chapter is organized as follows. First, an overview of the notation and essential concepts related to estimation and probability theory are presented as a foundation for nonlinear filtering development. Some of the concepts include recursive estimation frameworks, the implicit assumptions and limitations of traditional estimators, and the deleterious effects on performance when these assumptions are not satisfied. Next an overview of nonlinear estimation theory is presented with the goal of demonstrating and deriving three main classes of nonlinear recursive estimators. These include Gaussian sum filters, grid particle filters, and sampling particle filters. Each of these classes of nonlinear recursive estimators is demonstrated and evaluated using a simple navigation example. The chapter is concluded with a discussion regarding the strengths and weaknesses of the approaches discussed with an emphasis on helping navigation engineers decide which estimation algorithm to apply to a given problem of interest.

36.1.1 Notation

The following notation is used in this chapter:

- State vector: The state vector at time k is represented by the vector $\mathbf{x}_k$.
- State estimate: An estimated quantity is represented using the hat operator. For example, the estimated state vector at time k is $\hat{\mathbf{x}}_k$.
- A priori/a posteriori estimates: A priori and a posteriori estimates are represented using the + and – superscript notation. For example, the a priori state estimate at time k is $\hat{\mathbf{x}}_k^-$, and the a posteriori state estimate at time k is $\hat{\mathbf{x}}_k^+$.
- State error covariance estimates: The state error covariance matrix is represented using the matrix $\mathbf{P}$ with superscripts and subscripts as required. For example, the a priori state error covariance matrix at time k is given by $\mathbf{P}_k^-$.
- State transition matrix: The state transition matrix from time $k-1$ to k is given by $\mathbf{\Phi}_{k-1}^k$. Note that the time indices may be omitted when they are explained contextually.

Position, Navigation, and Timing Technologies in the 21st Century: Integrated Satellite Navigation, Sensor Systems, and Civil Applications, Volume 2, First Edition. Edited by Y. T. Jade Morton, Frank van Diggelen, James J. Spilker Jr., and Bradford W. Parkinson.

- Process noise vector and covariance: The process noise vector at time k is $\mathbf{w}_k$. The process noise covariance matrix at time k is $\mathbf{Q}_k$.
- Observation vector: The observation vector at time k is given by $\mathbf{z}_k$.
- Observation influence matrix: The observation influence matrix at time k is given by $\mathbf{H}_k$. Note that the time index may be omitted when contextually unnecessary.
- Measurement noise vector and covariance: The measurement noise vector at time k is represented by $\mathbf{v}_k$. The measurement noise covariance is represented by $\mathbf{R}_k$.
- Probability density function: Probability density functions are expressed as $p(\cdot)$.

36.2 Linear Estimation Foundations

The goal of any estimator is to estimate one (or more) parameters of interest based on a model of the system, observations from sensors, or both. Because the parameters are, by definition, random vectors, they can be completely characterized by their associated probability density function (pdf). If we define our parameter vector and observation vectors at time k as $\mathbf{x}_k$ and $\mathbf{z}_k$, respectively, the overarching objective of a recursive estimator is to estimate the pdf of all of the previous state vector epochs, conditioned on all observations received up to the current epoch. Mathematically, this is expressed as the following pdf:

$$p(\mathbb{X}_k|\mathbb{Z}_k) \tag{36.1}$$

where

$$\mathbb{X}_k \triangleq \{\mathbf{x}_0, \mathbf{x}_1, \cdots, \mathbf{x}_k\} \tag{36.2}$$

and

$$\mathbb{Z}_k \triangleq \{\mathbf{z}_0, \mathbf{z}_1, \cdots, \mathbf{z}_k\} \tag{36.3}$$

While this is the most general case, it should be noted that most online algorithms would only be concerned with the conditional state estimate at the current epoch. For this situation, Eq. 36.1 would be represented as

$$p(\mathbf{x}_k|\mathbb{Z}_k) \tag{36.4}$$

In the next section, we will present the typical recursive estimation framework which will serve as the foundations for developing the forthcoming nonlinear recursive estimation strategies to follow.

36.2.1 Typical Recursive Estimation Framework

In a typical recursive estimation framework, the system is represented using a process model and one (or more) observation models. The process model represents the internal dynamics of the system and can be expressed as a nonlinear, stochastic difference equation of the form

$$\mathbf{x}_k = f(\mathbf{x}_{k-1}, \mathbf{w}_{k-1}) \tag{36.5}$$

where $\mathbf{x}_k$ is the state vector at time $k \in \mathbb{N}$, and $\mathbf{w}_{k-1}$ is the process noise random vector at time $k - 1$. External observations regarding the system state are represented by an observation model. The generalized observation model is a function of both the system state and a random vector representing the observation errors:

$$\mathbf{z}_k = h(\mathbf{x}_k, \mathbf{v}_k) \tag{36.6}$$

In the above equation, $\mathbf{z}_k$ is the observation at time k, and $\mathbf{v}_k$ is the random observation error vector at time k. The objective of the recursive estimator is to estimate the posterior pdf of the state vector, conditioned on the observations

$$p(\mathbf{x}_k|\mathbb{Z}_k) \tag{36.7}$$

where $\mathbb{Z}_k$ is the collection of observations up to, and including, time k. This is accomplished by performing two types of transformations on the state pdf, propagation and updates. The result is a filter cycle given by

$$p(\mathbf{x}_{k-1}|\mathbb{Z}_{k-1}) \xrightarrow[\text{propagate}]{} p(\mathbf{x}_k|\mathbb{Z}_{k-1}) \xrightarrow[\text{update}]{} p(\mathbf{x}_k|\mathbb{Z}_k) \tag{36.8}$$

Note the introduction of the a priori pdf given by

$$p(\mathbf{x}_k|\mathbb{Z}_{k-1}) \tag{36.9}$$

Further examination of the propagation and update cycle in Eq. 36.8 provides insights into how our system knowledge and observations are incorporated into our understanding of the state vector. To begin, we consider the propagation step from epoch $k - 1$ to k. Time propagation begins with the posterior pdf $p(\mathbf{x}_{k-1}|\mathbb{Z}_{k-1})$. The process model defined in Eq. 36.5 is used to define the transition pdf $p(\mathbf{x}_k|\mathbf{x}_{k-1})$, which can then be used to calculate the a priori pdf at time k via the Chapman–Kolmogorov equation [2]:

$$p(\mathbf{x}_k|\mathbb{Z}_{k-1}) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbb{Z}_{k-1})p(\mathbf{x}_{k-1}|\mathbb{Z}_{k-1})d\mathbf{x}_{k-1} \tag{36.10}$$

Examination of the process model (Eq. 36.5) shows that the propagated state vector is a first-order Gauss–Markov random process and is dependent only on the previous state vector and the process noise vector. As a result, we can express the transition probability, which is independent of the prior observation, as

$$p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbb{Z}_{k-1}) = p(\mathbf{x}_k|\mathbf{x}_{k-1}) \tag{36.11}$$

Substituting Eqs. 36.11 into 36.10 results in the propagation relationship

$$p(\mathbf{x}_k|\mathbb{Z}_{k-1}) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbb{Z}_{k-1})d\mathbf{x}_{k-1} \tag{36.12}$$

An observation at time k can be incorporated by considering the posterior pdf $p(\mathbf{x}_k|\mathbb{Z}_k)$, which, given the definition of our observation sequence in Eq. 36.3, can be expressed equivalently as

$$p(\mathbf{x}_k|\mathbb{Z}_k) = p(\mathbf{x}_k|\mathbf{z}_k, \mathbb{Z}_{k-1}) \tag{36.13}$$

Applying Bayes' rule to Eq. 36.13 yields

$$p(\mathbf{x}_k|\mathbb{Z}_k) = \frac{p(\mathbf{z}_k|\mathbf{x}_k, \mathbb{Z}_{k-1})p(\mathbf{x}_k|\mathbb{Z}_{k-1})}{p(\mathbf{z}_k|\mathbb{Z}_{k-1})} \tag{36.14}$$

Observing the form of the previously defined observation, Eq. 36.6 shows that $\mathbf{z}_k$ is independent of $\mathbb{Z}_{k-1}$, and thus Eq. 36.14 can be simplified to

$$p(\mathbf{x}_k|\mathbb{Z}_k) = \frac{p(\mathbf{z}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbb{Z}_{k-1})}{p(\mathbf{z}_k|\mathbb{Z}_{k-1})} \tag{36.15}$$

As a final note, we observe that the normalizing term in the denominator, known as the evidence, can be expressed in a more directly obvious form by de-marginalizing about the state vector as follows:

$$p(\mathbf{x}_k|\mathbb{Z}_k) = \frac{p(\mathbf{z}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbb{Z}_{k-1})}{\int p(\mathbf{z}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbb{Z}_{k-1})d\mathbf{x}_k} \tag{36.16}$$

Thus, we have presented the mathematical form of both the propagation (Eq. 36.12) and update (Eq. 36.16) actions on the pdf representing the state random vector.

For a specific class of problems (e.g. linear Gaussian systems), the above equations can be solved in closed form. In this case, the generalized process model (Eq. 36.5) simplifies to

$$\mathbf{x}_k = \mathbf{\Phi}_{k-1}^{k}\mathbf{x}_{k-1} + \mathbf{w}_{k-1} \tag{36.17}$$

where $\mathbf{\Phi}_{k-1}^{k}$ is the state transition matrix from time $k-1$ to k, and $\mathbf{w}_{k-1}$ is a zero-mean, white Gaussian sequence with covariance $\mathbf{Q}_k$. Similarly, the generalized observation model 36.6 simplifies to

$$\mathbf{z}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k \tag{36.18}$$

where $\mathbf{H}_k$ is the observation influence matrix at time k, and $\mathbf{v}_k$ is a zero-mean, white Gaussian sequence with covariance $\mathbf{R}_k$.

Thus, both the a priori and posterior pdfs can be represented as the following Gaussian densities, respectively:

$$p(\mathbf{x}_k|\mathbb{Z}_{k-1}) \triangleq \mathcal{N}\left(\hat{\mathbf{x}}_k^-, \mathbf{P}_k^-\right) \tag{36.19}$$

$$p(\mathbf{x}_k|\mathbb{Z}_k) \triangleq \mathcal{N}\left(\hat{\mathbf{x}}_k^+, \mathbf{P}_k^+\right) \tag{36.20}$$

where $\mathcal{N}(\mu, \Lambda)$ represents a Gaussian density with μ mean and Λ covariance. In addition, the plus and minus superscripts are used to express an a priori or a posteriori quantity, respectively. Substituting the linear process model (Eq. 36.17) into our propagation relationship (Eq. 36.12) results in the linear Kalman filter propagation equations

$$\hat{\mathbf{x}}_k^- = \mathbf{\Phi}_{k-1}^{k}\hat{\mathbf{x}}_{k-1}^+ \tag{36.21}$$

$$\mathbf{P}_k^- = \mathbf{\Phi}_{k-1}^{k}\mathbf{P}_{k-1}^+\left(\mathbf{\Phi}_{k-1}^{k}\right)^T + \mathbf{Q}_{k-1} \tag{36.22}$$

Furthermore, substituting the linear observation model (Eq. 36.18) into our update relationship (Eq. 36.16) results in the linear Kalman filter update equations:

$$\hat{\mathbf{x}}_k^+ = \hat{\mathbf{x}}_k^- + \mathbf{K}_k\left(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_k^-\right) \tag{36.23}$$

$$\mathbf{P}_k^+ = \mathbf{P}_k^- - \mathbf{K}_k\mathbf{H}_k\mathbf{P}_k^- \tag{36.24}$$

where $\mathbf{z}_k$ is the realized measurement observation, and $\mathbf{K}_k$ is the Kalman gain at time k:

$$\mathbf{K}_k = \mathbf{P}_k^-\mathbf{H}_k^T\mathbf{S}_k^{-1} \tag{36.25}$$

and $\mathbf{S}_k$ is the residual covariance matrix, given by

$$\mathbf{S}_k = \mathbf{H}\mathbf{P}_k^-\mathbf{H}^T + \mathbf{R}_k \tag{36.26}$$

In many cases, systems can be accurately represented by linear Gaussian models. Unfortunately, there are a number of systems where these models are not adequate. This motivates the development of various algorithms that attempt to solve these equations for various classes of problems.

In the next section, we will present the fundamental concepts which will be used to derive various recursive nonlinear estimators.

36.3 Nonlinear Filtering Concepts

In the previous sections (Section 36.2.1), we have developed the generalized theory for recursive estimation problems. The theory is based on the fundamental need to determine the pdf of the state vector at an epoch of interest, conditioned on the observations up to, and including, the current epoch. Complete knowledge of the conditional state pdf represents maximum possible knowledge of the system. This is, in fact, the normal state of affairs for Gaussian systems, as the pdf can be completely described by a mean and covariance.

36.3.1 Effects of Nonlinear Operations on Random Processes – Breaking Up with Gauss

Consider a Gaussian random vector $\mathbf{x}$ with mean and covariance $\hat{\mathbf{x}}$ and $\mathbf{P}_x$, respectively. As mentioned previously, for Gaussian densities, these two parameters are sufficient to completely describe the full pdf of the random vector.

Next consider a linear transformation from $\mathbf{x}$ to $\mathbf{y}$ which is governed by the transformation matrix $\mathbf{H}$. The resulting equation for $\mathbf{y}$ is given by

$$\mathbf{y} = \mathbf{H}\mathbf{x} \tag{36.27}$$

In this case, the transformed random vector, $\mathbf{y}$, can be shown to be a Gaussian random vector with mean and covariance

$$\hat{\mathbf{y}} = \mathbf{H}\hat{\mathbf{x}} \tag{36.28}$$

$$\mathbf{P}_y = \mathbf{H}\mathbf{P}_x\mathbf{H}^T \tag{36.29}$$

This preservation of Gaussian nature when transformed via linear operations is an important property of Gaussian densities that makes the linear Kalman filter relatively simple to implement.

Now consider a generalized nonlinear transformation

$$\mathbf{y} = \mathbf{h}(\mathbf{x}) \tag{36.30}$$

In this case, the density of $\mathbf{y}$ can become difficult to calculate exactly. While we will address this issue in more detail later in the chapter, generally speaking, the resulting density function is clearly non-Gaussian, thus limiting the performance of the linear Kalman filter algorithm. Nonlinear estimators attempt to maintain a higher-fidelity estimate of the overall density function as it transforms over time.

In the next section, we present our first class of estimators designed to support systems with non-Gaussian pdfs.

36.3.2 Gaussian Sum Filters

One approach for modeling systems with non-Gaussian pdfs is the use of composite random variables expressed as a sum of Gaussian random variables. The generalized Gaussian sum can be expressed as

$$\mathbf{x} = \sum_{j=1}^{J} w^{[j]}\mathbf{y}_j \tag{36.31}$$

where $w^{[j]}$ is a scalar weighting factor, $\mathbf{y}_j$ is a Gaussian random variable with mean $\hat{\mathbf{y}}_j$, and covariance $\mathbf{P}_{y_j}$. These individually weighted Gaussian random variables can represent the overall distribution of the state vector. An example of a density function created using a sum of Gaussian random variables is shown in Figure 36.1.

36.3.2.1 Multiple Model Adaptive Estimation

One implementation of the Gaussian sum filtering approach is known as multiple model adaptive estimation (MMAE). The MMAE filter uses a weighted Gaussian sum

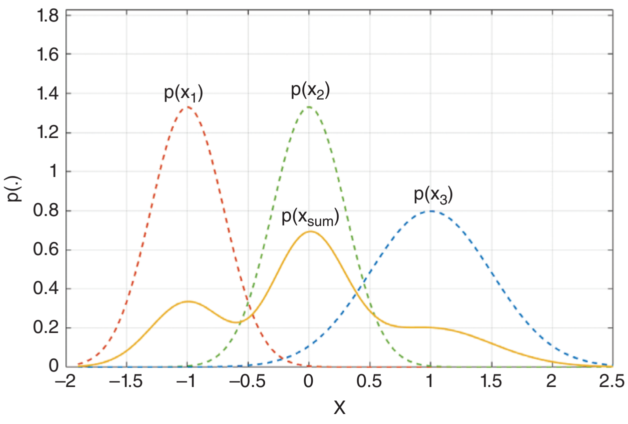

Figure 36.1 Gaussian sum illustration. The random variable x_{sum} is represented by a weighted sum of three individual Gaussian densities. In this example, $x_{sum} = 0.25x_1 + 0.5x_2 + 0.25x_3$.

to address the situation where unknown or uncertain parameters exist within the system model. Some examples of these types of situations include modeling discrete failure modes, unknown structural parameters, or processes with multiple discrete modes of operation (e.g. "jump" processes).

Consider our standard linear Gaussian process and observation models, repeated from Eqs. 36.17 and 36.18 for clarity:

$$\mathbf{x}_k = \mathbf{\Phi}_{k-1}^{k}\mathbf{x}_{k-1} + \mathbf{w}_{k-1} \tag{36.32}$$

$$\mathbf{z}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k \tag{36.33}$$

In the previous development, it was assumed that the system model parameters (i.e. $\mathbf{\Phi}_{k-1}^{k}$, $\mathbf{Q}_{k-1}$, $\mathbf{H}_k$, $\mathbf{R}_k$) were known. Let us now consider the situation where some of the system model parameters are unknown.

To address this situation, we can define a vector of the unknown system parameters, $\mathbf{a}$, and jointly estimate these parameters along with the state vector. In other words, we must now solve for the following density:

$$p(\mathbf{x}_k, \mathbf{a}|\mathbb{Z}_k) \tag{36.34}$$

which, after applying Bayes' rule, can be expressed as

$$p(\mathbf{x}_k, \mathbf{a}|\mathbb{Z}_k) = p(\mathbf{x}_k|\mathbf{a}, \mathbb{Z}_k)p(\mathbf{a}|\mathbb{Z}_k) \tag{36.35}$$

It is important to note that this expression is the product of the "known-system model" pdf, $p(\mathbf{x}_k|\,\mathbf{a}, \mathbb{Z}_k)$, and a new density function, $p(\mathbf{a}|\,\mathbb{Z}_k)$, which is the pdf of the unknown system parameters, conditioned on the observation set. Assuming $\mathbf{a} \in \mathbb{R}^n$, the parameter density can be written as

$$p(\mathbf{a}|\mathbb{Z}_k) = p(\mathbf{a}|\mathbf{z}_k, \mathbb{Z}_{k-1}) \tag{36.36}$$

Applying Bayes' rule yields

$$p(\mathbf{a}|\mathbb{Z}_k) = \frac{p(\mathbf{z}_k|\mathbf{a}, \mathbb{Z}_{k-1})p(\mathbf{a}|\mathbb{Z}_{k-1})}{p(\mathbf{z}_k|\mathbb{Z}_{k-1})} \tag{36.37}$$

Marginalizing the denominator about the parameter vector results in a more familiar form:

$$p(\mathbf{a}|\mathbb{Z}_k) = \frac{p(\mathbf{z}_k|\mathbf{a}, \mathbb{Z}_{k-1})p(\mathbf{a}|\mathbb{Z}_{k-1})}{\int p(\mathbf{z}_k|\mathbf{a}, \mathbb{Z}_{k-1})p(\mathbf{a}|\mathbb{Z}_{k-1})d\mathbf{a}} \tag{36.38}$$

where $p(\mathbf{z}_k|\ \mathbf{a}, \mathbb{Z}_{k-1})$ is the measurement prediction density, which, given our linear observation model, is expressed as the following normal distribution:

$$p(\mathbf{z}_k|\mathbf{a}, \mathbb{Z}_{k-1}) = \mathcal{N}\left(\mathbf{H}_k\hat{\mathbf{x}}_k^-, \mathbf{S}_k\right) \tag{36.39}$$

Unfortunately, the integral in the denominator is intractable in general, which requires an additional constraint. If the system parameters can be chosen from a finite set (e.g. $\mathbf{a} \in \{\mathbf{a}^{[1]}, \mathbf{a}^{[2]}, \cdots, \mathbf{a}^{[j]}\}$), the parameter density can be expressed as the sum of the individual probabilities of the finite set. This results in a system parameter pdf defined as

$$p(\mathbf{a}|\mathbb{Z}_{k-1}) = \sum_{j=1}^{J} w_{k-1}^{[j]}\delta\left(\mathbf{a} - \mathbf{a}^{[j]}\right) \tag{36.40}$$

where $w_{k-1}^{[j]}$ is the probability of the *j-th* parameter vector at time *k-1*, and $\delta(\cdot)$ is the delta function. It can be observed that the sum of the weights must be unity in order to represent a probability density. Substituting Eq. 36.40 into Eq. 36.38:

$$p(\mathbf{a}|\mathbb{Z}_k) = \frac{p(\mathbf{z}_k|\mathbf{a}, \mathbb{Z}_{k-1})\sum_{j=1}^{J} w_{k-1}^{[j]}\delta\left(\mathbf{a} - \mathbf{a}^{[j]}\right)}{\int p(\mathbf{z}_k|\mathbf{a}, \mathbb{Z}_{k-1})\sum_{n=1}^{J} w_{k-1}^{[n]}\delta(\mathbf{a} - \mathbf{a}^{[n]})d\mathbf{a}} \tag{36.41}$$

Moving the position of the summation operators and parameter weight vector:

$$p(\mathbf{a}|\mathbb{Z}_k) = \frac{\sum_{j=1}^{J} w_{k-1}^{[j]}p(\mathbf{z}_k|\mathbf{a}, \mathbb{Z}_{k-1})\delta\left(\mathbf{a} - \mathbf{a}^{[j]}\right)}{\sum_{n=1}^{J} w_{k-1}^{[n]}\int p(\mathbf{z}_k|\mathbf{a}, \mathbb{Z}_{k-1})\delta(\mathbf{a} - \mathbf{a}^{[n]})d\mathbf{a}} \tag{36.42}$$

The properties of the delta function can be exploited to rewrite the numerator and eliminate the integral from the denominator:

$$p(\mathbf{a}|\mathbb{Z}_k) = \sum_{j=1}^{J}\left(\frac{p\left(\mathbf{z}_k|\mathbf{a}^{[j]}, \mathbb{Z}_{k-1}\right)}{\sum_{n=1}^{J} w_{k-1}^{[n]}p(\mathbf{z}_k|\mathbf{a}^{[n]}, \mathbb{Z}_{k-1})}\right)w_{k-1}^{[j]}\delta\left(\mathbf{a} - \mathbf{a}^{[j]}\right) \tag{36.43}$$

At this point, we have established the posterior pdf of the parameter vector as a finite weighted set. Revisiting our system parameter pdf, now defined at time *k*

$$p(\mathbf{a}|\mathbb{Z}_k) = \sum_{j=1}^{J} w_k^{[j]}\delta\left(\mathbf{a} - \mathbf{a}^{[j]}\right) \tag{36.44}$$

and substituting into Eq. 36.43 yields the parameter density update relationship

$$\sum_{j=1}^{J} w_k^{[j]}\delta\left(\mathbf{a} - \mathbf{a}^{[j]}\right) = \sum_{j=1}^{J}\left(\frac{p\left(\mathbf{z}_k|\mathbf{a}^{[j]}, \mathbb{Z}_{k-1}\right)}{\sum_{n=1}^{J} w_{k-1}^{[n]}p(\mathbf{z}_k|\mathbf{a}^{[n]}, \mathbb{Z}_{k-1})}\right)w_{k-1}^{[j]}\delta\left(\mathbf{a} - \mathbf{a}^{[j]}\right) \tag{36.45}$$

In the above equation, the predicted measurement pdf, $p(\mathbf{z}_k|\ \mathbf{a}^{[j]}, \mathbb{Z}_{k-1})$, is evaluated at the measurement realization at time *k*, which yields the likelihood of realizing the current measurement, conditioned on the parameter set *j*. As mentioned previously, these likelihood values are based on the following evaluation of a normal density function:

$$p\left(\mathbf{z}_k = \mathbf{z}_k|\mathbf{a}^{[j]}, \mathbb{Z}_{k-1}\right) := \mathcal{N}\left(\mathbf{z}_k; \mathbf{H}_k\hat{\mathbf{x}}_k^{-[j]}, \mathbf{S}_k^{[j]}\right) \tag{36.46}$$

where $\mathbf{z}_k$ is the measurement realization at time *k*. This likelihood is equivalent to the likelihood of the residual from a Kalman filter tuned to the *j-th* parameter vector, $\mathbf{a}^{[j]}$.

Practically speaking, the parameter pdf consists of the discrete (fixed) parameter set and the associated weights (likelihood) at each epoch. The parameter density update shown in Eq. 36.45 shows the evolution of each parameter weight as a function of time, which can be rewritten as

$$w_k^{[j]} = \frac{p\left(\mathbf{z}_k = \mathbf{z}_k|\mathbf{a}^{[j]}, \mathbb{Z}_{k-1}\right)}{\sum_{n=1}^{J} w_{k-1}^{[n]}p(\mathbf{z}_k = \mathbf{z}_k|\mathbf{a}^{[n]}, \mathbb{Z}_{k-1})}w_{k-1}^{[j]}\forall j \in \{1, 2, \cdots, J\} \tag{36.47}$$

Our final task is to determine the overall posterior joint pdf of the system. Substituting Eq. 36.44 into Eq. 36.35, we obtain

$$p(\mathbf{x}_k, \mathbf{a}|\mathbb{Z}_k) = p(\mathbf{x}_k|\mathbf{a}, \mathbb{Z}_k)\sum_{j=1}^{J} w_k^{[j]}\delta\left(\mathbf{a} - \mathbf{a}^{[j]}\right) \tag{36.48}$$

which, when combined with knowledge of the delta function and implementing a straightforward rearrangement of terms produces the joint posterior density function

$$p(\mathbf{x}_k, \mathbf{a}|\mathbb{Z}_k) = \sum_{j=1}^{J} w_k^{[j]} p\left(\mathbf{x}_k|\mathbf{a}^{[j]}, \mathbb{Z}_k\right) \delta\left(\mathbf{a} - \mathbf{a}^{[j]}\right) \tag{36.49}$$

This pdf is clearly a weighted sum of Gaussian densities, each of these densities corresponding to the posterior state estimate of an individual Kalman filter, tuned to the parameter vector $\mathbf{a}^{[j]}$. The blended posterior state estimate and covariance are given by

$$\hat{\mathbf{x}}_k^+ = \sum_{j=1}^{J} w_k^{[j]} \hat{\mathbf{x}}_k^{+[j]} \tag{36.50}$$

$$\mathbf{P}_k^+ = \sum_{j=1}^{J} w_k^{[j]} \left[\left(\hat{\mathbf{x}}_k^{+[j]} - \hat{\mathbf{x}}_k^+ \right) \left(\hat{\mathbf{x}}_k^{+[j]} - \hat{\mathbf{x}}_k^+ \right)^T + \mathbf{P}_k^{+[j]} \right] \tag{36.51}$$

The MMAE filter can be visualized in block diagram form in Figure 36.2.

Additional forms that are very similar conceptually to the MMAE filter are known as interactive mixture model (IMM) estimators [8] and Rao-Blackwellized particle filters (RB-PFs) [9, 10], to name a few.

In the next section, we present a simple example to illustrate a potential application for Gaussian sum filters derived in this section.

36.3.3 MMAE Example – Integer Ambiguity Resolution

The benefits of the Gaussian sum filter can be illustrated using a simple example. Consider the following one-dimensional navigation scenario. A radio transmitter broadcasts a ranging signal from a fixed location, x_t. A ranging receiver is mounted on a vehicle that is free to move in the x-direction. The vehicle motion can be represented using a first-order Gauss–Markov velocity model [2] with uncertainty σ_v and time constant τ_v. The resulting state vector is given by

$$\mathbf{x}_k = \begin{bmatrix} p_k \\ v_k \end{bmatrix} \tag{36.52}$$

where p_k and v_k are the position and velocity of the vehicle at time k. The dynamics of the vehicle are given by

$$\mathbf{x}_{k+1} = \mathbf{\Phi}_k^{k+1} \mathbf{x}_k + \mathbf{w}_k \tag{36.53}$$

where

$$\mathbf{\Phi}_k^{k+1} = \begin{bmatrix} 1 & \Delta t \\ 0 & \exp(-\Delta t/\tau_v) \end{bmatrix} \tag{36.54}$$

and w_k is a zero-mean Gaussian random vector with

$$E\left[\mathrm{w}_j \mathrm{w}_k^T\right] = \begin{bmatrix} 0 & 0 \\ 0 & \sigma_v^2[1 - \exp(-2\Delta t/\tau_v)] \end{bmatrix} \delta_{jk} \tag{36.55}$$

The ranging signal consists of both a noise-corrupted measurement of the true range along with a measurement of the integrated carrier phase. The integrated carrier phase

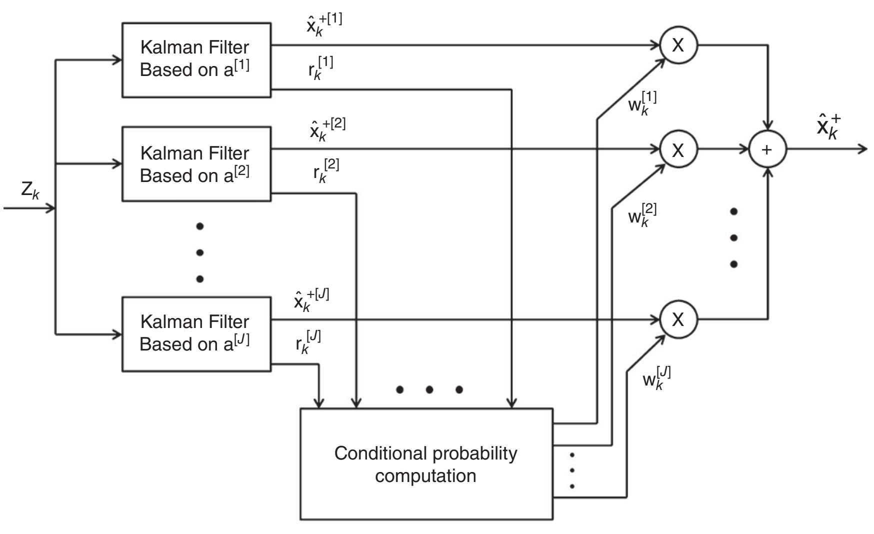

Figure 36.2 MMAE filter implementation. The MMAE filter constructs the state estimate by combining results from individual Kalman filters tuned to a parameter realization [7].

is a high-precision measurement, but is corrupted by an unknown integer ambiguity. The observation model is

$$\rho_k = x_k - x_t + \mathbf{v}_{\rho_k} \tag{36.56}$$

$$\phi_k = \lambda^{-1}(x_k - x_t) + N + \mathbf{v}_{\phi_k} \tag{36.57}$$

where λ is the carrier wavelength, and N is the integer ambiguity. Both observations are corrupted by zero-mean white Gaussian noise sequences with

$$E\left[\mathbf{v}_{\rho_j}\mathbf{v}_{\rho_k}\right] = \sigma_\rho^2\delta_{jk} \tag{36.58}$$

$$E\left[\mathbf{v}_{\phi_j}\mathbf{v}_{\phi_k}\right] = \sigma_\phi^2\delta_{jk} \tag{36.59}$$

$$E\left[\mathbf{v}_{\rho_j}\mathbf{v}_{\phi_k}\right] = 0 \tag{36.60}$$

Our goal is to use the MMAE estimator to accurately represent the (non-Gaussian) posterior pdf, thus maintaining a consistent overall state estimate and uncertainty, while incorporating all available information.

In this example, the integer ambiguity is the unknown parameter set, which in the previous development we designated as the vector $\mathbf{a}$. We choose a range of J plausible integers based upon any a priori knowledge or even the initial range observation itself, which results in the following unknown parameter vector:

$$\mathbf{a} = \left\{N^{[1]}, N^{[2]}, \cdots, N^{[J]}\right\} \tag{36.61}$$

with overall joint probability density

$$p(\mathbf{x}_k, \mathbf{a}|\mathbb{Z}_k) = \sum_{j=1}^{J} w_k^{[j]} p\left(\mathbf{x}_k|\mathbf{a}^{[j]}, \mathbb{Z}_k\right)\delta\left(\mathbf{a} - \mathbf{a}^{[j]}\right) \tag{36.62}$$

From this point forward, the implementation proceeds as outlined in the previous section. A total of J weighted Kalman filters are constructed, each with the assumption that $N^{[j]}$ is the correct integer ambiguity. The joint posterior density is given by

$$p(\mathbf{x}_k, \mathbf{a}|\mathbb{Z}_k) = \sum_{j=1}^{J} w_k^{[j]} p\left(\mathbf{x}_k|\mathbf{a}^{[j]}, \mathbb{Z}_k\right)\delta\left(\mathbf{a} - \mathbf{a}^{[j]}\right) \tag{36.63}$$

In order to demonstrate the performance of the Gaussian sum filter, the above scenario was implemented in a simulation environment. A trajectory and measurement set is randomly generated using the parameters specified in Table 36.1.

Note the carrier phase wavelength is 0.2 m, and the carrier phase measurement uncertainty is 0.1 cycles, which results in a measurement precision of 0.02 m, which is an improvement of 50 times over the pseudorange measurement errors.

Table 36.1 Simulation parameters

Parameter	Value	Units
σ_ρ	0.5	m
σ_ϕ	0.1	cycles
λ	0.2	m
σ_v	0.2	m/s
τ_v	500	s
x_t	0	m
Δt	1.0	s

The resulting trajectory, range observations, and phase observations are shown in Figure 36.3.

The MMAE global state estimate and density function of position after one observation ($t = 1$ s) are shown in Figure 36.4. The probability density function is clearly multi-modal, which accurately represents the range of solutions associated with the phase observation. As expected, the peaks are located at integer multiples of the carrier wavelength which corresponds to the most likely values of the unknown integer ambiguity. These peaks indirectly indicate the relative likelihood of the associated ambiguity being correct by exhibiting influence on the overall position density.

After 22 cycles, the position density shows a reduced number of peaks (see Figure 36.5). This indicates that the filter is incorporating sensor observations and the statistical dynamics model to effectively eliminate a number of potential ambiguity possibilities.

After 100 cycles (Figure 36.6), the filter has converged to a single ambiguity.

The global state estimate and associated standard deviation result for this simulation are shown in Figure 36.7. The shape of the uncertainty bound clearly shows the effects described above. As the likelihood of each integer ambiguity realization changes, the overall uncertainty changes and eventually collapses to the centimeter level.

Finally, the associated normalized filter weights for a subset of the integer ambiguity realizations are shown in Figure 36.8. As expected, the highly unlikely edge integers quickly collapse. The integers closer to the mean take longer to resolve. It is important to note that the resulting uncertainty is dependent on the *actual measurement realization sequence* received; thus, each realization would produce a different uncertainty (Table 36.2). This is a notable difference from the standard linear Kalman filter, where the uncertainty is independent of the observed measurements. Finally, it is important to note that, in this example, the state estimate and uncertainty of the MMAE filter are

Table 36.2 Summary of filter classes

Linear and extended Kalman filter		
Strengths • Optimal for linear Gaussian systems • Computationally simple	Weaknesses • Suboptimal approximation for nonlinear systems, can be prone to divergence	Use case • Linear, or close-to-linear, Gaussian problems
Gaussian sum filter		
Strengths • Optimal for linear Gaussian systems with discrete parameter vector	Weaknesses • If parameter vector is not discrete, the differences must be observable • Conservative tuning can mask difference between models and reduce performance • Increased computation requirements over simple Kalman filter	Use case • Linear, or close-to-linear, Gaussian problems with discrete parameters
Grid particle filter		
Strengths • Optimal solution when state space consists of discrete elements • Suitable for wide range of nonlinear conditions	Weaknesses • Computational requirements can be excessive • Processing requirements scale geometrically with the number of dimensions • Discretizing continuous state space results in suboptimal performance	Use case • Nonlinear problems with lower dimensionality
Sampling particle filter		
Strengths • Can produce nearly optimal solution for nonlinear problems • Computational requirements can be reduced over a grid particle filter via importance sampling strategies	Weaknesses • Maintaining good particle distribution can be difficult • Lack of repeatability from run to run • Computational requirements can still be large	Use case • Nonlinear problems with higher dimensionality

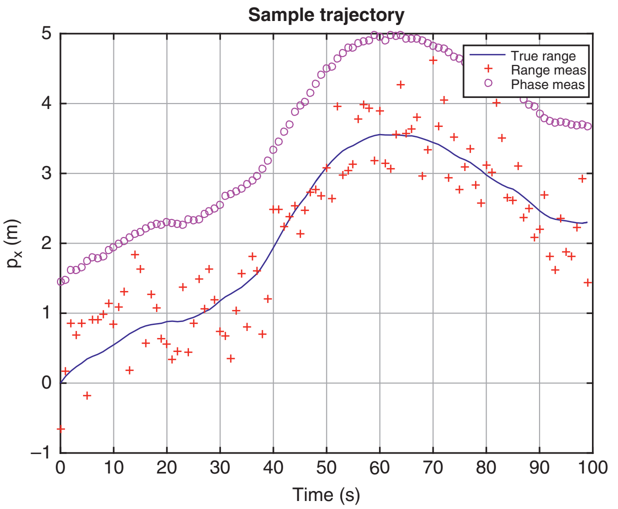

Figure 36.3 Sample vehicle trajectory and observations. Note that the range observations are accurate but not precise and the phase observations are precise but not accurate. Our goal is to accurately estimate the joint pdf of this system.

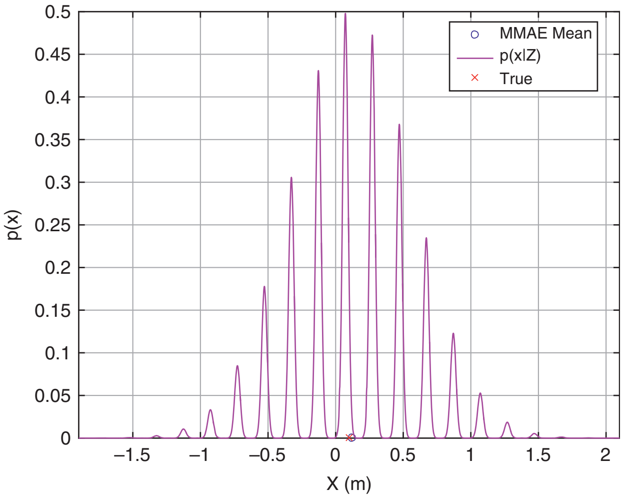

Figure 36.4 MMAE initial state estimate and position density function. Note the position density function is extremely multi-modal due to the limited information available at this point.

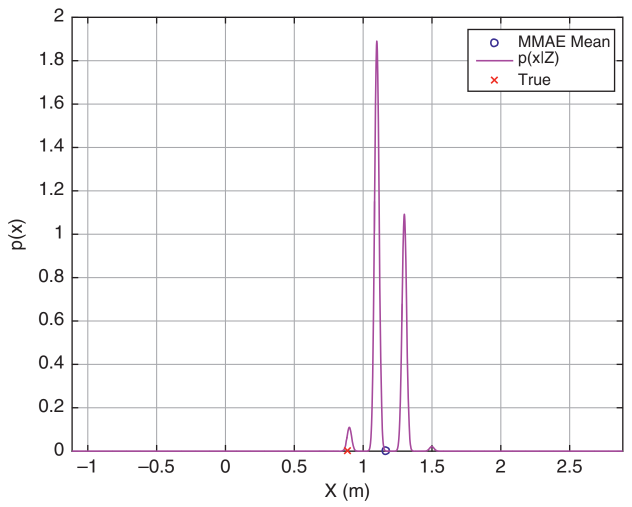

Figure 36.5 MMAE state estimate (after 22 observations). Range observations combined with the vehicle dynamics model are eliminating unlikely integer ambiguity values.

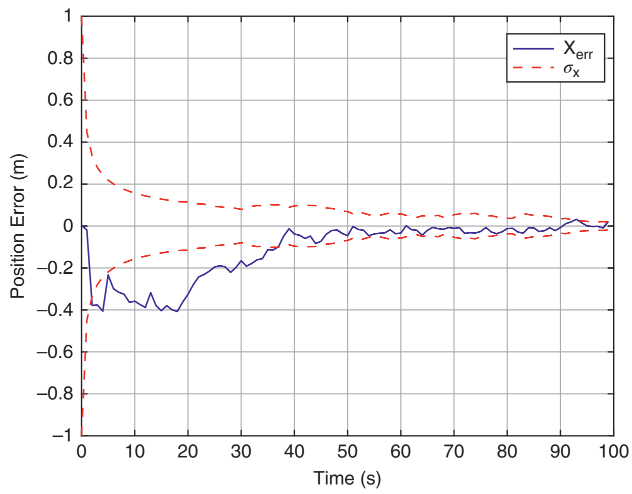

Figure 36.7 MMAE position error and one-sigma uncertainty. Note that the error uncertainty collapses once sufficient information is available to resolve the integer ambiguity.

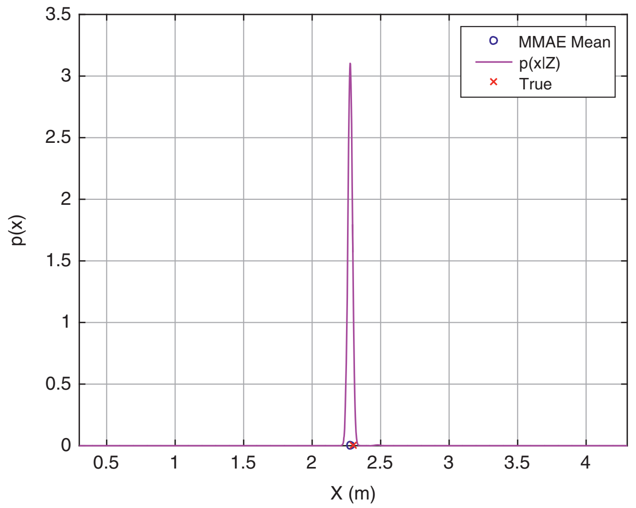

Figure 36.6 MMAE state estimate (after 100 observations). Note the state estimate is almost completely unimodal and has converged to the correct integer ambiguity.

truly optimal (i.e. minimum mean square error). This would not be the case if the integer ambiguity were resolved using a more traditional approach (e.g. float estimate with an ad hoc fixing stage). This is an interesting property of the Gaussian sum filter and sets the stage for us to investigate additional nonlinear estimation techniques.

36.3.4 Particle Filters

As mentioned in Section 36.3, the key requirement of a nonlinear filter is the ability to accurately represent arbitrary probability density functions. Particle filters accomplish this by representing density functions by using collections of discrete, weighted state vectors instances. These state vectors and associated weights are referred to as particles.

The development of the theory related to a particle-based representation of density functions begins by reviewing the essential properties of both the probability density function and the cumulative distribution function. An example cdf and pdf are shown in Figure 36.9. The cumulative distribution function is a monotonically increasing function which represents the probability of a random variable realization that is less than the operand and can be defined as the integral of the density function [11]:

$$Pr(\mathbf{x} < x_a) = F(x_a) \tag{36.64}$$

$$= \int_{-\infty}^{x_a} p(x)dx \tag{36.65}$$

Additionally, the probability of a random variable realization between a range x_a and x_b is expressed by

$$Pr(x_a \leq \mathbf{x} < x_b) = F(x_b) - F(x_a) \tag{36.66}$$

$$= \int_{x_a}^{x_b} p(x)dx \tag{36.67}$$

As a result, the density and cumulative distribution functions must have the following properties:

$$F(-\infty) = 0 \tag{36.68}$$

$$F(+\infty) = 1 \tag{36.69}$$

$$\int_{-\infty}^{\infty} f(x)dx = 1 \tag{36.70}$$

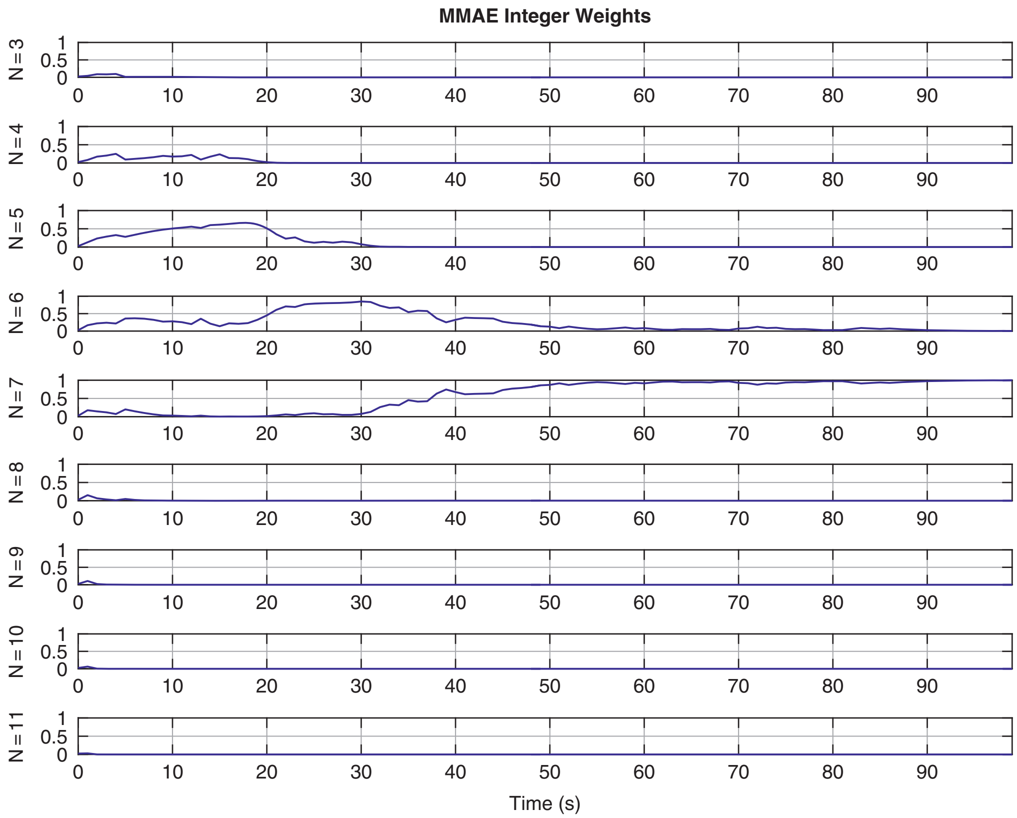

Figure 36.8 MMAE integer ambiguity particle weights (subset). The correct ambiguity particle (N=7) likelihood increases over time while the outliers are determined to be less likely.

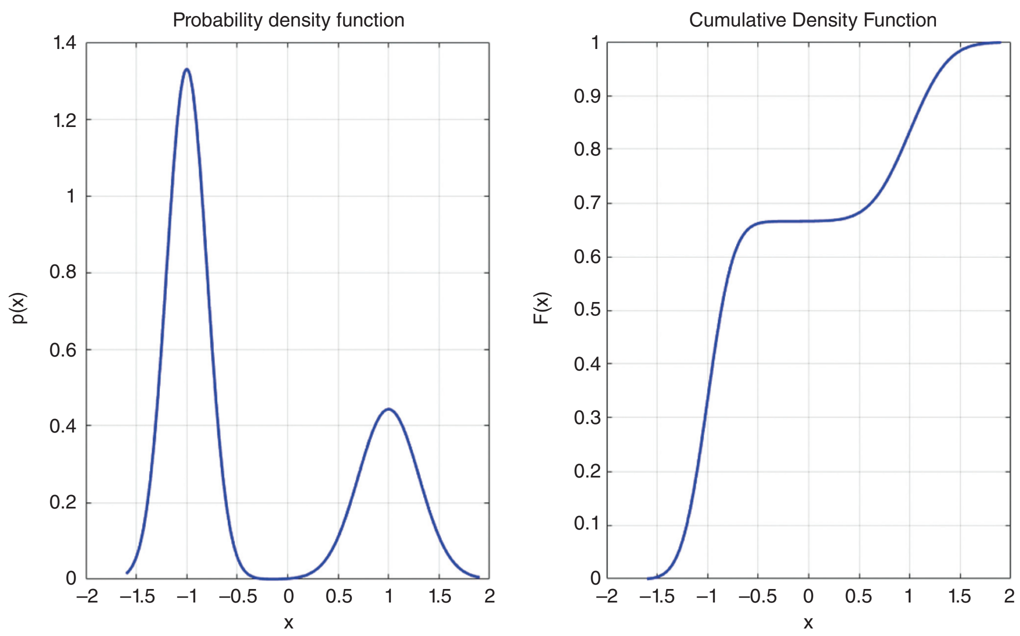

Figure 36.9 Probability density function (PDF) and cumulative density function example (CDF).

$$f(x) \geq 0 \tag{36.71}$$

The particle filter uses a collection of weighted delta functions to represent the pdf:

$$p(x) \approx \sum_{j=1}^{J} w^{[j]} \delta\left(x - x^{[j]}\right) \tag{36.72}$$

where $w^{[j]}$ is a scalar weighting value for the j-th particle with location $x^{[j]}$. As mentioned previously, the sum of weights must be unity:

$$\sum_{j=1}^{N} w^{[j]} = 1 \tag{36.73}$$

An example pdf represented by a collection of weighted particles is shown in Figure 36.10. This importance sampling strategy allows us to represent any pdf with a desired level of fidelity, given enough particles. Additional details regarding importance sampling are provided in Section 36.3.7.

In addition to representing arbitrary pdfs of random vectors, successful nonlinear estimation requires the ability to determine the resulting pdfs after applying nonlinear transformations to random vectors. In general, this can be intractable; however, representing the pdf using the collection of weighted particles makes the transformation relatively straightforward. An example of the effect of some sample nonlinear transformations is shown in Figure 36.11.

One of the most common functions necessary for filtering applications is calculation using the expectation operator. The expectation operator is defined as

$$E[g(\mathbf{x})] = \int_{-\infty}^{+\infty} g(\mathbf{x}) p(\mathbf{x}) d\mathbf{x} \tag{36.74}$$

where $E[\cdot]$ is the expectation operator, $g(\mathbf{x})$ is an arbitrary function of the random vector $\mathbf{x}$, and $p(\mathbf{x})$ is the pdf of the random vector.

Based on this definition, we can easily calculate some common expectations of the weighted particle pdf. The first is the mean, which is defined as $E[\mathbf{x}]$:

$$E[\mathbf{x}] = \int_{-\infty}^{+\infty} \mathbf{x} p(\mathbf{x}) d\mathbf{x} \tag{36.75}$$

Substituting the pdf of $\mathbf{x}$ from Eq. 36.72, rearranging the summation and integral, and then applying the sifting property:

$$E[\mathbf{x}] = \int_{-\infty}^{+\infty} \mathbf{x} \sum_{j=1}^{J} w^{[j]} \delta\left(\mathbf{x} - \mathbf{x}^{[j]}\right) d\mathbf{x} \tag{36.76}$$

$$= \sum_{j=1}^{J} w^{[j]} \int_{-\infty}^{+\infty} \mathbf{x} \delta\left(\mathbf{x} - \mathbf{x}^{[j]}\right) d\mathbf{x} \tag{36.77}$$

$$= \sum_{j=1}^{J} w^{[j]} \mathbf{x}^{[j]} \tag{36.78}$$

This shows that the mean of a weighted particle random variable can be calculated as the weighted sum of particles.

The above development can be applied identically to the general expectation function case with the following result:

$$E[g(\mathbf{x})] = \sum_{j=1}^{J} w^{[j]} g\left(\mathbf{x}^{[j]}\right) \tag{36.79}$$

This can easily be extended to represent a set of sufficient statistics for an arbitrary density function. As a result, it can be shown that any density function can be represented to arbitrary accuracy, given enough particles. Because we seek estimation methods that are computationally feasible, we are searching for methods that give us "good enough" performance (e.g. accuracy and stability) with limited computational resources.

In the next section, we investigate one approach, known as the grid particle filter, to representing the location of our particle collection.

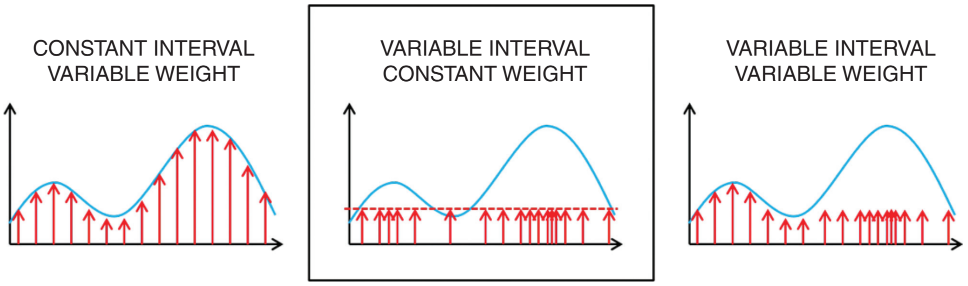

Figure 36.10 Importance sampling used to represent arbitrary density functions. The density function is represented by a combination of particle locations and weights (represented by arrows), which can be varied independently.

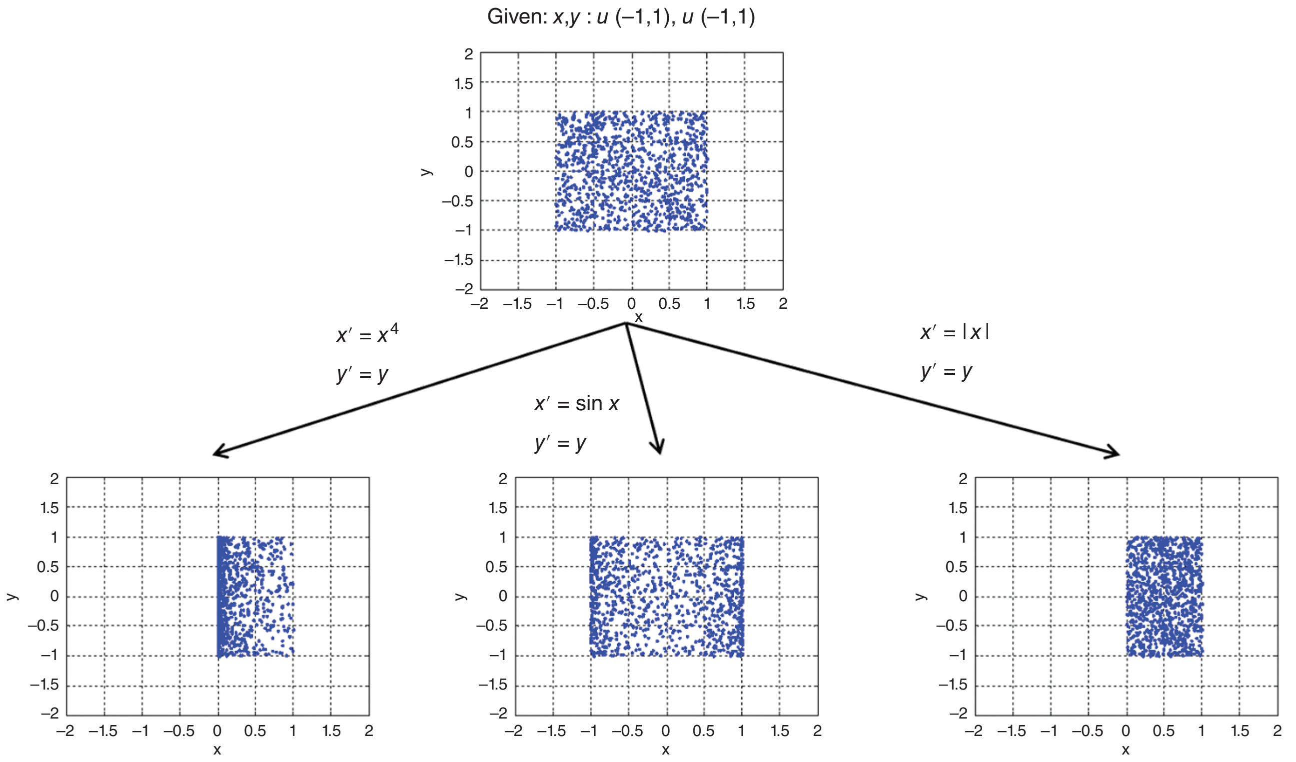

Figure 36.11 Visualization of nonlinear transformation on a random variable. Given uniform random variables x, y, the effects of three nonlinear transformations show that the density can change significantly during transformation.

36.3.5 Grid Particle Filtering

One approach to addressing the generalized nonlinear estimation requirement to maintain the full probability density is the so-called grid particle filter. The grid particle filter maintains a discrete collection of possible system states and associates a probability with each of those states (i.e. particles). This approach is optimal given systems with the following conditions:

1) The state vector is truly discrete or can be accurately approximated using a discretization of the state space.
2) The number of discrete states is computationally tractable.

Given these conditions, the state density function can be expressed as a weighted collection of particles (repeated from Eq. 36.72)

$$p(\mathbf{x}) = \sum_{j=1}^{J} w^{[j]}\delta\left(\mathbf{x}-\mathbf{x}^{[j]}\right) \tag{36.80}$$

where the particle weights, $w^{[j]}$, must sum to one. Because the particle locations are assumed to be static, the filtering operation is performed over the collection of weights. This allows the filter to maintain the density function as the collection of propagation and update steps are applied. At this point, it is relatively straightforward to derive the propagation and update relations for the collection of particles.

We begin with the propagation from time $k-1$ to k. Assume that the posterior density function at time $k-1$ is given by

$$p(\mathbf{x}_{k-1}|\mathbb{Z}_{k-1}) = \sum_{j=1}^{J} w_{k-1|k-1}^{[j]}\delta\left(\mathbf{x}_{k-1}-\mathbf{x}^{[j]}\right) \tag{36.81}$$

Substituting Eq. 36.81 into the Chapman–Kolmogorov equation (Eq. 36.12) and simplifying:

$$p(\mathbf{x}_k|\mathbb{Z}_{k-1}) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbb{Z}_{k-1})d\mathbf{x}_{k-1} \tag{36.82}$$

$$= \int p(\mathbf{x}_k|\mathbf{x}_{k-1}) \sum_{j=1}^{J} w_{k-1|k-1}^{[j]}\delta\left(\mathbf{x}_{k-1}-\mathbf{x}^{[j]}\right)d\mathbf{x}_{k-1} \tag{36.83}$$

$$= \sum_{j=1}^{J} w_{k-1|k-1}^{[j]} \int p(\mathbf{x}_k|\mathbf{x}_{k-1})\delta\left(\mathbf{x}_{k-1}-\mathbf{x}^{[j]}\right)d\mathbf{x}_{k-1} \tag{36.84}$$

$$= \sum_{j=1}^{J} w_{k-1|k-1}^{[j]} p\left(\mathbf{x}_k | \mathbf{x}^{[j]}\right) \tag{36.85}$$

The rightmost density in Eq. 36.85 is the transition probability function, which can be rewritten as

$$p\left(\mathbf{x}_k | \mathbf{x}^{[j]}\right) = \sum_{l=1}^{J} p\left(\mathbf{x}_k^{[l]} | \mathbf{x}_{k-1}^{[j]}\right) \delta\left(\mathbf{x}_k - \mathbf{x}^{[l]}\right) \tag{36.86}$$

Substituting Eq. 36.86 into Eq. 36.85 and simplifying

$$p(\mathbf{x}_k | \mathbb{Z}_{k-1}) = \sum_{j=1}^{J} w_{k-1|k-1}^{[j]} \sum_{l=1}^{J} p\left(\mathbf{x}^{[l]} | \mathbf{x}^{[j]}\right) \delta\left(\mathbf{x}_k - \mathbf{x}^{[l]}\right) \tag{36.87}$$

$$= \sum_{l=1}^{J} \left[\sum_{j=1}^{J} w_{k-1|k-1}^{[j]} p\left(\mathbf{x}^{[l]} | \mathbf{x}^{[j]}\right) \right] \delta\left(\mathbf{x}_k - \mathbf{x}^{[l]}\right) \tag{36.88}$$

$$= \sum_{l=1}^{J} w_{k|k-1}^{[l]} \delta\left(\mathbf{x}_k - \mathbf{x}^{[l]}\right) \tag{36.89}$$

where the new particle weight is given by

$$w_{k|k-1}^{[l]} = \sum_{j=1}^{J} w_{k-1|k-1}^{[j]} p\left(\mathbf{x}^{[l]} | \mathbf{x}^{[j]}\right) \tag{36.90}$$

Conceptually, this can be calculated as the sum of all posterior weights at time $k-1$ multiplied by the specific transition probability *into* state l from all possible prior states.

The development of the measurement update function proceeds in a similar fashion. Recalling our definition of the a priori density function (repeated from Eq. 36.89 for clarity):

$$p(\mathbf{x}_k | \mathbb{Z}_{k-1}) = \sum_{j=1}^{J} w_{k|k-1}^{[j]} \delta\left(\mathbf{x}_k - \mathbf{x}^{[j]}\right) \tag{36.91}$$

and substituting into the update equation (Eq. 36.16) yields

$$p(\mathbf{x}_k | \mathbb{Z}_k) = \frac{p(\mathbf{z}_k | \mathbf{x}_k) \sum_{j=1}^{J} w_{k|k-1}^{[j]} \delta\left(\mathbf{x}_k - \mathbf{x}^{[j]}\right)}{\int p(\mathbf{z}_k | \mathbf{x}_k) \sum_{j=1}^{J} w_{k|k-1}^{[j]} \delta(\mathbf{x}_k - \mathbf{x}^{[j]}) d\mathbf{x}_k} \tag{36.92}$$

which can be simplified by changing the order of integration and using the properties of the delta function

$$p(\mathbf{x}_k | \mathbb{Z}_k) = \frac{p(\mathbf{z}_k | \mathbf{x}_k) \sum_{j=1}^{J} w_{k|k-1}^{[j]} \delta\left(\mathbf{x}_k - \mathbf{x}^{[j]}\right)}{\int p(\mathbf{z}_k | \mathbf{x}_k) \sum_{l=1}^{J} w_{k|k-1}^{[l]} \delta(\mathbf{x}_k - \mathbf{x}^{[l]}) d\mathbf{x}_k} \tag{36.93}$$

$$= \frac{\sum_{j=1}^{J} w_{k|k-1}^{[j]} p(\mathbf{z}_k | \mathbf{x}_k) \delta\left(\mathbf{x}_k - \mathbf{x}^{[j]}\right)}{\sum_{l=1}^{J} w_{k|k-1}^{[l]} \int p(\mathbf{z}_k | \mathbf{x}_k) \delta(\mathbf{x}_k - \mathbf{x}^{[l]}) d\mathbf{x}_k} \tag{36.94}$$

$$= \frac{\sum_{j=1}^{J} w_{k|k-1}^{[j]} p\left(\mathbf{z}_k | \mathbf{x}^{[j]}\right) \delta\left(\mathbf{x}_k - \mathbf{x}^{[j]}\right)}{\sum_{l=1}^{J} w_{k|k-1}^{[l]} p(\mathbf{z}_k | \mathbf{x}^{[l]})} \tag{36.95}$$

Finally, recalling the grid particle filter form of the posterior density function

$$p(\mathbf{x}_k | \mathbb{Z}_k) = \sum_{j=1}^{J} w_{k|k}^{[j]} \delta\left(\mathbf{x}_k - \mathbf{x}^{[j]}\right) \tag{36.96}$$

and substituting into Eq. 36.95 yields

$$\begin{aligned} &\sum_{j=1}^{J} \left[w_{k|k}^{[j]} \right] \delta\left(\mathbf{x}_k - \mathbf{x}^{[j]}\right) \\ &= \sum_{j=1}^{J} \left[\frac{w_{k|k-1}^{[j]} p\left(\mathbf{z}_k | \mathbf{x}^{[j]}\right)}{\sum_{l=1}^{J} w_{k|k-1}^{[l]} p(\mathbf{z}_k | \mathbf{x}^{[l]})} \right] \delta\left(\mathbf{x}_k - \mathbf{x}^{[j]}\right) \end{aligned} \tag{36.97}$$

The bracketed areas show the final particle weight update equations

$$w_{k|k}^{[j]} = \frac{w_{k|k-1}^{[j]} p\left(\mathbf{z}_k | \mathbf{x}^{[j]}\right)}{\sum_{l=1}^{J} w_{k|k-1}^{[l]} p(\mathbf{z}_k | \mathbf{x}^{[l]})} \tag{36.98}$$

In the next section, we illustrate a potential application of the grid particle filter in a navigation context.

36.3.6 Grid Particle Filter Example Application

We return to the example presented in Section 36.3.3; however, in this case, we utilize a grid particle filter solution. The first step in the process is to determine the composition of the grid. In this case, there are two parameters we would like to estimate, position and velocity. Both of the parameters are continuous random variables, so we must quantize both of the parameters.

For this example, we are interested in centimeter-level positioning accuracy; thus, we divide the domain into 5 mm by 20 mm/s grids. For simplicity, we build a grid that is ±2 m in range and ±0.6 m/s in velocity. The absolute grid location is periodically adjusted based on the current estimated position and velocity of the vehicle.

An identical randomly generated trajectory and measurement set from the MMAE example (Section 36.3.3) is used

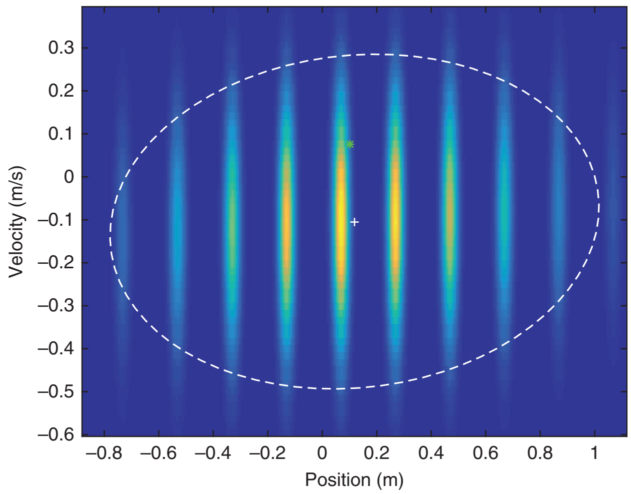

Figure 36.12 Grid particle filter state estimate and position density function after one observation. Note the density function is extremely multi-modal due to the limited information available at this point.

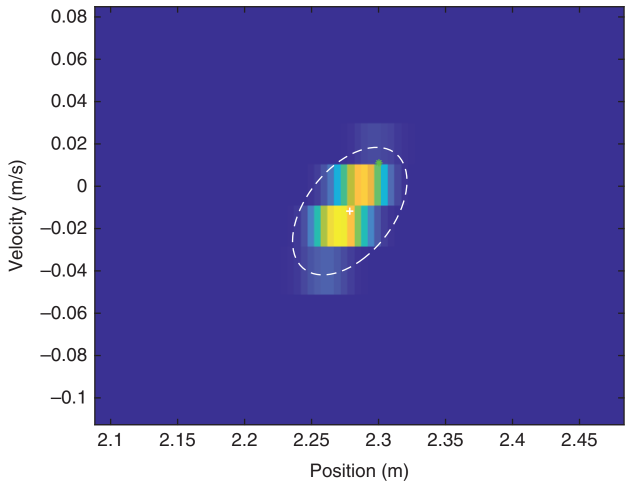

Figure 36.14 Grid particle filter state estimate (after 100 observations). Note that the state estimate is almost completely unimodal and has converged to the correct integer ambiguity.

as the inputs to the grid particle filter. For reference, the system parameters are specified in Table 36.1, and the resulting trajectory, range observations, and phase observations are shown in Figure 36.3.

The grid particle filter global state estimate and density function of position after one observation ($t = 1$ s) are shown in Figure 36.12. In this case, we present the probability density function using a two-dimensional array (position vs. velocity) of probabilities. The resulting pdf is clearly multi-modal, which accurately represents the range of solutions associated with the phase observation. As expected, the peaks are located as a function of the wavelength and represent the most likely values of integer ambiguity. These peaks indirectly indicate the relative likelihood of the associated ambiguity being correct by exhibiting influence on the overall position density. In each plot below, the calculated mean is represented by a white "plus," the true state is represented by a green asterisk, and the calculated 2-sigma uncertainty is represented as a white ellipse.

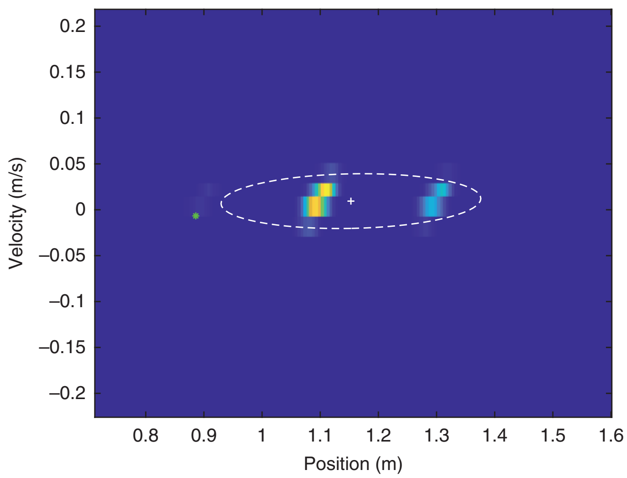

Figure 36.13 Grid particle filter state estimate (after 22 observations). Range observations combined with the vehicle dynamics model are eliminating unlikely integer ambiguity values.

After 22 cycles, the density shows a reduced number of peaks (see Figure 36.13). This indicates that the filter is incorporating sensor observations and the statistical dynamics model to effectively eliminate a number of potential ambiguity possibilities.

After 100 cycles (Figure 36.14), the filter has converged to a single ambiguity.

The global state estimate and associated standard deviation result for this simulation are shown in Figure 36.15. The shape of the uncertainty bound clearly shows the effects described above. As the likelihood of each integer ambiguity realization changes, the overall uncertainty changes and eventually collapses to the centimeter level.

In the next section, we will move to our final nonlinear filter algorithm, the sampling particle filter.

36.3.7 Sampling Particle Filter (SIS/SIR)

In a similar manner to the grid particle filter, the sampling particle filter, also known as a the sequential Monte Carlo (SMC) filter, represents the state density function using a weighted collection of particles. However, we seek to address the computational scaling problems inherent in grid-based approaches by exploiting an approach that

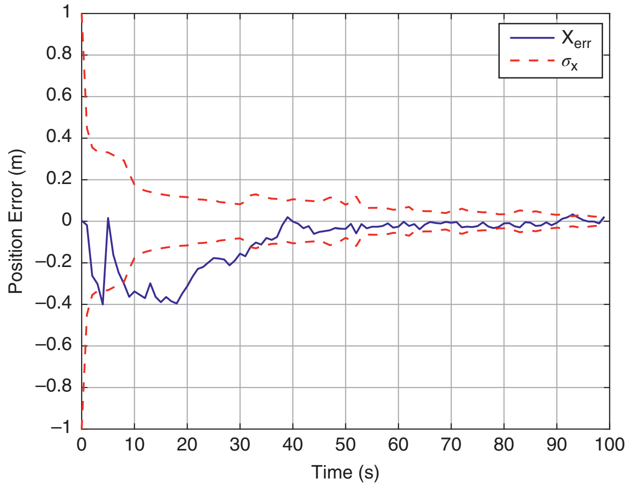

Figure 36.15 Grid particle filter position error and one-sigma uncertainty. Note that the error uncertainty collapses once sufficient information is available to resolve the integer ambiguity.

focuses computation on the regions of the state space with the highest likelihood. This is accomplished by randomly sampling the state space.

The main advantage of this approach is the potential to more completely sample the important areas of the state space, while limiting the total number of particles required. This is a useful advantage over the grid particle filter, which can require unreasonable numbers of particles as the state dimensionality and domain increase. While sampling particle filtering approaches are suboptimal, their computational advantages make them attractive for a larger range of applications.

We begin by describing the concept of Monte Carlo integration, which is subsequently used to develop a basic recursive estimation algorithm.

The fundamental enabling concept for the sampling particle filter is the concept of Monte Carlo integration. Given an integral in the following form:

$$I = \int_{\Omega} g(\mathbf{x}) d\mathbf{x} \tag{36.99}$$

where Ω is an n_x-dimensional region in $\mathbb{R}^{n_x}$ with volume

$$V = \int_{\Omega} d\mathbf{x} \tag{36.100}$$

If N independent samples are uniformly drawn from Ω, that is, $\{\mathbf{x}^{[1]}, \mathbf{x}^{[2]}, \cdots, \mathbf{x}^{[N]}\} \in \Omega$, then the integral can be approximated as

$$I \approx I_N = V \frac{1}{N} \sum_{i=1}^{N} g\left(\mathbf{x}^{[i]}\right) \tag{36.101}$$

which approaches equality as

$$\lim_{N \to \infty} I_N = I \tag{36.102}$$

Now consider the case where the function in the integrand, $g(\mathbf{x})$, can be expressed as the product

$$g(\mathbf{x}) = f(\mathbf{x}) p(\mathbf{x}) \tag{36.103}$$

where $p(\mathbf{x})$ is a probability density function; thus, $p(\mathbf{x}) \geq 0$ and $\int p(\mathbf{x}) d\mathbf{x} = 1$. If N independent samples, $\mathbf{x}^{[i]}$, can be drawn in accordance with $p(\cdot)$, then the integral can be estimated as the sample mean of the transformed particles:

$$I_N = \frac{1}{N} \sum_{i=1}^{N} f\left(\mathbf{x}^{[i]}\right) \tag{36.104}$$

The resulting error in the estimate is unbiased and, most importantly, scales as the reciprocal of the square root of N. This is an important result as it indicates that the error is independent of the dimensionality of the state, as long as the particles are properly sampled from the distribution of $\mathbf{x}$. This is an important distinction from the grid filter, which requires particles that increase geometrically with the number of dimensions in the state vector [6].

Unfortunately, it is not always possible to generate samples from arbitrary density functions. This motivates additional development of the concept known as importance sampling.

To further our discussion of importance sampling, it is convenient to introduce the concept of a proposal density, chosen to resemble (and provide support over) the true density of $\mathbf{x}$, while retaining the ability to generate samples. An illustration of a proposal-density sampling approach is shown in Figure 36.16.

Given a random vector with true density $p(\mathbf{x})$ and particles sampled from a proposal density, $q(\mathbf{x})$, Eq. 36.103 can be rewritten as

$$g(\mathbf{x}) = f(\mathbf{x}) \frac{p(\mathbf{x})}{q(\mathbf{x})} q(\mathbf{x}) \tag{36.105}$$

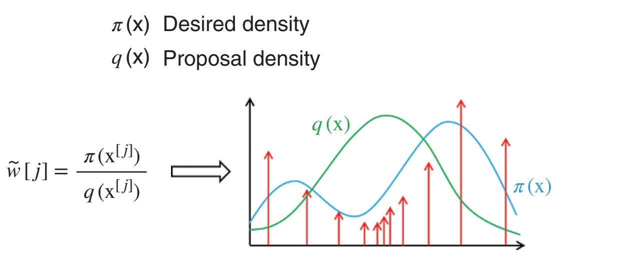

Figure 36.16 Proposal sampling illustration. In this example, the particles are generated using the proposal density (q) and subsequently weighted to represent the desired density (π).

The resulting estimate of the integral, assuming N independent particles sampled from $q(\cdot)$, is given by

$$I = \int_{i=1}^{N} f(\mathbf{x}) \frac{p(\mathbf{x})}{q(\mathbf{x})} q(\mathbf{x}) d\mathbf{x} \tag{36.106}$$

$$\approx \frac{1}{N} \sum_{i=1}^{N} f\left(\mathbf{x}^{[i]}\right) \frac{p\left(\mathbf{x}^{[i]}\right)}{q\left(\mathbf{x}^{[i]}\right)} \tag{36.107}$$

where the ratio between the true density and the proposal density can be expressed as particle importance weights:

$$\widetilde{w}^{[i]} = \frac{p\left(\mathbf{x}^{[i]}\right)}{q\left(\mathbf{x}^{[i]}\right)} \tag{36.108}$$

Substituting Eq. 36.108 into Eq. 36.107 yields

$$I_N = \frac{1}{N} \sum_{i=1}^{N} g\left(\mathbf{x}^{[i]}\right) \widetilde{w}^{[i]} \tag{36.109}$$

Finally, the collection of particle weights can be normalized via

$$w[i] = \frac{\widetilde{w}^{[i]}}{\sum_{i=1}^{N} \widetilde{w}^{[i]}} \tag{36.110}$$

then Eq. 36.109 becomes

$$I_N = \sum_{i=1}^{N} g\left(\mathbf{x}^{[i]}\right) w^{[i]} \tag{36.111}$$

which we will exploit to develop a recursive estimator.

36.3.8 Sequential Importance Sampling Recursive Estimator

In this section, we leverage the previously presented concept of importance sampling to derive the basis for a recursive nonlinear estimator using Monte Carlo integration [4]. This type of filter is generally referred to as a recursive particle filter.

Consider the following general system model:

$$\mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}, \mathbf{w}_{k-1}) \tag{36.112}$$

$$\mathbf{z}_k = \mathbf{h}(\mathbf{x}_k, \mathbf{v}_k) \tag{36.113}$$

where $\mathbf{x}_k$ is the state vector at time k, $\mathbf{f}(\cdot,\cdot)$ is the process model function at time $k-1$, $\mathbf{w}_{k-1}$ is the process noise vector, $\mathbf{h}(\cdot,\cdot)$ is the observation function, and $\mathbf{v}_k$ is the measurement noise vector at time k. The noise vectors are assumed to be independent of each other and in time with a known density function. Note that Gaussian densities are not required or assumed.

Assuming we begin with a known posterior density, $p(\mathbf{x}_{k-1}| \mathbb{Z}_{k-1})$. If N samples are drawn from an associated proposal density,

$$\mathbf{x}_{k-1|k-1}^{[i]} \tilde{q}(\mathbf{x}_{k-1}|\mathbb{Z}_{k-1}) \forall i \in \{1, \ldots, N\} \tag{36.114}$$

With normalized weights given by

$$w_{k-1|k-1}^{[i]} = \kappa \frac{p\left(\mathbf{x}_{k-1}^{[i]}|\mathbb{Z}_{k-1}\right)}{q\left(\mathbf{x}_{k-1}^{[i]}|\mathbb{Z}_{k-1}\right)} \tag{36.115}$$

where κ is the normalization factor required such that the sum of weights is unity, the posterior density function is expressed by the collection of particles and weights

$$p(\mathbf{x}_{k-1}|\mathbb{Z}_{k-1}) = \sum_{i=1}^{N} w_{k-1|k-1}^{[i]} \delta\left(\mathbf{x} - \mathbf{x}_{k-1|k-1}^{[i]}\right) \tag{36.116}$$

or, equivalently

$$p(\mathbf{x}_{k-1}|\mathbb{Z}_{k-1}) \leftrightarrow \left\{\mathbf{x}^{[i]}, w^{[i]}; i = 1, \ldots, N\right\}_{k-1|k-1} \tag{36.117}$$

Our goal is to estimate the posterior density at time k, $p(\mathbf{x}_k| \mathbb{Z}_k)$, by incorporating the statistical process model and the observation at time k. The density function of interest can be written as

$$p(\mathbf{x}_k|\mathbb{Z}_k) = \frac{p(\mathbf{z}_k|\mathbf{x}_k) p(\mathbf{x}_k|\mathbf{x}_{k-1})}{p(\mathbf{z}_k|\mathbb{Z}_{k-1})} p(\mathbf{x}_{k-1}|\mathbb{Z}_{k-1}) \tag{36.118}$$

$$\propto p(\mathbf{z}_k|\mathbf{x}_k) p(\mathbf{x}_k|\mathbf{x}_{k-1}) p(\mathbf{x}_{k-1}|\mathbb{Z}_{k-1}) \tag{36.119}$$

Assuming our proposal density can be factored:

$$q(\mathbf{x}_k|\mathbb{Z}_k) = q(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{z}_k) q(\mathbf{x}_{k-1}|\mathbb{Z}_{k-1}) \tag{36.120}$$

the posterior particle locations can be sampled from

$$\mathbf{x}_k^{[j]} \tilde{q}\left(\mathbf{x}_k|\mathbf{x}_{k-1}^{[j]}, \mathbf{z}_k\right) \tag{36.121}$$

Thus, the associated particle weights at time k can be calculated in a similar fashion as Eq. 36.108:

$$w_k^{[j]} \propto \frac{p\left(\mathbf{x}_k^{[j]}|\mathbb{Z}_k\right)}{q\left(\mathbf{x}_k^{[j]}|\mathbb{Z}_k\right)} \tag{36.122}$$

Substituting Eqs. 36.119 and 36.120 into Eq. 36.122 yields

$$w_k^{[j]} \propto \frac{p\left(\mathbf{z}_k|\mathbf{x}_k^{[j]}\right) p\left(\mathbf{x}_k^{[j]}|\mathbf{x}_{k-1}^{[j]}\right)}{q\left(\mathbf{x}_k^{[j]}|\mathbf{x}_{k-1}^{[j]}, \mathbf{z}_k\right)} \frac{p\left(\mathbf{x}_{k-1}^{[j]}|\mathbb{Z}_{k-1}\right)}{q\left(\mathbf{x}_{k-1}^{[j]}|\mathbb{Z}_{k-1}\right)} \tag{36.123}$$

Note that this equation is a function of the posterior weights at time $k - 1$; thus, the right-hand fraction of Eq. 36.123 can be replaced according to Eq. 36.108, which yields the final particle weight update equation from time $k - 1$ to time k:

$$w_k^{[j]} \propto \frac{p\left(\mathbf{z}_k | \mathbf{x}_k^{[j]}\right) p\left(\mathbf{x}_k^{[j]} | \mathbf{x}_{k-1}^{[j]}\right)}{q\left(\mathbf{x}_k^{[j]} | \mathbf{x}_{k-1}^{[j]}, \mathbf{z}_k\right)} w_{k-1}^{[j]} \tag{36.124}$$

which can be normalized such that the collection of weights sums to one, thus approximating the posterior density as

$$p(\mathbf{x}_k | \mathbb{Z}_k) \approx \sum_{j=1}^{N} w_k^{[j]} \delta\left(\mathbf{x}_k - \mathbf{x}_k^{[j]}\right) \tag{36.125}$$

In this manner, the particle locations and weights can be continuously maintained and updated using a recursive estimation framework.

36.3.9 Sampling Particle Filter Demo

In this section, we apply a sequential importance sampling particle filter design to our previous nonlinear estimation example. As before, an identical, randomly generated trajectory and measurement set from the MMAE example (Section 36.3.3) are used as inputs to the filter. Once again, for reference, the system parameters are specified in Table 36.1, and the resulting trajectory, range observations, and phase observations are shown in Figure 36.3. For this example, we use 10 000 two-dimensional particles. Finally, we exercise an importance resampling procedure [6] to ensure that the number of effective particles remains acceptable.

The SIS particle filter global state estimate and density function of position after one observation ($t = 1$ s) are shown in Figure 36.17. In this example, we show the location of the particles along with the estimated mean and one-sigma standard deviation calculated using the ensemble of particles. In the figures below, the estimated mean is represented as a magenta "plus," the true state is a green asterisk, and the estimated 2-sigma error bounds as a dashed ellipse. Each particle location is shown as a black dot.

After 22 cycles, the density shows a reduced number of peaks (see Figure 36.18) and is clearly multi-modal. Based on our knowledge of the true density functions developed in the previous examples, this indicates that the filter is incorporating sensor observations and the statistical dynamics model to effectively eliminate a number of potential ambiguity possibilities.

After 100 cycles (Figure 36.19), the filter has converged to a single ambiguity.

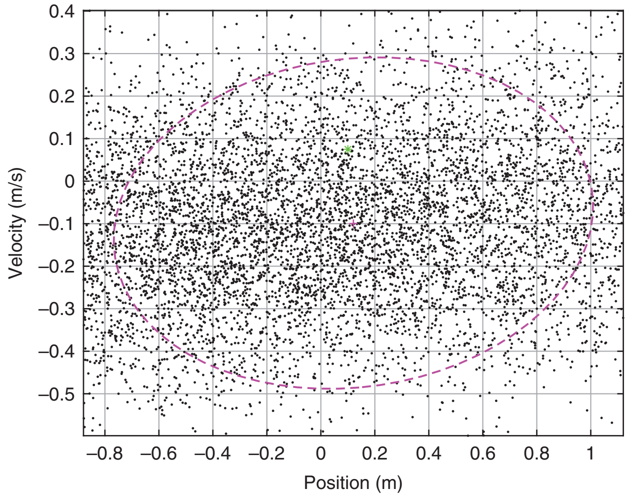

Figure 36.17 SIR particle filter initial state estimate and position density function. Note that the density function is extremely multi-modal due to the limited information available at this point.

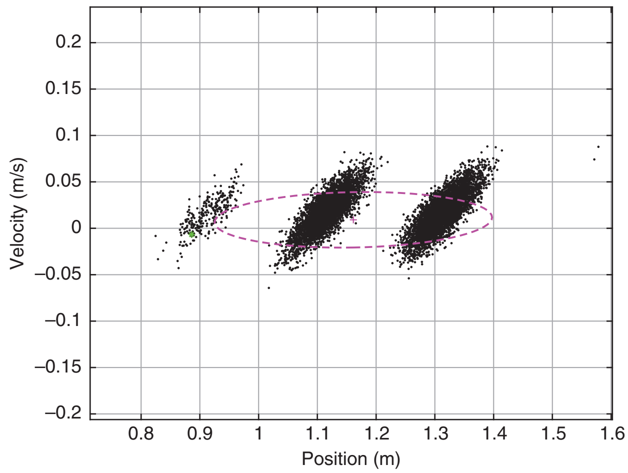

Figure 36.18 SIR particle filter state estimate (after 22 observations). Range observations combined with the vehicle dynamics model are eliminating unlikely integer ambiguity values.

The global state estimate and associated standard deviation result for this simulation is shown in Figure 36.20. The shape of the uncertainty bound clearly shows the effects described above. As the likelihood of each integer ambiguity realization changes, the overall uncertainty changes and eventually collapses to the centimeter level.

36.3.10 Strengths and Weaknesses of Approaches

In this chapter, we have presented three classes of nonlinear recursive estimation algorithms. While each algorithm offers improved performance over the linear and extended

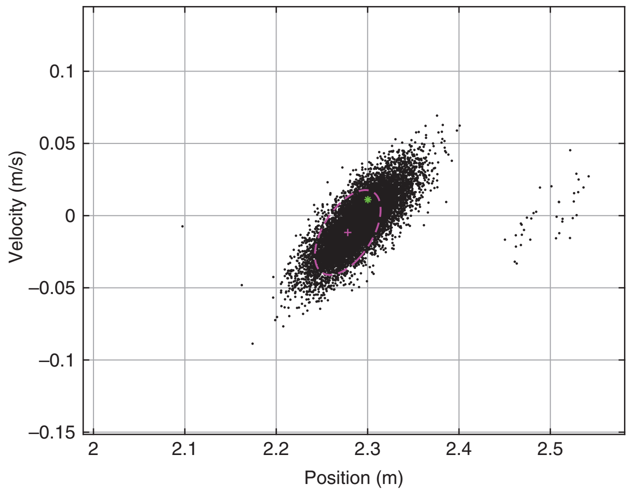

Figure 36.19 SIR particle filter state estimate (after 100 observations). Note that the state estimate is almost completely unimodal and has converged to the correct integer ambiguity.

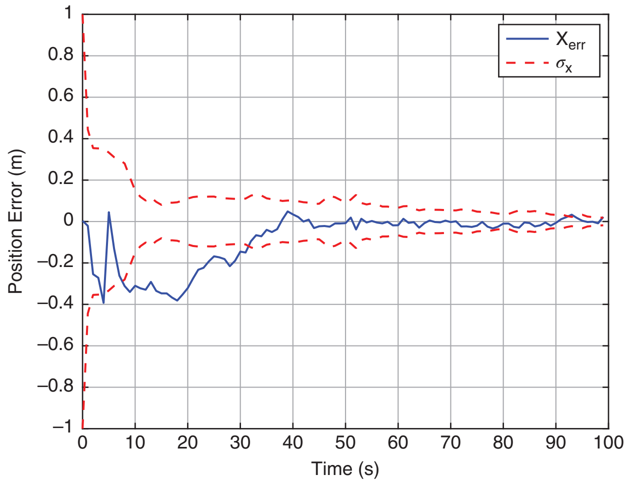

Figure 36.20 SIR particle filter position error and one-sigma uncertainty. Note that the error uncertainty collapses once sufficient information is available to resolve the integer ambiguity.

Kalman filter in the presence of nonlinearities and non-Gaussian systems, it is important to address the "strengths and weaknesses" of each. To accomplish this, we evaluate each estimation from this perspective, starting with the traditional approaches.

As expected, each approach has a set of associated strengths and weaknesses that can greatly influence the results for a given problem. Thus, the choice of estimator must be considered carefully based on the characteristics of the problem at hand. In cases where the constraints of the problem do not readily fit into the generalized categories above, there are many examples of hybrid estimation schemes that seek to synergistically combine the desirable properties of multiple estimator types. While it is beyond the scope of this chapter to explore the range of hybrid filtering approaches, the interested reader is referred to the references (e.g. [4, 5, 6, 9, 10]) for foundational concepts.

36.4 Summary and Conclusions

In this chapter, we have presented an overview of nonlinear estimation approaches suitable for navigation problems. Starting with first principles, three classes of nonlinear, recursive estimators were derived, the performance was demonstrated using a common navigation example application, and comparisons were made between the approaches.

The growing availability of a wide range of sensors and improved computational resources has heralded a new era of multisensor navigation. Because many of these sensors have nonlinear and non-Gaussian error models, researchers are developing a range of recursive navigation algorithms to meet these requirements.

When used within their associated limitations, nonlinear estimation algorithms hold enormous promise for addressing the most difficult navigation problems.

References

1. Kalman, R.E. (1960) A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 82 (Series D), 35–45.
2. Maybeck, P.S. (1979) *Stochastic Models, Estimation, and Control, Vol. I*, Academic Press, Inc., Orlando, Florida 32887.
3. Maybeck, P.S. (1979) *Stochastic Models, Estimation, and Control, Vol II*, Academic Press, Inc., Orlando, Florida 32887.
4. Gordon, N.J., Salmond, D.J., and Smith, A.F. (1993) Novel approach to nonlinear/non-Gaussian Bayesian state estimation, *in IEE Proceedings F-Radar and Signal Processing*, vol. 140, IET, vol. 140, pp. 107–113.
5. Doucet, A., De Freitas, N., and Gordon, N. (2001) Sequential Monte Carlo Methods in Practice. Series Statistics for Engineering and Information Science.
6. Ristic, B., Arulampalam, S., and Gordon, N. (2004) *Beyond the Kalman Filter: Particle Filters for Tracking Applications*, Artech House.
7. Sheldon, S. and Maybeck, P. (1990) An optimizing design strategy for multiple model adaptive estimation and control, in *Decision and Control, 1990, Proceedings of the 29th IEEE Conference on*, pp. 3522–3527, Vol. 6.

8 Bar-Shalom, Y., Li, X.R., and Kirubarajan, T. (2004) *Estimation with Applications to Tracking and Navigation: Theory Algorithms and Software*, John Wiley & Sons.

9 Doucet, A., de Freitas, N., Murphy, K., and Russell, S. (2000) Rao-Blackwellised particle filtering for dynamic Bayesian networks, in *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence*, Morgan Kaufmann Publishers Inc., San Francisco, California, UAI'00, pp. 176–183. URL http://dl.acm.org/citation.cfm?id=2073946.2073968.

10 Mustiere, F., Bolic, M., and Bouchard, M. (2006) Rao-Blackwellised particle filters: Examples of applications, in *Electrical and Computer Engineering, 2006. CCECE '06. Canadian Conference on*, pp. 1196–1200.

11 Papoulis, A. and Pillai, S.U. (2002) *Probability, Random Variables and Stochastic Processes*, McGraw-Hill, New York.

37

Overview of Indoor Navigation Techniques

Sudeep Pasricha

Colorado State University, United States

37.1 Introduction

While localization and navigation in outdoor environments can make use of global navigation satellite systems (GNSSs) such as Global Positioning System (GPS), this is not a viable solution for indoor localization because signals from GPS satellites are too weak to penetrate through buildings, obstacles, and into underground environments. Consequently, precise localization inside a closed structure, such as shopping malls, hospitals, airports, subways, and university campus buildings, require the use of alternative localization technologies. But indoor environments present unique challenges, particularly due to a diverse array of obstacles such as walls, doors, furniture, electronic equipment, and stationary or moving humans, all of which give rise to multipath effects in wireless signals due to signal reflection, attenuation, and noise interference. As a result, accurate indoor localization with wireless communication signals is a very complex problem. Moreover, indoor locales also require much higher levels of accuracy than outdoor environments; for example, while a 4–6 m accuracy is acceptable outdoors for vehicle navigation, it may not be acceptable for localization in many indoor contexts, where 4–6 m may be the difference between one room and the next.

Enabling location services for indoor locales has many potential applications. Buildings with awareness of the location of occupants can use this knowledge to optimize heating, lighting, and other resources toward saving energy costs. In emergency scenarios such as earthquakes and hurricanes, location services can allow emergency responders to determine where people are located at any time, potentially expediting evacuations as well as search and rescue efforts. Location awareness can be used as a backbone for smarter workplaces by allowing telephone calls to be routed to the nearest device in the proximity of a person, allowing colleagues to find each other, and helping guests navigate new buildings to reach their desired location. Services that utilize indoor location systems can also enable smart dynamic locking of sensitive rooms and resources if an owner is not present, to improve overall safety. Ubiquitous localization already plays a central role in social networking, for instance, to locate friends for coordinating joint activities or check into restaurants and other indoor locales via various smartphone apps, and is expected to play an even bigger role in the future. Indoor position awareness is also an essential component of industrial applications, such as for robot motion guidance, robot cooperation, and smart factories (e.g. the ability to find tagged maintenance tools and equipment scattered all over a plant in production facilities). Localization for cargo management systems at airports, ports, and for rail traffic enables unprecedented opportunities for increasing their efficiency.

Many different techniques have been proposed to enable indoor localization and navigation. The interest in indoor navigation systems is peaking because the crucial sensors necessary for localization have become sufficiently small and inexpensive to enable practical tracking of individuals (who must carry them at all times). A prime example of this is the inertial sensors that are part of inertial measurement units (IMUs) found in smartphones that can aid with localization. However, activity trackers, smart cards, and various types of wearable sensors can also play a crucial role to enable indoor navigation. The challenge today is to exploit these available sensors to achieve indoor tracking with acceptable robustness levels, similar to that demonstrated by GNSS in outdoor locales.

This chapter provides an overview of the state of the-art in the area of indoor localization and navigation. One can consider localization as an instantaneous process, providing the location of a user or object being tracked at a specific instance of time. In contrast, navigation can be considered as a form of continuous localization, where location estimates must be provided frequently and periodically over

Position, Navigation, and Timing Technologies in the 21st Century: Integrated Satellite Navigation, Sensor Systems, and Civil Applications, Volume 2, First Edition. Edited by Y. T. Jade Morton, Frank van Diggelen, James J. Spilker Jr., and Bradford W. Parkinson.

time to help a user while they navigate an indoor environment. Tracking can be considered to be similar to navigation, except that the location estimates of the mobile subject are provided not to the user but to some third party that is interested in the location information. For the sake of clarity and brevity, we will mainly use the term *localization* in the rest of the chapter to represent both instantaneous point estimates and continuous estimates (navigation or tracking), when discussing components and solutions that are relevant for indoor location estimation. We will use the term *navigation* sparingly, when necessary to discuss the unique aspects of continuous estimation of location.

The rest of the chapter is organized as follows. Section 37.2 discusses performance metrics that are necessary to understand, in order to compare and contrast the landscape of indoor localization approaches. Section 37.3 provides an easy reference to the key technical terms that are used throughout the rest of the chapter. Section 37.4 presents a review of the various signals that can be used to provide tracking in indoor locales, for the purpose of localization. Section 37.5 provides an overview of the vast landscape of solutions for indoor localization. Lastly, Section 37.6 discusses open research issues and challenges that still remain to be overcome for viable indoor localization.

37.2 Overview of Technical Terms

This section provides a brief overview of some of the commonly used technical terms that are relevant in the field of indoor localization [1].

Absolute and Relative Location. A location determined within the context of a global or large area reference grid obtained from GNSS satellites, markers, or landmarks is referred to as an absolute location. In contrast, relative positions depend on a local frame of reference, for example, coordinates within a small coverage area that represent displacement with respect to a local fixed reference (e.g. a fixed Wi-Fi access point with known global coordinates).

Anchor and Mobile Nodes. From a networking perspective, nodes in indoor environments that are part of the network and have a stationary (fixed) location are referred to as anchor nodes. In the literature, such nodes may also be referred to as beacons, fixpoints, access points (APs), base stations, or reference nodes. Typically, the coordinates of such anchor nodes are assumed to be known. In contrast, nodes that are part of the network and can move in the indoor environment are referred to as mobile nodes. Such nodes could represent people, robots, or other locomotion-capable devices (e.g. drones). In general, it is the job of the localization system to determine the (local or global) coordinates of such mobile nodes.

Centralized and Distributed Localization. In a centralized localization architecture, location estimation is carried out at a central server where all anchor and mobile node locations are stored and available to an administrator. The benefits of centralized architectures are simplicity, uniform service to all users, and lower expansion costs as most of the intelligence in the system is concentrated at one location, allowing the mobile and anchor nodes to be lower cost and contain fewer components. In a distributed system, location estimation is carried out on each mobile and anchor node based on local observations. The advantages of a distributed architecture are good system scalability and better guarantees of the user's privacy (as sensitive location information is not centrally stored, making it less susceptible to being compromised).

Line of Sight (LOS). When a signal can travel via a direct straight path from an emitter to a receiver, it is referred to as LOS transmission. Several localization techniques rely on LOS, for example, time of arrival (ToA)-based distance measurements with radio frequency (RF) signals. But due to occlusions from walls, furniture, and people, most indoor environments typically induce non-LOS (NLOS) propagation, which may cause inconsistent time delays at a radio receiver. These delays pose a challenge that can only be tackled by few localization techniques.

Multipath Environment. An environment in which a transmitted signal propagates along multiple paths (echoes), each of which arrives with different path delays at the receiver, is referred to as a multipath environment. Multipath propagation of signals is particularly problematic for time-based localization methods (Section 37.5.1.2) because signal paths from different directions degrade the ability to determine the travel time of the direct path. One way to distinguish the direct path from a non-LOS path is to move the receiver or transmitter. Non-LOS paths change erratically while in motion, allowing for separation and averaging, while the direct path is directly related to the motion of the object. Thus, averaging over time with a motion-tracking model is one effective way to mitigate multipath. Another way to overcome multipath is to switch to different frequency channels. Alternatively, radio signals with a large absolute frequency bandwidth such as Ultra-Wideband (UWB) have been shown to be advantageous for mitigation of multipath fading [2].

Received Signal Strength Indicator (RSSI). Signal attenuation can be used for distance estimation during localization, based on RSSI values. RSSI are observed RSS (received signal strength) values averaged over a specific sampling period and usually specified as received power P_R in decibels. Based on the attenuation model

$$P_R \propto P_T \frac{G_T G_R}{4\pi d^p}$$

the received signal power or signal strength P_R can help with the estimation of the distance d of a mobile user or object from the transmitter. In this model, P_T is the transmitted power at the transmitter, G_T and G_R are the antenna gains of the transmitter and receiver, and p is the path loss exponent. The path loss factor p characterizes the rate of attenuation with an increase in distance d. The free space model does not take into account that antennas are usually set up above the ground. In fact, the ground acts as a reflector, and thus the received power differs from that of free space. A mathematical formulation of such a path loss model, also known as open field model, can be found in [3]. Typically, in free space $p = 2$, whereas for environments with NLOS multipath, $p > 2$. For indoor environments, the path loss exponent typically takes higher values between 4 and 6. Theoretically, distances estimated from RSSI values to multiple anchor nodes can be used to determine the receiver position by multilateration techniques (see Section 37.5.1 for more details). However, interference, multipath propagation, and presence of obstacles and people results in a complex spatial distribution of RSSI values, which can make the estimation of distances using RSSI alone quite inaccurate. Therefore, fingerprinting has become more popular than propagation modeling (see Section 37.5.2 for more details).

37.3 Performance Metrics

Indoor localization solutions need to meet several goals if they are to be considered viable candidates for use in indoor environments. Here we review some of the more relevant performance metrics [4] that must be satisfied by any candidate indoor localization solution:

Accuracy: The location error of a positioning system is one of the most important metrics used to determine the effectiveness of a localization system. In its simplest form, localization accuracy can be reported as an error distance between the estimated location and the actual location of the user or object being tracked. For navigation systems, this may take the form of a running average of errors over a time period of interest, or the error could be calculated using geometric principles, to estimate the deviation of the predicted trajectory from the actual trajectory. Usually, the higher the accuracy, the better the system, but there is often a trade-off between accuracy and other characteristics. Therefore, a compromise between adequate accuracy and other characteristics described below is essential.

Timeliness: The timeliness or responsiveness of a solution determines how quickly the location estimate of a target is obtained. For simple indoor localization queries, a fast response to the query is important in most cases, but not crucial. However, for navigation systems, timeliness is a critical measure of effectiveness: if location estimates are not updated quickly in sync with the motion profile of the subject being tracked, the system will be ineffective for the purpose of navigation (regardless of the eventual accuracy of the estimates). Usually, the term *location lag* is used to refer to the delay between a mobile subject moving to a new location and the new location of that subject being reported by the system.

Coverage: Any indoor localization solution must work and be usable over the entire indoor environment of interest. Coverage defines the area over which a localization solution can provide estimates of sufficient accuracy, and possibly timeliness, to be considered useful. The physical environment (e.g. obstacles, walls, doors) plays a crucial role in limiting the availability of signals that are used by a given localization technique, consequently impacting the coverage achievable by the technique for that environment. Intuitively, it is possible to extend coverage by altering the physical environment or supplementing it with additional hardware, for example, wireless signal repeaters. Coverage can also be improved by enhancing the hardware carried by the user or object being tracked, for example, using mobile devices with more powerful and capable wireless radio antennas and chipsets.

Adaptiveness: Often, the physical environment around the subject to be tracked does not stay the same over time. For example, at different times of the day and days of the week, the number of people in a shopping mall varies quite significantly. In some environments, machinery, goods, containers, and other equipment may be repositioned constantly. Sometimes signals from wireless transmitters are temporarily blocked in an environment, or some transmitters may stop functioning due to unpredictable circumstances. These changes create a challenge for any indoor localization solution that relies on these signals. The ability of a solution to cope with these environmental changes represents its adaptiveness, or robustness. Obviously, a solution that is able to adapt to environmental changes can provide better localization accuracy than solutions that cannot adapt. An adaptive system can also prevent the need for repeated calibration of sensors used for localization.

Scalability: At a system level, solutions for localization may require supporting requests from multiple entities.

For instance, a system deployed in a shopping mall needs to be able to handle location queries from a few people, all the way up to thousands of people simultaneously. The ability to "scale up" and quickly respond to multiple location queries is of paramount importance in many indoor environments. Poor scalability can result in poor localization performance, necessitating the reengineering or duplication of systems, which can increase deployment overheads.

Integrity: The confidence that can be placed in the output of a localization solution can be termed its integrity. A solution with low integrity has a high probability that a malfunction will lead to an estimated position that differs from the required position by more than an acceptable amount and that the user will not be informed within a specified period of time about the malfunction. While regulatory bodies have studied and defined integrity performance parameters in some sectors such as civil aviation, for indoor localization it is more difficult to find well-quantified integrity parameters. At the very least, an indoor localization solution must provide an indication of some integrity parameters that are related to safety of life, economic factors, or convenience factors; thereby allowing consumers of the solution to understand its limits and capabilities under different usage scenarios.

Cost: An indoor localization system has costs associated with it that must be as low as possible, to incentivize widespread adoption and ease deployment overheads. These costs may include installation of localization solution-specific hardware and site survey time during the deployment period. If a positioning system can reuse an existing communication infrastructure (e.g. Wi-Fi APs already deployed in a building), some part of the infrastructure, equipment, and bandwidth costs can be saved. In addition to the infrastructure, there may also be costs associated with the mobile devices carried by the subject being tracked. For instance, such costs could represent monetary costs of the smartphone and any externally connected hardware sensors. However, the cost could also be calculated by considering other aspects, such as lifetime, weight, and energy consumption. For example, some mobile devices, such as electronic article surveillance (EAS) tags and passive radio frequency identification (RFID) tags, are energy passive (i.e. they only respond to external fields) and thus, can have an unlimited lifetime; however other mobile devices (e.g. smartphones with rechargeable battery) have a limited lifetime of several hours without recharging.

Complexity: Indoor localization solutions inevitably require hardware and software components that can have different complexities. Solutions may differ in the sophistication required from their associated signal processing software and hardware. While some techniques may involve very simple hardware (e.g. inertial sensors) and software (e.g. to implement simple filtering techniques), other techniques may require more complex custom hardware (e.g. for specialized digital signal processing) and complex software (e.g. sophisticated machine learning techniques). Also, if the computation of the localization algorithm is performed on a centralized server, the localization can be quickly estimated due to the powerful processing capability and the sufficient power supply; however if it is carried out on a mobile device, the effects of complexity can be much more apparent. Inevitably, complexity impacts the cost of the solution, and thus it is common practice to trade off the complexity with the other (non-cost) metrics.

37.4 Indoor Localization Signal Classification

GPS is the most popular wireless-signal-based positioning system in use today, and is extremely useful in outdoor environments [5]. GPS satellites broadcast microwave signals to enable GPS receivers on or near Earth's surface to determine location, velocity, and time. The GPS system itself is operated by the US Department of Defense (DoD) for use by both the military and the general public. Unfortunately, GPS signals cannot penetrate into indoor environments due to obstacles that spread and attenuate the electromagnetic radio signals [6]. Thus, GPS cannot be used for localization in indoor environments. Fortunately, there are many other signals available in indoor locales that can be leveraged by solutions intended for indoor localization. This section reviews some of the more relevant signals that can be used for indoor localization. Figure 37.1 shows a taxonomy of the signals that are covered in more detail in the rest of this section.

37.4.1 Infrared Radiation (IR) and Visible Light

Electromagnetic radiation at wavelengths within the visible range, which extends approximately between 380 and 750 nm, as well as in its lower or upper vicinity, known as ultraviolet (UV) and infrared (IR) light, are part of some of the most common indoor positioning systems that use wireless technology.

Visible light systems typically utilize general-purpose cameras and have been adopted particularly for indoor localization of robots. One common approach is to have a robot carry a camera to capture images of the environment

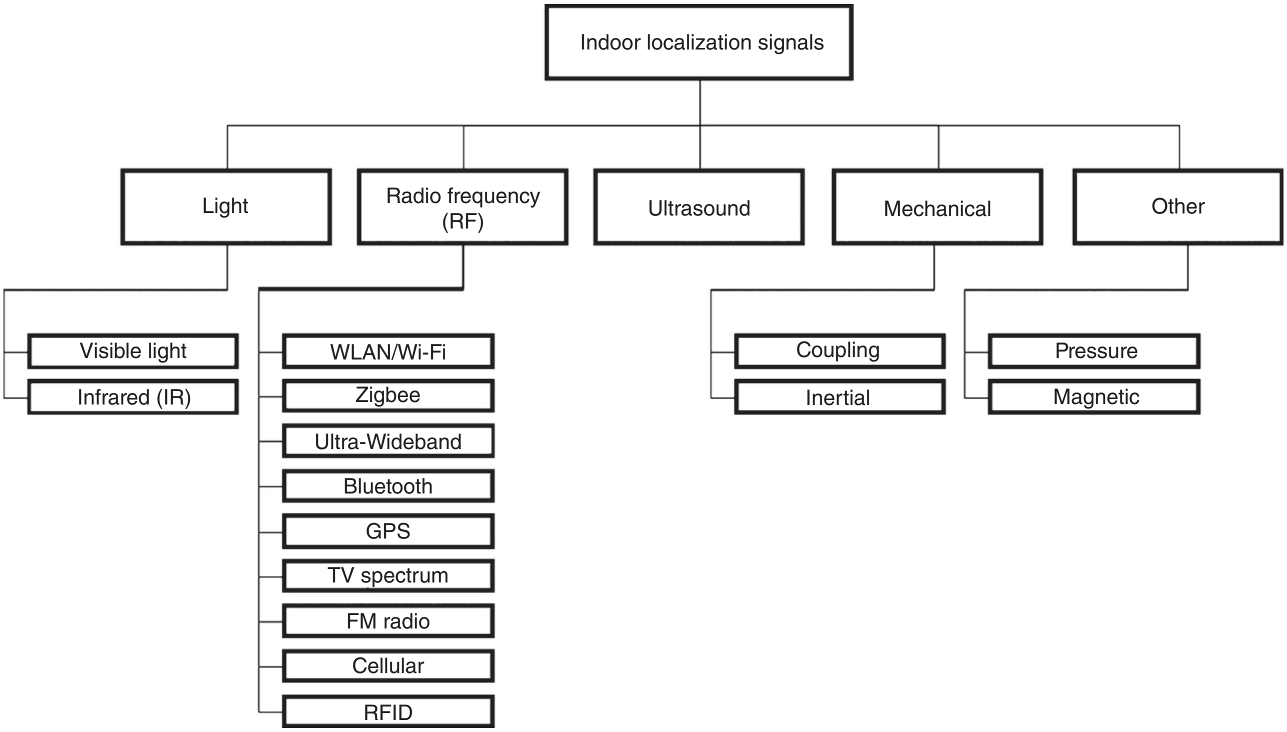

Figure 37.1 Taxonomy of signals for indoor localization.

that can then be processed to infer location with respect to the environment [7]. Other approaches deploy cameras in fixed locations across the environment, and if the salient features of the object to be tracked appear in the field of view of the camera, the location of the object can be calculated with respect to the camera's fixed position [8]. A key challenge is how location can be estimated in a 3D world when the primary observations are 2D, from an image sensor. Depth information can be obtained by making use of the motion of a camera. In such an approach – known as synthetic stereo vision – the scene is observed sequentially from different locations by the same camera (or by multiple fixed coordinated cameras) and image depths can be estimated in a manner similar to the well-known stereo-vision approach. Alternatively, distances can be directly measured with additional sensors, such as with laser scanners or range imaging cameras. The latter returns a distance value for every pixel of an image at a specific frame rate.

All visible-light-based localization approaches require some form of image processing, which is time consuming and can be particularly error prone in some dynamic environments, for example, due to illumination variability [8]. In the case of laser based-solutions, only class 1 laser devices should be used, which are classified as "eye-safe" by the IEC 60825-1 standard [9]. Another challenge arises due to occlusions caused by dynamic elements of the environment (e.g., moving objects or people). One way of reducing occlusion is to deploy sensors with overlapping coverage areas [8]. However, clinical settings and public indoor areas such as shopping malls are often densely populated, and therefore occlusion conditions can arise frequently even with ceiling-mounted sensors.

IR-based localization systems are also very popular, relying on a LOS communication mode between the transmitter and receiver. For instance, [10] presents an IR-based localization system for museums with IR emitters installed in the ceiling of the door frames of every room. Each emitter transmits a unique ID using the Infrared Data Association (IrDA) protocol. Visitors carry a personal digital assistant (PDA) with an infrared port. The PDA contains a database of visual and textual information of the exhibits, as well as maps of the museum. Upon reception of a new ID, the PDA automatically presents the map of the corresponding exhibit hall. The main advantage of using IR-based system devices is that they are small, lightweight, and easy to carry. However, IR-based indoor positioning systems also have several limitations for location estimation, such as interference from fluorescent light, sunlight, as well as noise and reflections [11].

It is important to mention the privacy issues that may arise when using imaging-based localization solutions. Typical solutions capture images of the environment, and thus can reveal important information about the person wearing the system or bystanders, for example, patients and health care personnel in the vicinity, in a hospital environment. This is particularly challenging because certain facilities (e.g., for health care) are required to protect the privacy of personnel, patients, and clients. The scenario becomes even more complex if image-processing-based mobile localization devices are designed to send captured images to central, computationally powerful servers for image processing. The confidentiality of the image is at great risk of being compromised while in transit over a network [12].

37.4.2 RF Signals

RF technologies [13] are commonly used in location position systems because radio waves can penetrate through obstacles such as building walls and human bodies easily. Moreover, they also enable a larger coverage area than other techniques. Localization solutions in this category estimate the location of a mobile user in the environment by measuring one or more properties of an electromagnetic wave radiated by a transmitter and received by a device carried by the mobile user. These properties typically depend on the distance traveled by the signal and the characteristics of the surrounding environment. RF localization systems can be categorized according to the underlying hardware technology and network type used: (i) personal and local area networks, and (ii) broadcast and wide area networks.

37.4.2.1 Personal and Local Area Networks

Personal and local area networks include technologies such as IEEE 802.11 (WLAN), Bluetooth, Zigbee, Ultra-Wideband (UWB), and RFID.

WLAN APs are ubiquitous in indoor environments, where they are used to provide Wi-Fi Internet services to users with a range of approximately 50–100 m per Wi-Fi access point. Therefore, Wi-Fi signals represent some of the most widely available RF signals in indoor environments. Not surprisingly, a majority of triangulation and fingerprint-based indoor localization techniques (discussed in more detail in Section 37.5) utilize Wi-Fi signals. The 802.11 Wi-Fi family consists of several standards. The 802.11b and 802.11g standards use the 2.4 GHz industrial, scientific, and medical radio (ISM) radio band, and employ direct-sequence spread spectrum (DSSS) and orthogonal frequency-division multiplexing (OFDM) signaling methods to limit occasional interference from microwave ovens, cordless telephones, and Bluetooth devices. The 802.11a standard uses the 5 GHz U-NII band, which for much of the world, offers at least 23 non-overlapping channels rather than the 2.4 GHz ISM frequency band offering only 3 non-overlapping channels, where other adjacent channels overlap. The 802.11n standard allows using either the 2.4 GHz or the 5 GHz band, while 802.11ac uses only the 5 GHz band. Note that as the segment of the RF spectrum used by 802.11 varies between countries, indoor localization solutions that utilize WLAN signals may need to be adjusted when deployed in different countries.

Bluetooth is a wireless standard for wireless personal area networks (WPANs), for exchanging data over short distances (using short-wavelength RF waves in the ISM band from 2.4 to 2.485 GHz). Almost all Wi-Fi-enabled mobile devices available in the market today (e.g. smartphones, tablets, laptops) have an embedded Bluetooth module. Bluetooth has a smaller coverage area than Wi-Fi (typically 10–20 m). In [14] it was shown that the Bluetooth Low Energy (BLE; Bluetooth 4.0) propagation model can better relate RSSI to range than Wi-Fi, which indicates that BLE can be more accurate when used in localization scenarios. However, Wi-Fi has a much wider coverage than BLE, so BLE-based localization solutions will require more anchors/beacons compared to Wi-Fi APs, for the same coverage area.

Zigbee is another short-range wireless technology based on the IEEE 802.15.4 specification and mainly designed for applications which require low power consumption but do not require large data throughput. It operates in the 2.4 GHz ISM band in most jurisdictions worldwide, 784 MHz in China, 868 MHz in Europe, and 915 MHz in the United States and Australia. Zigbee implementations are typically more economical, more energy efficient, and have higher coverage (~100 m) than Bluetooth implementations. However, Zigbee support is less common in mobile devices than Bluetooth support. Moreover data rates for Zigbee vary from 20 kbit/s (868 MHz band) to 250 kbit/s (2.4 GHz band), which are much lower than the data rates achievable with Bluetooth (1–25 Mbit/s).

RFID systems are commonly composed of one or more reading devices that can wirelessly obtain the ID of tags present in the environment. When the reader transmits an RF signal, RFID tags in the environment reflect the signal, modulating it by adding a unique identification code [15]. The tags can be active, that is, powered by a battery, or passive, drawing energy from the incoming radio signal. The detection range of passive tags is therefore more limited compared to that of active tags. RFID technology is used in a wide range of tracking applications in the automobile assembly industry, warehouse management systems, and across supply chain networks, where LOS contact is difficult or even impossible [16]. Passive RFID systems typically make use of four frequency bands: LF (125 kHz), HF (13.56 MHz), UHF (433, 868–915 MHz), and microwave frequency (2.45 GHz, 5.8 GHz). Active RFID systems use similar frequency ranges, except for the low-frequency and high-frequency ranges.

UWB radio technology is designed for short-range, high-bandwidth communication, with the desirable properties of strong multipath resistance. Unlike narrowband wireless technologies such as Wi-Fi, which beam signals within a defined frequency band (e.g. the 2.4 GHz or 5 GHz band), UWB scatters its transmissions over several gigahertz of the spectrum (from 3.1 to 10.6 GHz in the United States as restricted by the FCC; 6.0 to 8.5 GHz in Europe as restricted by the ECC) using short pulses (typically <1 ns). UWB waves typically occupy a much larger frequency bandwidth (>500 MHz) than narrowband operation. Due to its spectrum-scattered approach to communication, UWB is theoretically less susceptible to interference. UWB short-duration pulses are easy to filter, to determine which signals are correct and which are generated from multipath. UWB signals also pass easily through walls, equipment, and clothing, but metallic and liquid materials can still cause UWB signal interference. A major advantage of using UWB for distance measurements is that large bandwidth translates into a higher resolution in time and consequently in distance than other technologies.

37.4.2.2 Broadcast and Wide Area Networks

Broadcast and wide area networks include networks designed for localization purposes, such as GPS, and broadcast networks not originally intended for localization purposes, such as television broadcast signals [17], cellular phone networks [18], and FM radio signals [19].

As the signal properties and geometrical arrangement of the digital TV broadcast network have been designed to penetrate indoors, they offer significantly greater indoor coverage than GPS-based solutions. For instance, [17] proposed using synchronization signals already present in the Advanced Television Signal Committee (ATSC) standard for compliant digital TV signals to perform indoor localization. Emitters of digital television are synchronized with GPS time, allowing the data to be time-stamped, which can be useful for distance estimation with ToA techniques (Section 37.5.1.2). Digital TV signals also have a wide bandwidth of 5–8 MHz that can theoretically help reduce multipath mitigation. However, the weak density of terrestrial emitters causes the direct signal to arrive at low elevation angles near the horizon. As such, only 2D positioning is feasible, and multipath is severe because the direct signal is usually blocked.

Similar to digital TV networks, cellular networks have a wider range than, say, Wi-Fi signals, and can also be used for indoor localization, much like with Wi-Fi. With the Federal Communications Commission's Enhanced-911 (E-911) mandate, cellular networks include positioning information as part of their standards. The E-911 mandate requires mobile phones to be locatable within a 50 m accuracy for 67% of emergency calls. Such accuracies are usually achievable with the help of GPS signals. However, an accuracy of 50 m is insufficient for indoor areas. Therefore, cellular networks such as 2G/GSM, 3G UMTS, 4G LTE, and emerging standards must be utilized with other technologies for finer-grained localization resolution in indoor environments.

FM radio is another possible candidate, utilizing frequency-division multiple access (FDMA) to split the wireless band into a number of separate frequency channels that are used by stations. FM band ranges and channel separation distances vary in different regions, but the pervasiveness of FM signals can enable their use for indoor localization. Typically, radio waves operating in the frequency band 87.5 to 108.0 MHz are part of the FM spectrum. Due to the passive nature of the client devices, FM can be used in sensitive areas where other RF technologies are prohibited for safety or security reasons. Unfortunately, FM signals lack timing information, which limits their use in certain localization techniques (such as the time-based trilateration techniques discussed in Section 37.5.1.2).

37.4.2.3 Challenges

In general, the propagation of RF signals in indoor environments faces several challenges. Certain materials in the indoor environment affect the propagation of RF waves. For example, materials such as wood or concrete attenuate RF signals, while materials such as metals or water cause reflections, scattering, and diffraction of RF waves. These effects lead to multipath radio wave propagation, which prevents accurate calculation of the distance between the transmitter and the receiver in indoor environments. The propagation of RF waves is also affected by changes to the physical indoor environment (e.g. movement of people, rearrangement of furniture, structural modifications). In these environments, the RF properties are highly dynamic, and a radio map captured at a certain point in time cannot be used reliably for localization without accounting for these dynamic changes [12]. Moreover, while some solutions operate within a reserved radio band [18], most solutions utilize open spectrum bands. This means that these solutions must account for the increased risk of interference due to other systems sharing the same frequency bands of the radio spectrum. The usage of radio transmitting devices is also restricted in some cases, for example, in critical areas of most healthcare facilities, according to recommendations made by the Association for the Advancement of Medical Instrumentation (AAMI) [20] and other standards or regulatory bodies. Such restrictions limit the deployment of localization systems based on non-broadcast RF waves.

37.4.3 Ultrasound

Sonic waves are mechanical vibrations transmitted over a solid, liquid, or gaseous medium. The sonic waves produced by vibrations below and above the threshold of human hearing are known as infrasonic and ultrasonic waves, respectively. There are indoor localization solutions that propose the use of ultrasonic range finders and sonars [21]. The relative distance between two devices can be estimated from ToA measurements (see Section 37.5.1.2) of ultrasound pulses from the emitter to the receiver, and thus ultrasound signals can be used to estimate the position of the emitter tags from the receivers. Typical ultrasound systems operate in the low-frequency band compared to the other RF signaling technologies. In contrast to RF waves, the ultrasound ToA operating range is 10 m or less due to the specific decay profile of the airborne acoustic channel. Doubling the distance causes the signal's sound pressure level to attenuate by 6 dB due to radial intensity attenuation and absorption, which translates to an inverse quadratic attenuation in 3D space. In general, ultrasound signals are unable to penetrate walls, and they reflect off most indoor obstructions (furniture, people), resulting in echoes that can lead to localization inaccuracies. It has also been observed that high levels of ambient noise prevent accurate detection of the sonic signal; co-interference caused by the presence of multiple sonic emitters in the environment also leads to errors. Variations in the speed of sound over air are another challenge: for instance, temperature variations are known to affect the speed of sound in air [22]. Therefore, sonic-wave-based systems cannot be used in environments with frequent and drastic temperature changes.

37.4.4 Inertial and Mechanical

Whenever energy is exerted due to the mechanical movement of a moving subject, the energy can be measured and used for localization in indoor environments. As an example, a "Smart Floor" [23] was proposed with metallic plates that were instrumented with load cells, which used mechanical coupling between a moving person and the load cells. The plates were laid on the floor, and the signal captured via the load cell was processed in order to identify the person walking over a plate and the path they were taking. A more common example of exploiting mechanical energy for localization is via accelerometer (to measure acceleration) and gyroscope (to measure angular rotation) sensors. Such "inertial" sensors that are part of IMUs commonly found in smartphones can be used to estimate the trajectory of motion for a moving person or object, which can help with their localization in indoor environments [24]. In particular, such sensors are very useful to estimate the stride length and step counts for a person in motion, to determine their displacement over time. Techniques that use inertial sensors for localization are often referred to as "dead reckoning" techniques, as the location estimates provided by the sensors depend on previous measurements to estimate the absolute position or orientation of the object being tracked at any given instant. A challenge with using inertial measurements for localization is that the inertial sensors are susceptible to drift due to thermal changes in the circuitry of the sensors, calibration issues, and inherent noise [25].

37.4.5 Other Signals

There are a few other signals that can aid with indoor localization. Atmospheric pressure can be captured using barometric/altitude/pressure sensors, and used to provide estimates of the altitude of the person or object to be tracked. Magnetic readings captured by a (digital) compass can also be used for heading (direction) estimation. Most IMUs today include three perpendicular magnetometer sensors to measure the strength and/or direction of a magnetic field, along with traditional 3-axis accelerometer sensors for motion estimation, and 3-axis gyroscope sensors to measure angular rotation. However, spurious electromagnetic field disturbances can affect the readings of the magnetometer sensors when in proximity to metallic structures or radio-wave-emitting devices.

In general, the signals discussed above can help improve the accuracy of indoor localization when used in tandem with other more robust and comprehensive localization signals, for example, dead reckoning or RF-signal-based localization.

37.5 Indoor Localization Techniques

Having identified the commonly used signals for indoor localization, we now present a survey of various indoor localization techniques that have been proposed and evaluated to date. We classify these techniques in this section based on the measuring principles used: triangulation, fingerprinting, proximity, dead reckoning, map matching, and hybrid techniques. Typically, for all of these different types of techniques, there are two main approaches for deployment: (i) developing a custom signaling and network infrastructure, and (ii) reusing an existing network infrastructure (e.g. existing Wi-Fi APs in a building). With the first approach, it is possible to control the physical specification and, consequently, the quality of the location sensing results; whereas the second approach has much lower costs as it avoids expensive and time-consuming deployment of infrastructure [26].

37.5.1 Triangulation

Triangulation is a family of wireless radio-signal-based methods that use the geometric properties of triangles to determine location. The methods can be broadly classified as angulation-based and lateration-based [27]. Angulation locates an object by computing its angles relative to multiple fixed reference points. In contrast, lateration estimates the position of an object by measuring its distances from multiple reference points (the general term *multilateration is* often used whenever two or more reference points are used). As a proxy for directly using distance, some methods use the RSS, ToA or time difference of arrival (TDoA). In these methods, the distance is derived by computing the attenuation of the signal strength or by multiplying the radio signal velocity and the signal travel time. A few methods also use the round-trip time of flight (RToF) or received signal phase for distance estimation. We describe the major triangulation-based methods for indoor localization in the rest of this section.

37.5.1.1 Angle-Based Methods

The angle of arrival (AoA) technique estimates the location of the desired target by analyzing the intersection of several pairs of angle direction lines, each formed by the circular radius from a base station or a beacon station to the mobile target. Figure 37.2 shows how AoA methods may use at least two known reference points (A, B), and two measured angles (θ_1, θ_2) to derive the 2D location of the subject *P*. The actual estimation of AoA can be accomplished with directional antennas or an array of antennas. The AoA between a UWB pulse arriving at multiple sensors has been used for

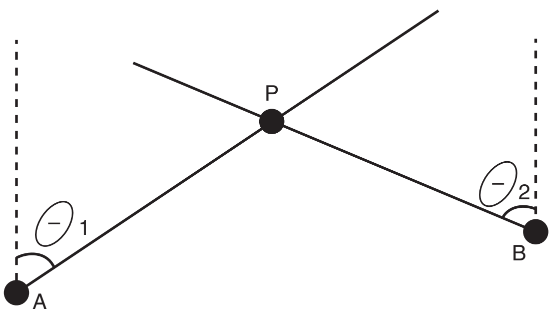

Figure 37.2 Positioning based on angle of arrival (AoA) measurement [27]. *Source:* Reproduced with permission of IEEE.

real-time 3D location positioning in [28]. The advantages of the AoA approach are that a location estimate may be determined with just three measuring units for 3D positioning or two measuring units for 2D positioning, and that no time synchronization between measuring units is required [29]. The disadvantages are primarily due to the large and complex hardware needed (e.g. Quuppa's HAIP system [30] uses AoA for indoor localization but requires a specific hardware device including 16 array antennas with a transmitter as nearby anchors and special tags for positioning), and location estimate degradation as the mobile subject moves farther from the measuring units [31]. The angle measurements need to be accurate for accurate positioning, but this is challenging with wireless signals due to limitations imposed by shadowing, multipath reflections arriving from misleading directions, or by the directivity of the measuring aperture [32].

AoA-based methods have been used in several light-based localization solutions. PIXEL [33] is an indoor localization solution that uses AoA methods to determine localization and orientation of mobile devices. The system consists of beacons that periodically send out their identity via visible light communication, which are captured by the mobile devices, followed by AoA-based post-processing. Luxapose [34] also uses visible light and employs AoA techniques for indoor localization. In [35], an AoA-based localization solution was proposed based on passive thermal IR sensors to detect thermal radiation of the human skin. The system is passive as it uses natural infrared radiation without any active IR signal emitters. The approach used thermophiles (a series of thermocouple-based temperature sensor elements) with a lower resolution compared to IR cameras. Multiple sensors were placed in the corners of a room from where the angles relative to the radiation source were measured. The position of human subjects was then roughly estimated via the principle of AoA, using triangulation from multiple thermophile arrays. However, the effects of dynamic background radiation need to be carefully considered before the method is considered for use in real-world environments.

A somewhat different technique from AoA that also exploits angular information was proposed in [36]. The system uses a fixed beacon composed of an active infrared (IR) light source and an optical polarizing filter, which only passes light through that oscillates along a single plane. The mobile receiver consists of a photo detector and a rotating polarizer that causes attenuation of the signal intensity depending on the horizontal angle. The phase of the time-varying signal is then translated into the angle of the polarizing plane. This allows estimation of the absolute azimuth angle with an accuracy of 2% (or a few degrees).

37.5.1.2 Time-Based Methods

ToA-based localization solutions are based on the synchronization of the arrival time of a signal transmitted from a mobile subject (P) to at least three receiving beacons, as shown in Figure 37.3. The underlying idea is that the distance from the mobile subject to the beacons is directly proportional to the propagation time. This distance between the mobile subject and beacons is calculated based on one-way propagation time measurements [37]. Several methods have been proposed for such measurements using DSSS [38] or UWB [39, 40]. In general, short-pulse UWB waveforms permit accurate determination of the precise ToA and time of flight of a burst transmission from a short-pulse transmitter to a corresponding receiver, which has been utilized in UWB-based indoor localization solutions that can achieve very high indoor location accuracy (down to 20 cm in some cases) [41, 42]. But care needs to be taken in ToA systems so that that all transmitters and receivers in the system are precisely synchronized. Also, the transmitting signal must send a timestamp for the receiver to discern the distance the signal has traveled. The Active Bat positioning system [43] uses ultrasound signaling and ToA triangulation to measure the location of a tag carried by a person. The tag periodically broadcasts a short pulse of ultrasound that is received by a matrix of ceiling-mounted receivers at known positions. The distances between the tag and the receivers can be measured by the ToA of the ultrasonic waves. The Hexamite system [44] also uses ToA triangulation-based localization with ultrasound signaling. A hybrid ToA/AoA approach was introduced in [45]. By utilizing the information measured from AoA and ToA, the number of beacons (anchors) required can be reduced. In [46], another hybrid approach is proposed, in which a hybrid AoA/ToA scheme is used for localization if only one Wi-Fi AP is available; however, if more APs are available, then a multiple-message-based AoA scheme is used to obtain higher accuracy location. This design aims to provide accurate localization even when the number of nearby anchors (Wi-Fi APs) is limited.

TDoA techniques determine the relative position of a mobile transmitter by analyzing the difference in time at which the signal arrives at multiple measuring units, rather than the absolute arrival time of ToA. A 2D target location can be estimated from the two intersections of two or more TDoA measurements, as shown in Figure 37.4. Two hyperbolas are formed from TDoA measurements at three fixed measuring units (A, B, C) to provide an intersection point and locate the subject P. The conventional methods for obtaining TDoA estimates are to use correlation techniques, for example, by the cross-correlation between the signals received at a pair of measuring units. With TDoA, a transmission with an unknown starting time is received at various receiving nodes, with only the receivers requiring time synchronization [47]. TDoA does not need a synchronized time source of transmission in order to resolve timestamps and find the location. A delay-measurement-based TDoA estimation method was proposed in [38] for Wi-Fi signals, which eliminates the requirement of initial synchronization in conventional methods. The TDoA between a UWB pulse arriving at multiple sensors has been used for high-precision real-time 3D location positioning [28]. Wi-Fi-based TDoA was proposed by [48], for indoor location estimation. The approach requires the same radio signal to be received at three or more separate points, timed very accurately (to a few nanoseconds) and processed using a TDoA algorithm to determine the location. A TDoA system with a proprietary RF signal (from the 2.4 GHz band) was proposed [49], with an emphasis on power efficiency. The approach used a dedicated standard protocol (ANSI 371.1)

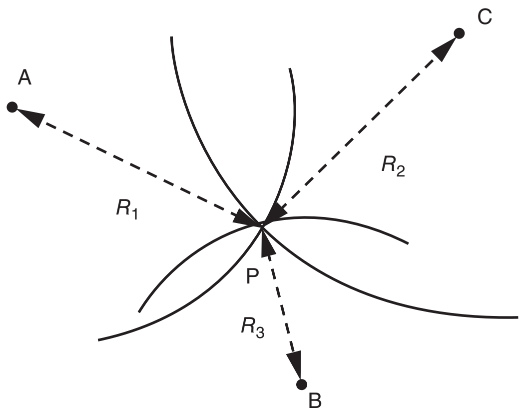

Figure 37.3 Positioning based on ToA/RToF measurements [27]. *Source:* Reproduced with permission of IEEE.

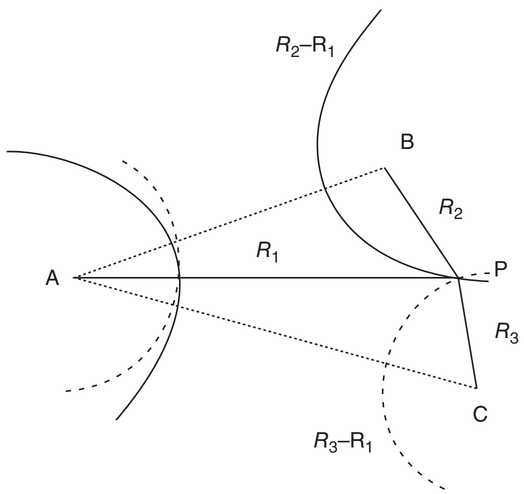

Figure 37.4 Positioning based on time difference of arrival (TDoA) [27]. *Source:* Reproduced with permission of IEEE.

optimized for low-power spread-spectrum localization, which works by timing the signals transmitted from tags to a network of receivers.

RToF techniques measure the time of flight of a signal from the transmitter to the measuring unit and back. While ToA techniques calculate delay by using two local clocks in two different measuring nodes, RToF techniques use only one node to record the transmitting and arrival times. Thus, RToF techniques are less susceptible to synchronization problems than the other time-based methods. An algorithm to measure the RToF of Wi-Fi packets is presented in [9], with the results indicating measurement errors of a few meters. The positioning algorithms for ToA can often be directly applied in RToF techniques. Typically, the mobile subject responds to a received signal from the measuring units, and these measuring units calculate the RToF; however, it is difficult for the measuring unit to know the exact processing/response delay time at the mobile subject. Another challenge is that the measuring node may become overloaded when tracking multiple mobile subjects moving quickly. The 3D-ID system [50] uses the RToF for distance estimation during localization. In the proposed approach, whenever a mobile tag receives a broadcast, the tag immediately rebroadcasts it on a different frequency, modulated with the tag's ID. A cell controller cycles through the antennas, collecting a set of ranges to the tag. With a 40 MHz signal, this system was shown to achieve a 30 m range, 1 m precision, and 5 s location update rate.

It should be noted that radar (RAdio Detection And Ranging) systems exploit time-based methods, such as the ones discussed above, for localization of an object. The original principle of radar was to measure the propagation time and direction of radio pulses transmitted by an antenna and then bounced back from a distant passive target. If the object returns some of the wave's energy to the antenna, the radar can measure the elapsed time (i.e. RToF) to estimate the distance, as well as the angle of incidence (by using a directional antenna). This original concept of radar assumes passive object reflection and involves only one station with both a transmitter and sensor. But in such systems, most of the signal energy gets lost due to reflection, and the use of steerable directional antennas is impracticable. Therefore, the concept of radar has been extended to include more than one active transmitter (secondary radar). Instead of passive reflection, the single-way travel time of the radar pulse is measured by ToA and then returned actively. Frequency-Modulated Continuous Wave (FMCW) radar is a short-range localization technique, where the transmitter frequency is linearly increased with the time [1]. The returned echo is received with a constant offset, which relates to the distance traveled. An advantage of FMCW is its resistance to the Doppler effect. The Doppler movement introduces a shift in the frequency which is canceled out by differencing. Most FMCW radar implementations make use of multilateration based on RToF-based distance estimates between a mobile transmitter and multiple fixed transponders. The transponder broadcasts a radio signal in the free ISM band (5.725 GHz to 5.875 GHz) which is received, processed, and echoed back to the transponder by each transmitter without time delay. The echo is coded with the respective transponder's identification in order to allow the transmitter to separate each transponder´s answer. A localization system based on FMCW radar was proposed in [51] that consists of multiple fixed base stations and a lightweight mobile transponder operating at 5.8 GHz. Based on TDoA ranges at centimeter-level precision measured under LOS conditions, a positioning accuracy of 10 cm over 500 m was shown to be achieved.

37.5.1.3 Signal-Property-Based Methods

The triangulation-based localization techniques discussed previously compute the distance to the mobile subject using either timing or angle information. But in the absence of LOS channels between transmitters and receivers, the underlying mechanism in both types of techniques (time and angle) are impacted by multipath effects, which can reduce the accuracy of the estimated location.

An alternative approach to measuring the distance of a mobile subject to some reference measuring nodes involves using the attenuation of the emitted (radio) signal strength. Theoretical and empirical models are usually used to translate the difference between the transmitted signal strength and the RSS into a range estimate. Such an RSSI is the most widely used signal-related feature [52]. Typically, RSSI measurement estimations depend heavily on the environment, and are also nonlinear. Several techniques make use of RSSI with Wi-Fi technology for indoor localization. As path loss models that are essential for such techniques are also impacted by multipath fading and shadowing effects [27], often indoor site-specific parameters need to be used for these models. Some efforts have been proposed to improve accuracy in such cases; for example, [53] uses pre-measured RSSI contours centered at the receiver to improve localization accuracy with cellular network signals, while [54] employs a fuzzy logic algorithm to improve Wi-Fi RSSI-based localization. In [55, 56], Bluetooth RSS was used to estimate distances and then an extended Kalman filter (EKF) algorithm was applied to obtain 3D position estimates.

Another approach to estimating distance is to use the signal phase (or phase difference) property [57]. As an example, assuming that all transmitting stations emit pure sinusoidal signals that are of the same frequency, with zero phase offset; then the receiver can measure the phase difference between the signals transmitted by the stations, which is a function of its location with respect to the stations. It is possible to use the signal phase approach together with ToA/TDoA or RSSI techniques to fine-tune the location positioning. However, the signal phase approach is susceptible to interference along NLOS paths that can introduce errors.

37.5.2 Fingerprinting

Fingerprinting techniques refer to algorithms that estimate the location of a person or object at any time by matching real-time signal measurements with unique location-specific “signatures” of signals (e.g. Wi-Fi RSSI). Typically, fingerprinting can be performed analytically or empirically.

Analytical fingerprinting, for example, RSSI-based, involves using propagation models such as the radial symmetric free-space path loss model to derive the distance between a radiating source and a receiver by exploiting the attenuation of RSSI with distance. Unfortunately, this simplistic model is rarely applicable in indoor environments, where the signals do not attenuate predictably with the distance due to shadowing, reflection, refraction, and absorption by the indoor building structures. Therefore, other models have been proposed, such as the Indoor Path Loss Model [58] and the Dominant Path Model [59], which takes into account only the strongest path, which is not necessarily identical to the direct path.

Empirical fingerprinting is more commonly used in various indoor localization techniques due to the difficulty in analytically modeling unpredictable multipath effects. There are typically two stages involved in such empirical location fingerprinting: an offline (calibration) stage and an online (run-time) stage. The offline stage involves a site survey in an indoor environment, to collect the location coordinates/landmarks/labels and strengths (or other features) of signals of interest at each location. This procedure of site survey is time consuming and labor intensive. However, such a survey can account for static multipath effects much more easily than with analytical fingerprinting (although dynamic effects, e.g. due to different number of moving people are still problematic and can cause variations in readings for the same location). Several public Wi-Fi APs (and also cellular network ID) databases are readily available [60–63] that can somewhat reduce survey overheads for empirical-fingerprinting-based indoor localization solutions; however, the limited quantity and granularity of fingerprint data for building interiors remains a challenge. In the run-time stage, the localization technique uses the currently observed signal features and previously collected information to figure out an estimated location, with the underlying premise that the locations of interest each have unique signal features.

RSSI-based empirical fingerprinting is used extensively in several indoor localization techniques. Many RSSI-based fingerprinting solutions aim to utilize the existing infrastructure to minimize costs, for example, Wi-Fi APs [65] and GSM/3G/4G cellular networks [66], while a few approaches advocate for custom beacon deployment for RF signal generation to support RSSI-based localization [67, 68]. A GSM cellular network RSSI-based indoor localization system is presented in [66]. Indoor localization based on a cellular network is possible if the locale is covered by several base stations or one base station with strong RSSI received by indoor cellular devices. The approach uses wide signal strength fingerprints, which includes the six strongest GSM cells and readings of up to 29 additional GSM channels, most of which are strong enough to be detected but too weak to be used for efficient communication. The additional channels help improve localization accuracy, with results showing the ability to differentiate between floors in three multi-floor buildings, and achieving median within-floor accuracy as low as 2.5 m in some cases. Typically, Wi-Fi is the most common signal type used in RSSI-based fingerprinting approaches. Figure 37.5 illustrates how different locations often are covered by different Wi-Fi APs or have different signal strength characteristics for the same Wi-Fi AP, as can be seen from the two plots in the figure, allowing for unique fingerprinting and consequently, localization [64].

RADAR [65] was one of the first approaches to use Wi-Fi RSSI for indoor localization. An offline phase was used to measure Wi-Fi AP signal strengths across locations. Signal attenuations due to floors, walls, and other obstructions were also modeled, to improve the accuracy of the system to about 2–3 m in some cases. Several other indoor localization approaches use a similar strategy, for example, PlaceLab [69] and Horus [70]. Probabilistic (Bayesian-network based) methods were employed in [71, 72] to improve the correlation between locations and Wi-Fi RSSI during fingerprinting. In [73], a neural network classifier was used for Wi-Fi RSSI-based location estimation, with a reported error of 1 m with 72% probability. Wi-Fi RSSI fingerprinting has also been widely used in the area of mobile robotics, to determine the location of a mobile robot assuming the availability of inputs from robot-mounted sensors. A Bayesian robot localization algorithm was proposed in [74] that first computed the probability of a robot’s location based on the RSS from nine Wi-Fi APs, and then as a second step exploited the limited maximum speed of mobile robots to refine the results (of the first step) and reject solutions with significant change in the location of the mobile robot. Depending on whether the second step is used or not, 83%

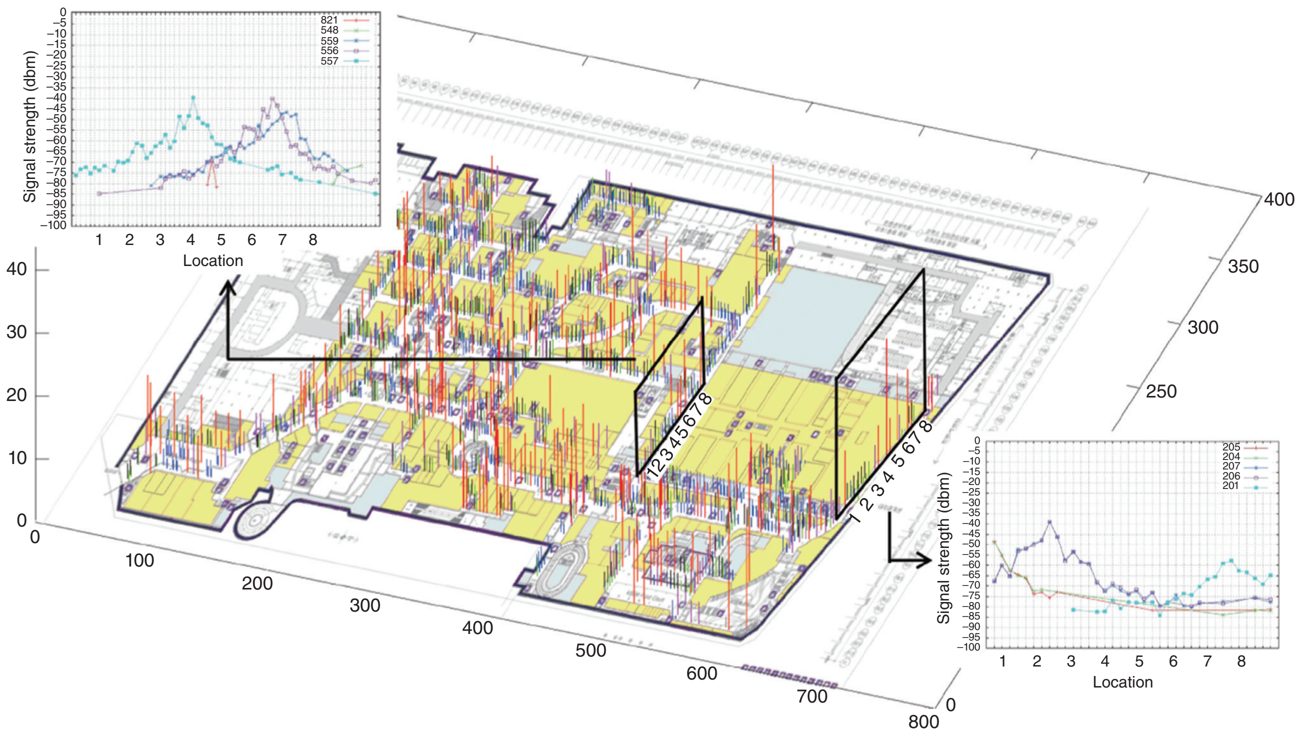

Figure 37.5 Measured accuracy and Wi-Fi signal distributions excerpted for an indoor location. Different locations are often covered by different Wi-Fi access points (APs) and or have different signal strength characteristics for the same Wi-Fi AP (as can be seen from the two plots in the figure), allowing for unique fingerprinting. Black, blue, violet, and red bars on the map represent 3-, 6-, 9-, and more than 12 m error distances, respectively, when using Wi-Fi RSSI fingerprinting for localization [64]. *Source:* Reproduced with permission of IEEE.

and 77% of the time, mobile robots could be located within 1.5 m in their studies. A challenge in these fingerprinting-based techniques is that two distant locations in an open indoor space may have similar signal fingerprints, which can be very problematic during localization [75]. Fingerprinting using the RSS of FM radio signals was proposed in [76]. As FM signals do not carry any timing information, which is a critical factor in range calculation using ToA, TDoA, and AoA methods, RSS-based fingerprinting is the most viable approach when using FM signals for indoor localization. In [77], it was demonstrated that FM and Wi-Fi signals are complementary; that is, their localization errors are independent. Further, experimental results indicated that when FM and Wi-Fi signals were combined to generate fingerprints, the localization accuracy increased by 11% (without accounting for temporal variation) or up to 83% (when accounting for wireless signal temporal variation) compared to when Wi-Fi RSSI only is used as a signature. A few efforts have also proposed Bluetooth RSSI-based localization. Gimbal [67] and iBeacon [68] allow users to instrument their environment with custom Bluetooth-based beacons and then use RSSI values received from these beacons on the user's mobile device for indoor localization. Zigbee RSSI-based localization has also been proposed in several works [2, 78, 79].

Although empirical RSSI-based localization schemes are very popular, a disadvantage of these methods, as mentioned earlier, is that they require labor-intensive surveying of the environment to generate radio maps. Crowdsourcing is one possible solution to simplify the radio map generation process, by utilizing data from multiple users carrying smartphones [80]. Utilizing principles from Simultaneous Localization and Mapping (SLAM) approaches proposed for robot navigation in a priori unknown environments can also be beneficial for quickly building maps of indoor locales [81]. Another challenge is that RSSI readings are susceptible to wireless multipath interference as well as shadowing or occlusions created by walls, windows, or even the human body; thus, in dynamic environments such as a shopping mall with moving crowds, the performance of fingerprinting can degrade dramatically. To overcome multipath interference, a recent effort [82] proposes using the energy of the direct path (EDP), and ignoring the multipath reflections between the mobile client and APs. EDP can improve performance over RSSI because RSSI includes the energy carried by multipath reflections which travel longer distances than the actual distance between the client and the APs.

Several approaches have proposed computer-vision-based fingerprinting [83–86]. These techniques require the mobile subject to either carry a camera or use a camera embedded in a handheld device, such as a smartphone. When the subject moves around in an indoor environment, the camera captures images (visual fingerprints) of the environment, and then determines the subject's position and orientation by matching the images against a database of images with known location. One challenge with such an approach is the high storage capacity required for storing the images of an indoor environment. Significant computing power may be required to perform the image matching, which may be challenging to implement on mobile devices. If subjects are required to carry supporting computing equipment [85], it may impede their mobility. In [87, 88], it is assumed that the illumination intensity and ambient color vary from room to room and thus can be used as fingerprints for room-level localization. Light matching [89] utilizes the position, orientation, and shape information of various indoor luminaires, and models the illumination intensity using an inverse-square law. To distinguish different luminaires, the work relies on asymmetries/irregularities of luminaire placement. Absolute and relative intensity measurements for localization have been proposed in [90, 91], respectively, with knowledge of the receiver orientation assumed to be available to solve for a position. These techniques employ received intensity measurements to extract position information from multiple transmitters using a suitable channel model. Other approaches exploit the directionality of free space optics, where angular information is encoded by discrete emitters (e.g. modulated LED-based beacons) [92–94]. The Xbox Kinect [95] uses continuously projected infrared (IR) structured light to fingerprint and capture 3D scene information with an infrared camera. The 3D structure can be computed from the distortion of a pseudorandom pattern of structured IR light dots. People can be tracked simultaneously up to a distance of 3.5 m at a frame rate of 30 Hz. An accuracy of 1 cm at 2 m distance has been reported.

Some approaches utilize dedicated coded markers or tags in the environment for visual fingerprinting, to help with localization [96–100]. Such approaches can overcome the limitations of traditional vision-based localization systems that rely entirely on natural features in images which often lack of robustness, for example, under conditions with varying illumination. Common types of markers include concentric rings, barcodes, or patterns consisting of colored dots. Such markers greatly simplify the automatic detection of corresponding points, allow determination of the system scale, and enable distinguishing between objects by using a unique code for different types of objects. An optical navigation system for forklift trucks in warehouses was proposed in [96]. Coded reference markers were deployed on ceilings along various routes. On the roof of each forklift, an optical sensor took images that were forwarded to a centralized server for processing. Another low-cost indoor localization system was proposed in [97] that made use of phone cameras and bar-coded markers in the environment. The markers were placed on walls, posters, and other objects. If an image of these markers was captured, the pose of the device could be determined with an accuracy of "a few centimeters." Additional location-based information (e.g. about the next meeting in the room) could also be displayed.

An alternative approach to physically deploying markers is to project reference points or patterns onto the environment. In contrast to systems relying only on natural image features, the detection of projected patterns is facilitated by the distinct color, shape, and brightness of the projected features. For example, in [101], the TrackSense system was proposed, consisting of a projector and a simple webcam. A grid pattern is projected onto plain walls in the camera's field of view. Using an edge detection algorithm, the lines and intersection points are determined. By the principle of triangulation – analogous to stereo vision – the distance and orientation to each point relative to the camera are computed. With a sufficient number of points, TrackSense is able determine the camera's orientation relative to fixed large planes, such as walls and ceilings.

37.5.3 Proximity

Techniques that are based on the presence of the mobile subject in the vicinity of a sensor (with a finite range and analysis capabilities) are referred to as proximity-based localization approaches. The proximity of the mobile subject can be detected via physical contact or by monitoring a physical quantity in the vicinity of the sensor, such as a magnetic field. When a mobile subject is detected by a single sensor, it is considered to be collocated with it. Several proximity-based localization techniques have been implemented, involving IR, RFID, and cell identification (Cell-ID).

One of the first IR-based proximity indoor positioning systems was the Active Badge system [102] designed at AT&T Cambridge in the 1990s. By estimating the location of active badges worn by people in the building, the Active Badge system was able to locate persons in a building. The active badges would transmit a globally unique IR signal every 15 s (with a battery life of 6 months to a year). In each room, one or more sensors were fixed and detected the IR signal sent by an active badge. Using the measured location of the people in the building, the system was able to track the employees, their location (room numbers), and the nearest telephone to reach them. The accuracy of the system was driven by the operating range of the IR sender, which was 6 m. The proximity system based on wearable IR emitters was small, lightweight, and easy to carry; however, the network of fixed sensors deployed across the building had a substantial cost associated with it. Moreover, the update rate of 15 s is too large for real-time localization (navigation).

An RFID system consists of RFID readers (scanners) and RFID tags. The RFID reader is able to read the data from RFID tags that are either passive or active. Passive RFID tags rely on inductive coupling and operate without a battery, reflecting the RF signal transmitted to them from a reader and adding information by modulating the reflected signal. Inductive coupling allows the passive tags to receive sufficient energy in the form of RF waves from the nearby RFID reader to perform signal modulation, to transfer their unique serial ID (or other information) back to the reader. But the range of passive RFID tags is very limited (1–2 m), and the cost of the readers is relatively high. Active RFID tags are small transceivers that can actively transmit their ID (or other additional data) in response to an interrogation. Systems based on active RFID use smaller antennas and have a much longer range (tens of meters). LANDMARC [103] utilizes active RFID-based fixed location reference tags for proximity-based indoor location calibration.

The Cell-ID (or Cell-of-Origin) method is based on the principle of capturing the ID of an anchor node that is generating the RF signal with the highest RSSI, and then identifying the mobile subject's position as having the same coordinates as the anchor node. For example, mobile cellular networks can identify the approximate position of a mobile handset by knowing which cell site the device is using at a given time. Wi-Fi APs can also be used to obtain the ID of the AP with the highest RSSI and perform localization with respect to that AP. In general, the localization accuracy when using cell-ID is quite low (50–200 m), depending on the cell size (or Wi-Fi coverage). The accuracy is often higher in densely covered areas (e.g. urban places) and much lower in rural environments [104].

37.5.4 Dead Reckoning

Dead reckoning refers to the use of sensors that provide location updates, calculated based on the last determined position and incrementing that position based on known or estimated speeds over elapsed time. Position and speed estimation is typically based on IMUs, which include multi-axis accelerometers, gyroscopes, and possibly magnetometers. A disadvantage of dead reckoning is that the inaccuracy of the estimation process is cumulative, so any deviations in the position estimates become larger with time. This is because new positions are calculated entirely from previous positions. Thus, these inertial navigation systems (INSs) are often used to estimate relative rather than absolute location, that is, the change in position since the last update, with some other localization technology (e.g. Wi-Fi fingerprinting) for obtaining periodic position fixes (absolute location estimates).

The first task in an INS that is used for human localization is the identification of steps or strides from the sensor data. A step is the period between two footfalls on opposite feet, whereas a stride is the same quantity but between the same foot. At a minimum, there is a need for accurate step detection and step counting for most INS-based indoor localization approaches. Step cycle detection algorithms detect cycles in the INS data caused by the repetitive motion of walking, which may involve searching for repeating data patterns or for repeating events (e.g. the heel strike). This information can be used for step detection and counting. Figure 37.6 shows an example of extracting the cycle (and thus step counts) by seeking maxima in the mean-adjusted autocorrelation of a sequence of accelerometer magnitude data [105]. Usually a low-pass filter is used on the raw accelerometer data to remove noise as a pre-processing step before the algorithm is used. Filter cutoff frequencies of around 20 Hz retain the step periodicity, although filtering down to 2 or 3 Hz has also been used successfully [106]. In general, the cyclic property of walking is directly reflected in the acceleration trace in the time domain. As heel strikes tend to introduce sharp changes, numerous schemes propose to detect magnitude peaks [107, 108], local variance peaks [109, 110], local minima [111, 112], zero crossings [113], or level crossings [114] (where levels are defined by historical mean and variance) from the low-pass filtered acceleration trace.

Knowing the step count, it is possible to estimate displacement of a mobile subject if information related to the stride length is available. Pedestrians typically have a natural walking pace with a surprisingly constant stride length. However, this natural walking pace is altered when rushing, ambling, or walking with others. The stride length can vary by as much as 40% between pedestrians walking at the same speed, and up to 50% across the range of walking speeds of a pedestrian [106]. Direct measures involving additional sensors can help estimate the length with high accuracy (e.g. using ultrasonic sensors mounted on the front and back of each shoe [115], or electromyography sensors attached to the calf [116]), but such methods also come at high costs and inconvenience to the mobile user. Several studies in human kinematics correlate stride length with step frequency [117–119]. The major observation is that stride length tends to be shorter when walking slowly rather than fast [120]. A simple linear relationship often suffices [121], but the model parameters that are trained offline are specific to walking conditions, such as wearing

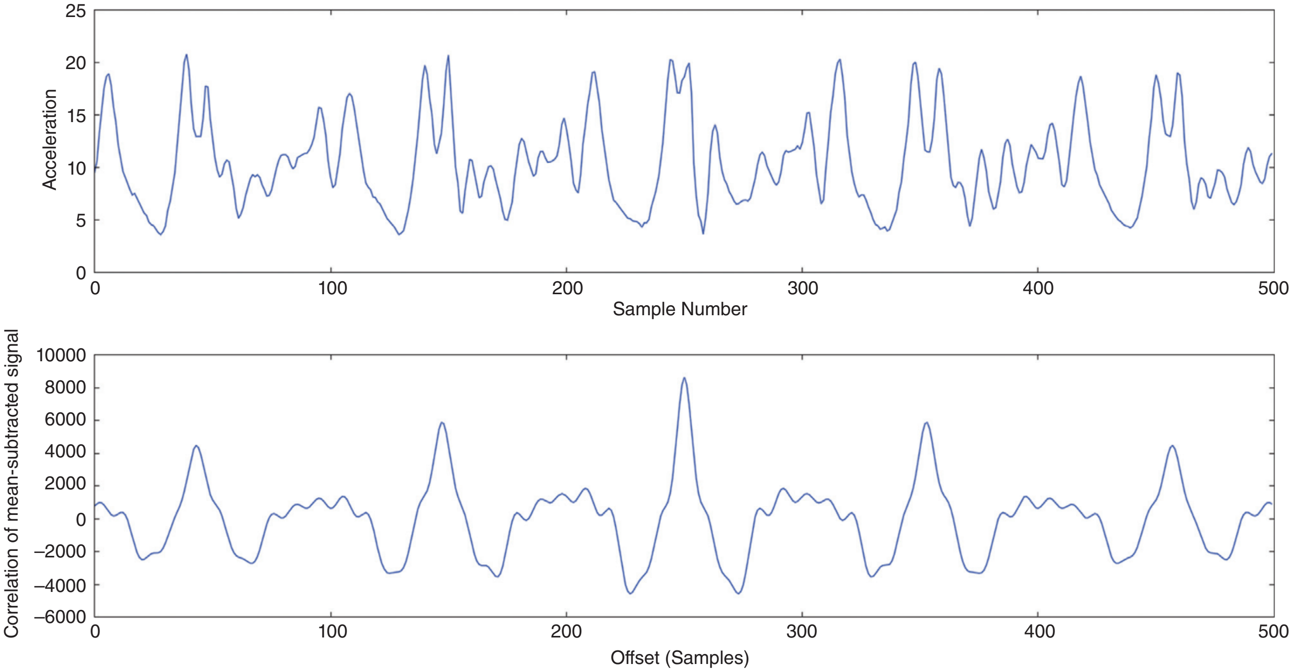

Figure 37.6 Autocorrelation-based step cycle detection. The top graph shows the raw acceleration magnitude during five sample strides. The autocorrelation of the mean-subtracted signal is shown in the bottom graph, with strong peaks associated with each stride [105]. *Source:* Reproduced with permission of IEEE.

sport shoes or high heels [122]. Techniques to detect locomotion modalities (e.g. walking or running) can help construct more elaborated motion models (e.g. adjust stride length estimation or step counting according to varying speeds), to improve localization accuracy. Utilizing accelerometers in smartphones to distinguish different locomotion modalities has been proposed in a few works [123–125].

While accurate stride length improves displacement estimation, the accuracy increase is often marginal as drifts in heading (the direction of motion) typically dominate errors [126]. The heading direction of steps during motion can be obtained with a gyroscope or a compass (magnetometer). Gyroscopes output angular velocities in 3D, which are integrated over time to obtain direction change information. A turn can be detected when the relative orientation measured by a gyroscope changes abruptly. To distinguish between changes due to turns and changes caused by noise, only heading changes exceeding a predefined threshold are determined as turns [127]. A compass can measure the absolute orientation (heading) of the mobile device (e.g. smartphone) with respect to the magnetic north. However, Earth's magnetic field is relatively weak at the surface, and buildings that are filled with metal and conducting wires can overpower the natural signal, leading to local "disturbances" (e.g. location-specific magnetic offsets that can cause heading errors of up to 100^{o} [128]). Some efforts attempt to filter the magnetic offset on consecutive compass readings, to improve accuracy [129]. An increasingly popular solution to overcome the offset is to combine gyroscope and magnetometer readings as the two sensors have complementary error characteristics: gyroscopes provide poor long-term orientation, while magnetometers are subject to short-term orientation errors [130]. In general, multiple types of inertial sensors perceive similar movements during walking, which can be used to overcome errors; for example, a compass value can be considered valid if the readings of the compass and gyroscope in the INS unit experience a correlated trend [111], which can help discard compass values containing a severe magnetic offset.

Today's smartphones include IMUs, and the fact that they are carried by people almost everywhere makes INS-based indoor localization particularly attractive. However, one important challenge is to account for the manner in which the smartphone is carried: in front pockets, back pockets, side pockets, shirt pockets, backpacks, handbags, on belt clips, or in the hand. A few efforts on activity recognition have explored estimating phone placements [125], which may help improve the performance of dead-reckoning-based localization systems. However, studies have shown that even if a smartphone is located in a single location (e.g. trouser pocket), notable errors are accrued (about 14.4% [131]) when estimating distance traveled, compared to foot-mounted ground truth sensors.

37.5.5 Map Matching

Accurate trajectory estimation is a major goal of most indoor localization and navigation systems. A pedestrian trajectory consists of a sequence of step vectors. Techniques that utilize an electronic map to determine the position of a mobile person or object along a trajectory in the context of locations provided on the map are referred to as map matching techniques. The idea of applying electronic maps to adjust a mobile subject's positions has been used in outdoor localization schemes [132]. Similarly, integrating the geometric constraints of floor plans in indoor environments can help improve indoor localization accuracy (e.g. when used in tandem with dead reckoning or Wi-Fi fingerprinting). In general, the overall geometric shape of a mobile subject's trajectory should be similar to that of the floor plan, and any deviations can point toward an error in a localization scheme. Various geometric abstraction models have been proposed for map matching, for example, link-node models [133] and stress-free floor plans [108]. Particle filtering techniques can additionally be used to exclude unlikely positions for mobile subjects, such as obstacles and walls [134, 135].

LiFS [110] is an example of a framework for matching sensor/signal readings to a physical floor plan. First, continuous measurement of acceleration readings and RSS readings is performed with the aid of smartphone users during their routine work and occupancy of buildings. Footsteps are then detected and counted, and these are then used as the inter-fingerprint distance measurements. Feeding the inter-fingerprint distances to a multidimensional scaling (MDS) algorithm results in a high-dimension space called the fingerprint space, where the mutual distances between points (fingerprints) are preserved. The fingerprint space is then mapped to the physical floor plan to associate fingerprints with their corresponding physical locations in the indoor environment. The mapping is achieved by exploring the spatial similarity between the fingerprint space and a transformed floor plan, called the stress-free floor plan. The stress-free floor plan is a space that transforms a normal floor plan into a high-dimension space using MDS, in such a way that the geometrical distances between the points in the new space reflect walking distances instead of straight distances. The rationale behind such transformation is that, due to the presence of obstacles (e.g. walls), the walking distance between two locations is not necessarily equal to the geographical distance between them. LiFS was shown to achieve good performance, with the 95th percentile mapping error being lower than 4 m and an average error of 1.33 m. The radio map generated using LiFS can be used as a starting point for various fingerprint-based localization techniques.

Several other efforts have addressed map matching. In [136], a framework was proposed to combine a backtracking particle filter (BPF) with different levels of building plan detail to improve the indoor localization performance via dead reckoning. Particle filters are able to take into account building plan information during indoor localization with a technique called map filtering [137]. With map filtering, new particles are not allowed to occupy impossible positions given the map constraints. For example, particles are not allowed to cross directly through walls. Particles that transition through such obstacles are deleted from the set of particles or downweighted, as shown in Figure 37.7. BPF further exploits particle trajectory histories to improve upon simple particle filters, by recalculating previous state estimates after invalid particles are detected. In order to enable backtracking, each particle has to remember its state history or trajectory. Mean location estimation errors when using dead reckoning, dead

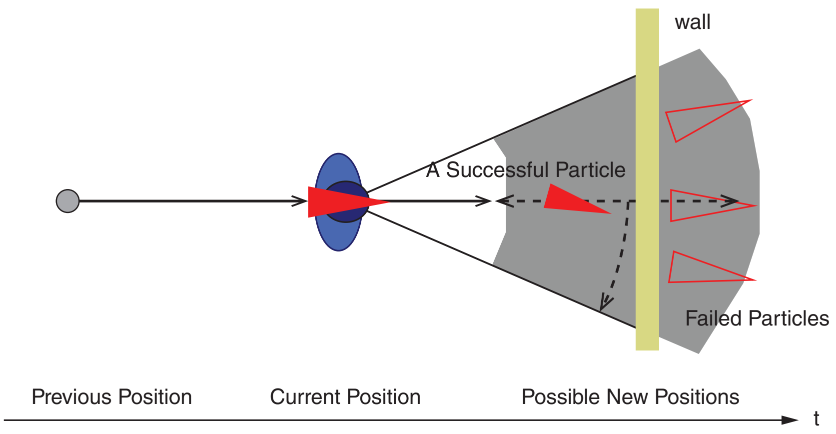

Figure 37.7 Particle transition near obstacles: if a particle tries to move to an impossible location, for example, across walls defined in the map, it will be killed off [136]. *Source:* Reproduced with permission of IEEE.

reckoning with particle filters, and dead reckoning with BPF were shown to be 7.7, 3.1, and 2.6 m, respectively [136].

Predicting the trajectory of a mobile subject can also help reduce ambiguity when using fingerprinting for localization [138]. As an example, displacement and direction information obtained with dead reckoning impose relative geometrical constraints between consecutive location queries along a trajectory. These constraints transform the fingerprint matching from essentially being a point matching process to one that now involves line fitting by embedding the entire trajectory into the radio map. ACMI [139] employs FM broadcast signal fingerprinting for localization, and uses trajectory predictions for localization accuracy improvement. Experimental results have demonstrated that localization errors decreased from 10–18 m to 6 m, along with an increase in the room identification accuracy from 59% to 89%, when trajectory matching was used.

Certain indoor landmarks and contexts also possess distinctive sensor signatures. For example, accelerometer readings on an elevator exhibit a sharp surge and drop at the start and the stop of the elevator. An investigation of such unique acceleration patterns of stairs, elevators, escalators, and so on, was performed in [111], and it was concluded that if the locations of these structures were known previously, they could serve as landmarks to improve indoor localization accuracy (e.g. to overcome dead reckoning drifts).

The techniques discussed so far address the problem of positioning a mobile subject in an indoor environment with a known map or landmarks. A more difficult problem that has been studied by the robotics community involves SLAM for robots to navigate in a priori unknown environments [81]. In SLAM, a moving robot explores its environment and uses its sensor information and odometry control inputs to build a "map" of landmarks or features, while also estimating its position in reference to the map [140]. Odometry refers to the control signals given to the driving wheels of the robot. Simple integration of these odometry signals can be considered to be a form of dead reckoning. EKF-SLAM [81] employs an EKF to represent the large joint state space of robot pose (position and orientation) and all landmarks identified so far. The approach known as FastSLAM uses a Rao-Blackwellized particle filter (RBPF) [141] where each particle effectively represents a pose and set of independent compact EKFs for each landmark. The conditioning on a pose allows the landmarks to be estimated independently, leading to lower complexity. SLAM implementations for robot positioning always build on sensors and robot odometry that are readily available on robot platforms. The sensors can consist of laser rangers or a single or multiple cameras mounted on the robot platform, and the features are extracted from the raw sensor data. SLAM is considered to be a "hard" problem, in contrast to the two easier special cases: positioning in an environment with known landmarks or building a map of features given the true pose of the robot. In [140], a SLAM approach was proposed for learning building paths/maps automatically by observing data from a mobile subject, which can either be used to localize the subject or provide maps for others. The approach made use of inertial sensors together with principles derived from the FastSLAM framework [141] and dynamic Bayesian networks.

37.5.6 Hybrid Techniques

Each of the five classes of techniques discussed in this section so far has drawbacks when used in isolation. Therefore, a recent trend has been to combine various techniques together, to successfully bridge the differences among different types of techniques and overcome the limitations of a single type of localization strategy to improve accuracy. Some of these hybrid techniques can also be used in both indoor and outdoor environments.

37.5.6.1 GPS-Based Techniques

The wireless-assisted GPS (A-GPS) was pioneered by SnapTrack (now part of Qualcomm) and can be used for indoor locales. The approach leverages the cellular network together with GPS signals. Many cellular network towers have GPS receivers (or a base station nearby), and those receivers often constantly collect satellite information to detect the same satellites as cellular phones. This data is sent to the cellular phone (when requested), speeding up the time to first fix (TTFF; to acquire the orbit and clock data of relevant GPS satellites), which on a mobile device without assistance can take a long time (minutes) in some cases. Not only does the TTFF get reduced, but the approach can enable localization in indoor environments, where the GPS signals detected by the cellular phone are often very weak, with accuracies ranging from 5–50 m.

37.5.6.2 Techniques Fusing RF Signals with Dead Reckoning

Several techniques have been proposed that combine inertial sensor readings with data from RF signals for indoor localization. For example, in [142], an indoor localization framework is proposed that combines Wi-Fi RSSI fingerprint-based positioning and dead reckoning data, with the help of a Hidden Markov Model (HMM). The dead reckoning consists of an accelerometer-driven step length estimation and a magnetic-field-based heading calculation. While dead reckoning achieves high precision over short time periods, it suffers from error accumulation over longer durations. In contrast, the positioning error with Wi-Fi fingerprints does not increase with time, but has less accuracy over the short term. Thus, the sensor data fusion of dead reckoning and Wi-Fi positioning yields a synergistic effect, resulting in higher robustness and precision. The proposed HMM is based on the discrete positions of the Wi-Fi fingerprints as the hidden states and the RSSI Wi-Fi measurements as the observable states (a Markov model is called hidden if it contains an underlying stochastic process that is not directly observable, but can be observed through another stochastic process [143]). State transitions depend on movement inputs from dead reckoning. The use of the HMM makes it possible to deal with ambiguities resulting from Wi-Fi fingerprinting. The HMM approach is also computationally less expensive than the filtering schemes used in other efforts. For instance, particle filters are used in [144, 145] for the integration of Wi-Fi positioning and dead reckoning, but the particle filters have high computational costs, depending on the number of particles computed. Kalman filters and EKFs are also not well suited for such sensor data fusion, as the assumption of Gaussian distributions is in conflict with the ambiguous outputs of Wi-Fi fingerprinting algorithms. In [146], another HMM-based indoor localization approach was proposed that fused Wi-Fi fingerprints and dead reckoning. In the work, the HMM is augmented to take into account vector (instead of scalar) observations, and prior knowledge about user mobility drawn from personal electronic calendars (e.g. a calendar entry of "meeting in conference room C103A at 1 p.m." can be useful to estimate the probability associated with positioning of the subject in room C103A). An extension of the Baum–Welch algorithm [147] is used to learn the parameters of the augmented HMM.

The LearnLoc framework [148] fuses Wi-Fi fingerprinting with dead reckoning to create a low-cost, infrastructure-less indoor navigation solution. The framework adapted and enhanced three machine learning techniques that took inputs from inertial sensors and Wi-Fi fingerprinting to make predictions about indoor location on a map in the presence of noise (e.g. due to incorrect sensor readings). The three supervised learning algorithms used to assist with indoor localization were based on KNN, linear regression (LR), and nonlinear regression with neural networks (NL-NN). Regression-based variants of these algorithms were used instead of the more traditional classification-based variants. This is because a classification technique requires dividing the entire indoor map area into a fine-grained grid for classification toward accurate localization, which creates a prohibitively large input space that is impractical to process on resource-constrained mobile devices. For instance, their efforts to implement SurroundSense [149] that proposes an SVM-based classification technique for indoor localization for real-time localization on a smartphone were not successful because of the large memory footprint and slow performance (taking close to a minute for each prediction) with the approach. In contrast, regression can allow fast predictions with much lower resource demands, which is what is needed for real-time indoor localization with mobile devices. Figure 37.8(a) shows a detailed look at the predicted paths by the KNN-based LearnLoc variant for different Wi-Fi scan intervals. Not surprisingly, the lowest Wi-Fi scan interval (1 s) results in the highest accuracy, but also incurs a very high energy consumption overhead because scanning is performed very frequently (as can be seen by the high density of green dots that represent Wi-Fi scan instances in Figure 37.8(a) for the 1 s interval case). As the Wi-Fi scan interval increases, the paths traced start deviating notably from the actual path, and the estimation errors increase. A scan interval of 4 s was chosen for all three LearnLoc variants to balance energy consumption on a smartphone with localization accuracy. Figure 37.8(b) summarizes the paths traced by the three LearnLoc variants and the Footpath [150] inertial navigation (Inertial_Nav) technique. It can be observed that the path traced by the Inertial_Nav technique greatly deviates from the actual path due to error accumulation over time. The sequence alignment algorithm in the Inertial_Nav technique aims to overcome this error with periodic recalibration, but is not always successful in doing so. For the LearnLoc variants, the green points in the figure indicate instances where a Wi-Fi scan was performed. The KNN variant performs best, with an average error of 2.23 m. The accuracy can be improved much further if scan intervals smaller than 4 s are chosen. LearnLoc is one of the very few techniques to explore trade-offs between energy consumption and accuracy during indoor localization, and also consider realistic resource constraints when devising algorithms meant for execution on resource-constrained mobile devices. A more recent work, CNNLoc [151], improves upon LearnLoc by using a more sophisticated convolutional neural network (CNN) machine learning algorithm deployed on smartphones.

An indoor localization system was proposed in [152] that does not depend on a centrally established database of signals, nor on a pre-supplied building map. It combines inertial sensor data (from the accelerometer and compass), as well as RSSI measurements from Wi-Fi and GSM cellular radios. It divides the building area into a regular grid and applies a SLAM technique to correct any observed drift. Apple's WiFiSLAM system [153] also utilizes the above combination of signals and sensors for indoor localization. SignalSLAM [154] extends these efforts by combining readings from many more sources: time-stamped Wi-Fi and Bluetooth RSS, 4G LTE Reference Signal Received Power (RSRP), magnetic field magnitude, near-field

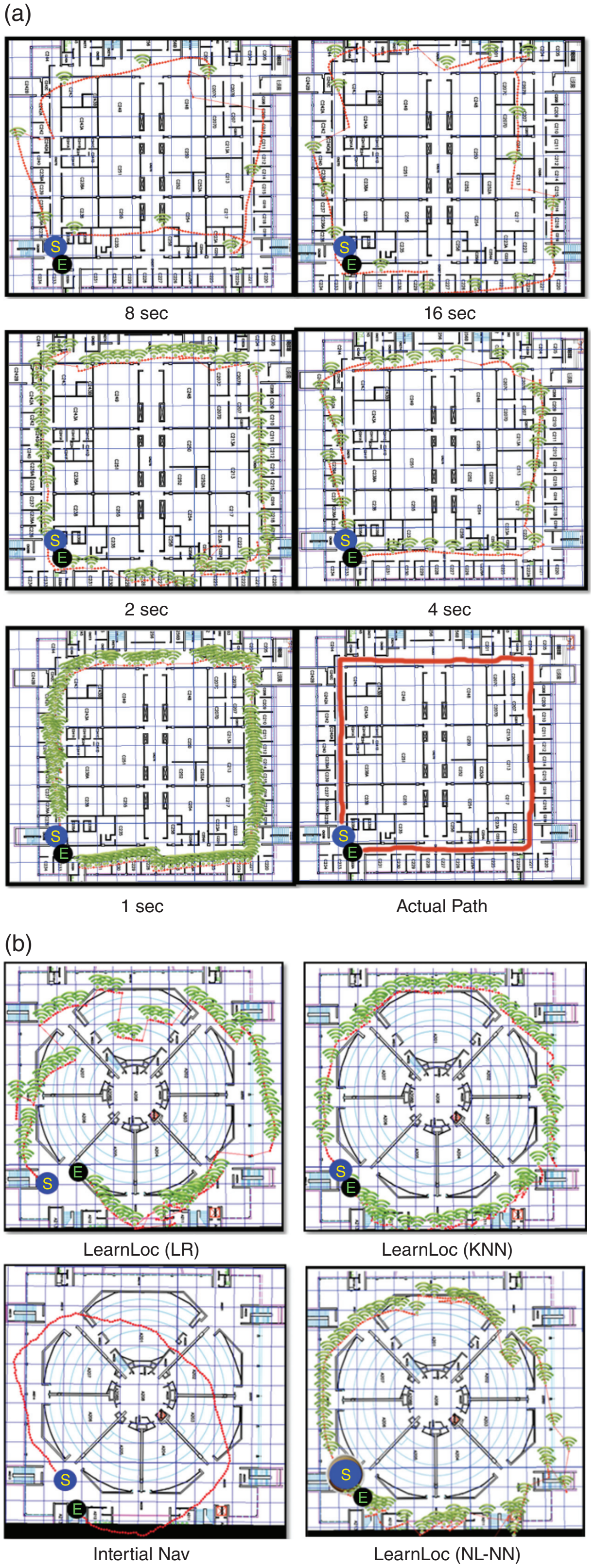

Figure 37.8 (a) Paths traced for various Wi-Fi scan intervals for LearnLoc using K-nearest neighbor (KNN) along the Clark L2 South path; green dots represent an instance of a Wi-Fi scan along the path; (b) paths traced by indoor localization techniques along the Clark L2 North building benchmark path [148]. *Source:* Reproduced with permission of IEEE.

communication (NFC) readings at specific landmarks, and dead reckoning based on inertial data. The location of a mobile subject is resolved by using a modified version of GraphSLAM optimization [155] of the user's poses with a collection of absolute location and pairwise constraints that incorporate multi-modal signal similarity.

37.5.6.3 Techniques Fusing RF Signals with Other Signals

Many techniques propose to combine RF signal data with readings from other sources beyond inertial sensors. SurroundSense [149] utilizes fingerprints of a location based on RF (GSM, Wi-Fi) signals as well as ambient sound, light, color, and the layout-induced user movement (detected by an accelerometer). Cameras, microphones, and accelerometers on a Wi-Fi-enabled Nokia N95 phone were used to sense the fingerprint information. The sensed values are recorded, pre-processed, and transmitted to a remote SurroundSense server. The goal of pre-processing on the phone is to reduce the data volume that needs to be transmitted. Once the sensor values arrive at the server, they are separated by the type of sensor data (sound, color, light, Wi-Fi, accelerometer) and distributed to different fingerprinting modules. These modules perform a set of appropriate operations, including color clustering, light extraction, and feature selection. The individual fingerprints from each module are logically inserted into a common data structure, called the ambience fingerprint, which is forwarded to a fingerprint matching module for localization. Support vector machines (SVMs), color clustering, and other simple methods were used for location classification.

The Acoustic Location Processing System (ALPS) [156] combines BLE transmitters with ultrasound signals to improve localization accuracy and also help users configure indoor localization systems with minimal effort. ALPS consists of time-synchronized beacons that transmit ultrasonic chirps that are inaudible to humans, but are still detectable by most modern smartphones. The phone uses the TDoA of chirps to measure distances. ALPS uses BLE on each node to send relevant timing information, allowing for the entire ultrasonic bandwidth to be used exclusively for ranging. The platform requires a user to place three or more beacons in an environment and then walk through a calibration sequence with a mobile device where they touch key points in the environment (e.g. the floor and the corners of the room). This process automatically computes the room geometry as well as the precise beacon locations without needing auxiliary measurements. Once configured, the system can track a user's location referenced to a map. Other techniques such as SmartLOCUS [157] and Cricket [158] also use a combination of RF and ultrasound technologies, where the TDoA between RF and ultrasound signals (generated by wall- and ceiling-mounted beacons) is used to measure distance and localize mobile subjects.

Radianse [159] and Versus [160] use a combination of RF and IR signals to perform location positioning. Their tags emit IR and RF signals containing a unique identifier for each person or asset being tracked. The use of RF allows coarse-grain positioning (e.g. floor level granularity), while the IR signals provide additional resolution (e.g. room granularity). The EIRIS local positioning system [161] uses an IRFID triple technology that combines IR, RF (UHF), and LF (RF low-frequency transponder) signals. It combines the advantages of each technology, that is, the room location granularity of IR, the wide range of RF, and the tailored range sensitivity of LF.

The CUPID2.0 indoor positioning system [162] combines ToF-based localization with signal strength information to improve indoor localization with Wi-Fi RF signals. The proposed architecture consists of a location server and multiple Wi-Fi APs, each of which talks to the mobile device. ToA-based trilateration methods are used to determine the device location. In particular, the time of flight of the direct path (TFDP), as calculated from the data-ACK exchange between the AP and the device, is used for distance estimation. TFDP is then combined with measurements of signal strength, particularly the EDP [82], to improve accuracy and also ensure scalability. The system was implemented, deployed, and analyzed at six cities across two different continents for more than 14 months with 40 different mobile devices and more than 2.5 million location fixes, and was shown to achieve a mean localization error of 1.8 m.

37.5.6.4 Techniques Fusing Dead Reckoning with Non-RF Signals

A few indoor localization techniques combine inertial sensors with non-RF signals. In [163], the IDyLL indoor localization system is proposed that combines dead reckoning with light measurements from photodiode sensors on smartphones. Typical luminaire sources (including incandescent, fluorescent, and LED) are often uniquely (sometimes evenly) spaced in many indoor environments. Moreover, most smartphones have light sensors (photodiodes) for automatic brightness adjustment that can theoretically sample at a high rate (e.g. 1.17 MHz for APDS-9303 on Nexus 5 and 7 devices), although they are often constrained either by the hardware interface or the OS-level support to a few hertz to up to 100 Hz. IDyLL samples the light sensors at 10 Hz, and uses an illumination peak detection algorithm to gather light readings. The readings are combined with those obtained from inertial sensors, as well as knowledge of the floor map and luminary placement, to achieve fine-grained indoor localization. The approach in [164] combines dead reckoning, laser scanners, and image-based localization, all integrated in a human-carried backpack which can be used to generate 3D models of complex indoor environments. The locations are determined from data capture based on two laser scanners and an inertial measurement unit. The localization performance could be improved by making use of camera images that have been taken in an offline phase. The images can be used to refine the six parameters of the camera pose and improve the quality of the 3D textured model.

37.6 Open Research Issues

Indoor localization systems are steadily becoming more mature, but there are still several challenges that must be addressed, as outlined in [165], which discusses the experiences and lessons learned from Microsoft's indoor localization competition. Below we provide a holistic overview of some of the key open research challenges in the area of indoor localization.

Evaluation methodologies. The outcomes of studies to determine the efficacy of an indoor localization solution can be impacted by several factors, such as the building type and size, construction materials and layout along the analyzed indoor paths, lengths of the indoor paths, characteristics of test subjects, and the test procedure followed (including duration and the degree of "natural" activity) [126]. There is currently very little consensus on how to evaluate various indoor localization solutions, which hinders an appropriate comparison. Because of stark differences in the above-listed factors (that are also not often clearly presented) across evaluation studies, claims made in literature about the accuracy of a particular solution are often difficult to reproduce. Many solutions in literature are content with a very simple proof-of-concept evaluation, with contrived walking tests along indoor locales that are limited in scope (e.g. testing with a single subject). Moreover, manually evaluating indoor localization technologies is a tedious and time-consuming process. It may be possible to reduce evaluation overhead with an automated robot-based benchmarking platform that can also improve the fidelity of the evaluation process.

Evaluation metrics. Indoor localization solutions in the literature are compared using various metrics such as the average location error, RMSE, 95th percentile, and so on. However these metrics often do not capture real-world variations. For instance, [165] discussed how certain indoor locales were very easy to localize by even the simplest of techniques; however, some other points were extremely difficult to accurately localize. The way in which evaluation points are selected and weighted in the evaluation metric is therefore crucial, and a lot of work needs to be done in terms of standardizing the evaluation metrics of indoor localization technologies to properly capture these parameters.

Sensor positioning. Many indoor localization techniques rely on readings from sensors that are carried by the person or object to be tracked. It is possible for the orientation and position of the sensors to change over time and across tracked subjects; for example, a person may carry a smartphone with inertial sensors in different pockets, or hold it in their hand when moving. There may also be other types of positioning issues; for example, the direction a smartphone is facing may be different from the direction the subject is moving. Indoor localization techniques should take such factors into account and compensate for positioning variations. For better accuracy in the estimation of the step length or even the heading direction, it may be preferable to use foot-mounted sensors [166]; however, this usually comes at the cost of user inconvenience.

Sensor calibration. Many of the sensors used for indoor localization have an inherent bias and variations in sensitivity to environmental factors. In particular, the low-cost and compact MEMS inertial sensors found in smartphone IMUs have inferior sensor screening, installation error calibration, cross axis error calibration, zero point correction, temperature drift compensation, and so on, compared to the more accurate (and thus bulkier and more expensive) IMUs used in unmanned aerial vehicles (UAVs) and other industrial applications [138]. The IMU sensors must therefore be individually re-calibrated when in use, to avoid drifts in outputs that result in increasing errors over time.

Battery power. Indoor localization techniques that rely on mobile devices carried by moving subjects need to be aware of the battery life constraints of the device. If excessive computation or sensing is performed with the mobile device, the battery of the mobile device can drain quickly, and this is an extremely undesirable scenario, especially if it happens during navigation. For example, if smartphones are utilized for indoor localization, care should be taken to limit the use of CPU, GPU, or DSP processing; wireless radio modules (e.g. Wi-Fi, 4G/5G cellular, GPS); and inertial sensors, as all of these when used continuously or in combination can cause the smartphone battery to drain very quickly. Techniques to optimize energy efficiency in mobile devices [167–170] will be key to achieving cost-effective and practical indoor localization solutions.

Processing capability and memory constraints. Many indoor localization techniques rely on algorithms that must be run on resource-constrained mobile devices carried by the moving subject. For example, many techniques require the use of machine learning algorithms, image processing, signal processing, bandpass filters, peak detectors, autocorrelators or particle filters, and so on. In general, mobile devices have limited computational capabilities, and therefore algorithms that are shown to work correctly on laptops or desktops may not run fast enough on mobile devices for reasonable indoor localization (especially for navigation situations). In some cases, the limited memory in mobile devices may restrict the types of algorithms that can be deployed. Even though over time mobile devices such as smartphones are becoming more powerful (with the size of integrated memory also increasing steadily), the sophistication of localization algorithms has also increased over time, driven by the need for greater levels of accuracy. Thus, limits on processing capabilities and memory with mobile devices cannot be overlooked.

Portability. Techniques for indoor localization that require mobile devices to be carried by the moving subject must ensure that such devices are not heavy, too large, or cumbersome to carry. For instance, requiring subjects to wear foot-mounted/strapped sensors is inconvenient and unlikely to result in a solution that is accepted by a large population of users. Similarly, using proprietary wireless signals that require custom (and possibly bulky) hardware to be attached to smartphones for indoor localization may not be ideal for many people who carry smartphones in their pockets. Thus, portability concerns cannot be ignored, as they can make the crucial difference between indoor localization solutions being widely accepted or largely ignored.

Device heterogeneity. Indoor localization techniques must be able to cope with the heterogeneity of devices on which they may eventually be deployed. Such heterogeneity can be a function of differences in models/vendors of Wi-Fi, IMU, or other wireless/sensor interfaces used across devices. These differences can cause significant accuracy variations when deploying indoor localization techniques across devices. For example, analysis in [171] shows a localization error of as much as 8× due to mobile device heterogeneity. Approaches such as the SHERPA framework [172], which proposes heterogeneity-resilient fingerprint pattern matching, are needed to enable heterogeneity resilience.

Initialization and deployment costs. Many indoor localization techniques require an initialization phase, for example, training machine learning algorithms, war-driving (i.e. site surveying involving searching for

Wi-Fi wireless networks with a moving vehicle or person to create a map of Wi-Fi APs in an area), or calibrating sensors. Other techniques may require infrastructure enhancements, for example, deploying custom radio beacons across an indoor environment, before the technique can be used for localization. All such initialization is time consuming and usually has costs associated with it. Care needs to be taken to ensure that the initialization phases of localization techniques are short and manageable; and that deployment costs for any custom components are not too high. Novel techniques, such as crowdsourcing with multiple users and heterogeneous devices to create radio maps for fingerprinting-based indoor localization, and variants of the SLAM techniques discussed earlier, can significantly reduce initialization time and costs. Such approaches can overcome limitations that arise due to changing infrastructure, for example, adding or removing of Wi-Fi APs in indoor environments over time.

Application-domain specific requirements. The requirements from indoor localization solutions vary quite significantly across application domains. A few studies have quantified acceptable values for localization performance metrics (Section 37.3) across application domains. In [173], the requirements for indoor localization for the mass market were discussed, emphasizing the use of standard devices (e.g. smartphones) and existing infrastructure (e.g. Wi-Fi APs) without significant supplementary sensors, beacons, or additional wearable components. In [174], indoor localization requirements for underground construction sites are discussed, with an emphasis on high accuracy (~centimeter level). In [175], indoor localization requirements for enforcement officers, firefighters, and military personnel are presented, with an emphasis on encrypted communication, uncertainty estimation, fast real-time response, and robustness of devices.

References

1 R. Mautz, “Indoor Positioning Technologies,” 2012.

2 A. Molisch, “Ultrawideband propagation channels,” *Proceedings of IEEE, Special Issue on UWB*, Vol. **97**, pp. 353–371, 2009.

3 A. Bensky *Wireless Positioning Technologies and Applications*, Artech House Publishers, 2007, **311** p.

4 Z. Farid, R. Nordin, and M. Ismail, “Recent advances in wireless indoor localization techniques and system,” *Journal of Computer Networks and Communications*, 2013.

5 https://en.wikipedia.org/wiki/Global_Positioning_System,2016.

6 F. Seco, A. R. Jim’enez, C. Prieto, J. Roa, and K. Koutsou, “A survey of mathematical methods for indoor localization,” in *Proceedings of the 6th IEEE International Symposium on Intelligent Signal Processing (WISP ‘09)*, August 2009, pp. 9–14.

7 O.A. Aider, P. Hoppenot, and E. Colle, “A model-based method for indoor mobile robot localization using monocular vision and straight-line correspondences,” *Robotics and Autonomous Systems*, Vol. **52**, No. 2-3, pp. 229–246, August 31, 2005.

8 V.A. Petrushin, Gang Wei, and A.V. Gershman, “Multiple-camera people localization in an indoor environment,” *Knowledge and Information Systems*, Vol. **10**, No. 2, pp. 229–241, August 2006.

9 IEC, “Safety of laser products – part 1: Equipment classification and requirements,” Technical Information Report 2nd edition, International Electrotechnical Commission, 2007.

10 C. Ciavarella and F. Paterno, “The design of a handheld, location-aware guide for indoor environments,” *Personal and Ubiquitous Computing*, Vol. **8**, No. 2, pp. 82–91, 2004.

11 J. Xiao, Z. Liu, Y. Yang, D. Liu, and H. Xu, “Comparison and analysis of indoor wireless positioning techniques,” in *Proceedings of the International Conference on Computer Science and Service System (CSSS ‘11)*, June 2011, pp. 293–296.

12 J. Torres-Solis, T.H. Falk, and T. Chau, “A review of indoor localization technologies: towards navigational assistance for topographical disorientation,” in *Ambient Intelligence* (ed. F.J.V. Molina), INTECH Open Access Publisher, 2010.

13 P. Vorst, J. Sommer, C. Hoene et al., “Indoor positioning via three different RF technologies,” in *Proceedings of the 4th European Workshop on RFID Systems and Technologies (RFID SysTech ‘08)*, June 2008, pp. 1–10.

14 X. Zhao, Z. Xiao, A. Markham, N. Trigoni, and Y. Ren, “Does BTLE measure up against WiFi? A comparison of indoor location performance,” in *European Wireless 2014; 20th European Wireless Conference; Proceedings of*, VDE, May 2014, pp. 1–6.

15 M. Philipose, J.R. Smith, B. Jiang, A. Mamishev, S. Roy, and K. Sundara-Rajan, “Battery-free wireless identification and sensing,” *IEEE Pervasive Computing*, Vol. **4**, No. 1, pp. 37–45, January/March 2005.

16 M. Vossiek, L. Wiebking, M. Gl¨anzer, D. Mastela, and M. Christmann, “Wireless local positioning—concepts, solutions, applications,” in *Proceedings of the IEEE Radio and Wireless Conference (RAWCON ‘03)*, August 2003, pp. 219–224.

17 M. Rabinowitz and J.J. Spilker Jr., "A new positioning system using television synchronization signals," *IEEE Transactions on Broadcasting*, Vol. **51**, No. 1, pp. 51–61, March, 2005.

18 H. Sheng, L. Jian-jun, X. Ying, G. Jun-wei, and B. Hae-Young, "Hybrid location determination technology within urban and indoor environment based on cell-id and path loss," *Journal of Chongqing University of Posts and Telecommunication (Natural Science Edition)*, Vol. **16**, No. 5, 46–49, October 2004.

19 A. Popleteev, V. Osmani, and O. Mayora, "Investigation of indoor localization with ambient FM radio stations," in *Proceedings of the IEEE International Conference on Pervasive Computing and Communications (PerCom '12)*, 2012.

20 AAMI, "Technical information report (tir) 18, guidance on electromagnetic compatibility of medical devices for clinical/biomedical engineers," Technical Information Report TIR-18, Association for the Advancement of Medical Instrumentation, Arlington, Virginia, 1997.

21 M. Hazas and A. Hopper, "Broadband ultrasonic location systems for improved indoor positioning," *IEEE Transactions on Mobile Computing*, Vol. **5**, No. 5, pp. 536–547, September/October, 2006.

22 R. Casas, D. Cuartielles, A. Marco, H.J. Gracia, and J.L. Falc´o, "Hidden issues in deploying an indoor location system," *IEEE Pervasive Computing*, Vol. **6**, No. 2, pp. 62–69, 2007.

23 R.J. Orr and G.D. Abowd, "The smart floor: A mechanism for natural user identification and tracking," in *CHI '00: CHI '00 Extended Abstracts on Human Factors in Computing Systems*, New York, NY, USA, 2000, ACM, pp. 275–276.

24 G. Retscher, "Test and integration of location sensors for a multi-sensor personal navigator," *Journal of Navigation*, Vol. **60**, No. 1, pp. 107–117, January 2007.

25 F. Evennou and F. Marx, "Advanced integration of WiFi and inertial navigation systems for indoor mobile positioning," *EURASIP Journal on Applied Signal Processing*, Vol. **2006**, No. 17, 11 pp., 2006.

26 M. Vossiek, M. Wiebking, L. Gulden, P. Weighardt, and J. Hoffmann, "Wireless local positioning—Concepts, solutions, applications," in *Proceedings of the IEEE Wireless Communications and Networking Conference, August* 2003, pp. 219–224.

27 H. Liu, H. Darabi P. Banerjee, and J. Liu, "Survey of wireless indoor positioning techniques and systems," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, Vol. **37**, No. 6, 1067–1080, November 2007.

28 R.J. Fontana, E. Richley, and J. Barney, "Commercialization of an ultra wideband precision asset location system," in *Proceedings of the IEEE Ultra Wideband Systems Technologies Conference, Reston, Virginia, November* 2003, pp. 369–373.

29 B.D. Van Veen and K.M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Magazine*, Vol. **5**, No. 2, pp. 4–24, April 1988.

30 F. Belloni et al., "Angle-based indoor positioning system for open indoor environments," 6th Wksp. Positioning, Navigation and Communication, WPNC 2009, March 2009, pp. 261–265.

31 P. Stoica and R.L. Moses, *Introduction to Spectral Analysis*, Englewood Cliffs, NJ: Prentice-Hall, 1997.

32 B. Ottersten, M. Viberg, P. Stoica, and A. Nehorai, "Exact and large sample ML techniques for parameter estimation and detection in array processing," in Radar Array Processing (eds. S.S. Haykin, J. Litva, and T.J. Shepherd), pp. 99–151, New York: Springer-Verlag, 1993.

33 Z. Yang, Z. Wang, J. Zhang, C. Huang, and Q. Zhang, "Wearables can afford: Light-weight indoor positioning with visible light," in *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services*, ACM, May 2015, pp. 317–330.

34 Y.-S. Kuo, P. Pannuto, K.-J. Hsiao, and P. Dutta, "Luxapose: Indoor positioning with mobile phones and visible light," *MobiCom '14, New York, NY*, USA, ACM, 2014, pp. 447–458.

35 D. Hauschildt and N. Kirchhof, "Advances in thermal infrared localization: Challenges and solutions," *Proceedings of the 2010 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, September 15–17, 2010 Campus Science City, ETH Zurich, Switzerland, 2010.

36 K. Atsuumi and M. Sano, "Indoor IR Azimuth sensor using a linear polarizer," *Proceedings of the 2010 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, September 15–17, 2010 Campus Science City, ETH Zurich, Switzerland, 2010.

37 B. Fang, "Simple solution for hyperbolic and related position fixes," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. **26**, No. 5, pp. 748–753, September 1990.

38 X. Li, K. Pahlavan, M. Latva-aho, and M. Ylianttila, "Comparison of indoor geolocation methods in DSSS and OFDM wireless LAN," in *Proceedings of the IEEE Vehicular Technology Conference*, Vol. **6**, September 2000, pp. 3015–3020.

39 N.S. Correal, S. Kyperountas, Q. Shi, and M. Welborn, "An ultrawideband relative location system," in *Proceedings of the IEEE Conference on Ultra Wideband Systems and Technology*, November 2003, pp. 394–397.

40 D. Dardari et al., “Ranging with ultrawide bandwidth signals in multipath environments,” *Proceedings of IEEE*, Vol. **97**, No. 2, February 2009, pp. 404–426.

41 S. Gezici, Z. Tian, G. V. Giannakis, H. Kobaysahi, A.F. Molisch, H.V. Poor, and Z. Sahinoglu, “Localization via ultra-wideband radios: A look at positioning aspects for future sensor networks,” *IEEE Signal Processing Magazine*, Vol. **22**, No. 4, pp. 70–84, July 2005.

42 R.J. Fontana, “Recent system applications of short-pulse ultra-wideband (UWB) technology,” *IEEE Transactions on Microwave Theory and Techniques*, Vol. **52**, No. 9, pp. 2087–2104, September 2004.

43 Active Bat website, http://www.cl.cam.ac.uk/research/dtg/attarchive/bat/

44 Hexamite: http://www.hexamite.com/

45 S. Venkatraman and J. Caffery, “Hybrid TOA/AOA techniques for mobile location in non-line-of-sight environments,” *IEEE WCNC 2004*, Vol. **1**, pp. 274–278, March 2004.

46 C. Yang and H.-R. Shao, “WiFi-based indoor positioning,” *IEEE Communications Magazine*, Vol. **53**, No. 3, 150–157, 2015.

47 D. Zhang, F. Xia, Z. Yang, L. Yao, and W. Zhao, “Localization technologies for indoor human tracking,” in *Proceedings of the 5th International Conference on Future Information Technology (FutureTech ‘10)*, May 2010.

48 AeroScout. [Online]. Available: http://www.aeroscout.com/

49 Wherenet, https://www.zebra.com/us/en/solutions/location-solutions/enabling-technologies/wherenet.html

50 J. Werb and C. Lanzl, “Designing a position system finding things and people indoors,” *IEEE Spectrum*, Vol. **35**, No. 9, pp. 71–78, September 1998.

51 A. Stelzer, K. Pourvoyeur, and A. Fischer, “Concept and application of LPM – A novel 3-D local position measurement system,” *IEEE Transactions on Microwave Theory and Techniques*, Vol. **52**, No. 12, pp. 2664–2669, 2004.

52 S. Gezici, “A survey on wireless position estimation,” *Wireless Personal Communications*, Vol. **44**, No. 3, pp. 263–282, 2008.

53 J. Zhou, K.M.-K. Chu, and J.K.-Y. Ng, “Providing location services within a radio cellular network using ellipse propagation model,” in *Proceedings of the 19th International Conference on Advanced Information Networking and Application*, March 2005, pp. 559–564.

54 A. Teuber and B. Eissfeller, “A two-stage fuzzy logic approach for wireless LAN indoor positioning,” in *Proceedings of the IEEE/ION Position Location Navigation Symposium*, Vol. **4**, April 2006, pp. 730–738.

55 A. Kotanen, M. Hannikainen, H. Leppakoski, and T.D. Hamalainen, “Experiments on local positioning with Bluetooth,” in *Proceedings of the IEEE International Conference on Information Technology: Computer and Communication, April* 2003, pp. 297–303.

56 J. Hallberg, M. Nilsson, and K. Synnes, “Positioning with Bluetooth,” in *Proceedings of the IEEE 10th International Conference on Telecommunication*, Vol. **2**, March 2003, pp. 954–958.

57 K. Pahlavan, X. Li, and J. Makela, “Indoor geolocation science and technology,” *IEEE Communication Magazine*, Vol. **40**, No. 2, pp. 112–118, February 2002.

58 T. Chrysikos, G. Georgopoulos, and S. Kotsopoulos, “Site-specific validation of ITU indoor path loss model at 2.4 GHz,” *10th IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM), June* 2009, pp. 1–6.

59 B.B. Parodi, H. Lenz, A. Szabo, H. Wang, J. Horn, J. Bamberger, and J. Obradovic, “Initialization and online-learning of RSS maps for indoor/campus localization,” *Proceedings of Position Location and Navigation Symposium (PLANS 06), IEEE/ION*, Myrtle Beach, South Carolina, 2006, pp. 24–27.

60 https://combain.com/

61 https://unwiredlabs.com/

62 https://location.services.mozilla.com/

63 https://www.navizon.com/wifi-cell-tower-location-database.php

64 D. Han, S. Jung, M. Lee, and G. Yoon, “Building a practical Wi-Fi-based indoor navigation system,” *IEEE Pervasive Computing*, Vol. **13**, No. 2, pp. 72–79. 2014.

65 P. Bahl and V.N. Padmanabhan, “RADAR: An in-building RF-based user location and tracking system,” in *Proceedings of the IEEE INFOCOM 2000*, Vol. **2**, March 2000, pp. 775–784.

66 V. Otsason, A. Varshavsky, A. LaMarca, and E. de Lara, “Accurate GSM indoor localization,” *UbiComp 2005, Lecture Notes Computer Science, Springer-Verlag*, Vol. **3660**, 2005, pp. 141–158.

67 Gimbal: http://www.gimbal.com/

68 iBeacon: https://developer.apple.com/ibeacon/

69 A. LaMarca, Y. Chawathe, S. Consolvo, J. Hightower, I. Smith, J. Scott, T. Sohn, J. Howard, J. Hughes, F. Potter, J. Tabert, P. Powledge, G. Borriello, and B. Schilit, “Place lab: Device positioning using radio beacons in the wild,” in *Proceedings of the 3rd International Conference on Pervasive Computing*, 2005, pp. 116–133.

70 M. Youssef and A. Agrawala, “The Horus WLAN location determination system,” in *Proceedings of the Annual International Conference on Mobile Systems, Applications and Services*, 2005, pp. 205–218.

71 T. Roos, P. Myllymaki, H. Tirri, P. Misikangas, and J. Sievanan, "A probabilistic approach to WLAN user location estimation," *International Journal of Wireless Information Networks*, Vol. **9**, No. 3, pp. 155–164, July 2002.

72 P. Castro, P. Chiu, T. Kremenek, and R.R. Muntz, "A probabilistic room location service for wireless networked environments," in *Proceedings of the 3rd International Conference on Ubiquitous Computing*, Atlanta, GA, September 2001, pp. 18–34.

73 S. Saha, K. Chaudhuri, D. Sanghi, and P. Bhagwat, "Location determination of a mobile device using IEEE 802.11b access point signals," in *Proceedings of the IEEE Wireless Communications and Networking Conference*, Vol. **3**, March 2003, pp. 1987–1992.

74 A.M. Ladd, K.E. Bekris, A. Rudys, L.E. Kavraki, and D.S. Wallach, "On the feasibility of using wireless Ethernet for indoor localization," *IEEE Transactions on Robotics and Automation*, Vol. **20**, No. 3, pp. 555–559, June 2004.

75 W. Sun, J. Liu, C. Wu, Z. Yang, X. Zhang, and Y. Liu, "MoLoc: On distinguishing fingerprint twins," in *Proceedings of IEEE International Conference on Distributed Computing Systems (ICDCS)*, 2013a.

76 V. Moghtadaiee, A.G. Dempster, and S. Lim, "Indoor localization using FM radio signals: A fingerprinting approach," in *IPIN, September* 2011, pp. 1–7.

77 Y. Chen, D. Lymberopoulos, J. Liu, and B. Priyantha, "Indoor localization using FM signals," *IEEE Transactions on Mobile Computing*, Vol.**12**, No. 8, 1502–1517, 2013.

78 A.S.-I. Noh, W.J. Lee, and J.Y. Ye, "Comparison of the mechanisms of the Zigbee's indoor localization algorithm," in *Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, 2008, SNPD '08. Ninth ACIS International Conference on*, 2008, pp. 13–18.

79 J. Yang and R. Rao, "Multi-subnetwork switching mechanism in the large-scale Zigbee mesh network for the real-time indoor positioning system," in *Multimedia Information Networking and Security (MINES), 2011 Third International Conference on*, 2011, pp. 100–104.

80 A. Anshul Rai and K.K. Chintalapudi, P. Venkat, and R. Sen, "Zee: Zero-effort crowdsourcing for indoor localization," in *Proceedings of the 18th Annual International Conference on Mobile Computing and Networking (MobiCom)*, August 2012.

81 R. Smith, M. Self, and P. Cheeseman, "Estimating uncertain spatial relationships in robotics," in *Autonomous Robot Vehicles* (eds. I.J. Cox and G.T. Wilfong), pp. 167–193, Springer-Verlag New York, Inc., 1990.

82 S. Sen et al., "Avoiding multipath to revive inbuilding Wi-Fi localization," in *MobiSys*, 2013.

83 A.R. Golding, and N. Lesh, "Indoor navigation using a diverse set of cheap, wearable sensors," *ISWC '99: Proc. 3rd IEEE Int. Symp. Wearable Computers*, San Francisco, CA, USA, IEEE Computer Society, 1999, p. 29.

84 A. Hub, J. Diepstraten, and T. Ertl, "Design and development of an indoor navigation and object identification system for the blind," *Proceedings of the 6th International ACM SIGACCESS Conference on Computers and Accessibility*, Atlanta, GA, USA, 2004, pp. 147–152.

85 L. Ran, S. Helal, and S. Moore, "Drishti: An integrated indoor/outdoor blind navigation system and service," *PERCOM '04 Proc. Second IEEE International Conf. Pervasive Computing and Communications (PerCom'04)*, Orlando, FL, USA, 2004, pp. 23–30.

86 G. Retscher and M. Thienelt, "Navio—a navigation and guidance service for pedestrians," *Journal of Global Positioning Systems*, Vol. **3**, No. 1–2, 208–217, 2004.

87 A.R. Golding and N. Lesh, "Indoor navigation using a diverse set of cheap, wearable sensors," in *Wearable Computers, The Third International Symposium on*, IEEE, pp. 29–36, 1999.

88 N. Ravi and L. Iftode, "Fiatlux: Fingerprinting rooms using light intensity." In Pervasive Computing, in *Adjunct Proceedings of the Fifth International Conference on*, 2007.

89 A.R. Jimenez, F. Zampella, and F. Seco, "Light-matching: A new signal of opportunity for pedestrian indoor navigation," in *Indoor Positioning and Indoor Navigation (IPIN), 2013 International Conference on*, IEEE, 2013, pp. 1–10.

90 S.-Y. Jung, C.-K. Choi, S. Hu Heo, S. Ro Lee, and C.-S. Park, "Received signal strength ratio based optical wireless indoor localization using light emitting diodes for illumination," in *Consumer Electronics (ICCE)*, 2013 IEEE, 2013.

91 W. Zhang and M. Kavehrad, "A 2D indoor localization system based on visible light LED," in *Photonics Society Summer Topical Meeting Series*, IEEE, 2012, pp. 80–81.

92 M. Bilgi, M. Yuksel, and N. Pala, "3D optical wireless localization," in *GLOBECOM Workshops (GC Wkshps)*, IEEE, 2010, pp. 1062–1066.

93 P. Hu, L. Li, C. Peng, G. Shen, and F. Zhao, "Pharos: Enable physical analytics through visible light based indoor localization," in *Proceedings of the Twelfth ACM Workshop on Hot Topics in Networks, ACM*, 2013, **5**.

94 L. Li, P. Hu, C. Peng, G. Shen, and F. Zhao, Epsilon: A visible light based positioning system, in *11th USENIX Symposium on Networked Systems Design and Implementation (NSDI 14)*, USENIX Association, 2014, pp. 331–343.

95 Microsoft Kinect: http://www.xbox.com/de-DE/kinect/

96 Sky-Trax Inc., 2011, http://www.sky-trax.com/

97 A. Mulloni, D. Wgner, D. Schmalstieg, and I. Barakonyi, "Indoor positioning and navigation with camera phones," *Pervasive Computing, IEEE*, Vol. **8**, pp. 22–31, 2009.

98 AICON 3D Systems: http://www.aicon.de

99 Hagisonic, "User's Guide Localization System StarGazerTM for Intelligent Robots," 2008, http://www.hagisonic.com/

100 S. Lee and J.B. Song, "Mobile Robot localization using infrared light reflecting landmarks," *Proceedings of the International Conference on Control, Automation and Systems (ICCAS'07)*, 2007, pp. 674–677.

101 M. Köhler, S. Patel, J. Summet, E. Stuntebeck, and G. Abowed, "TrackSense: Infrastructure free precise indoor positioning using projected patterns," *Pervasive Computing, LNCS*, Vol. **4480**, pp. 334–350, 2007.

102 R. Want, A. Hopper, V. Falcao, and J. Gibbons, "The Active Badge Location System," *ACM Transactions on Information Systems*, Vol. **10**, No. 1, January 1992, pp. 91–102.

103 L.M. Ni, Y. Liu, Y.C. Lau, and A.P. Patil, "LANDMARC: Indoor location sensing using active RFID," *Wireless Networks*, Vol. **10**, No. 6, pp. 701–710, November 2004.

104 J.J. Caffery and G.L. Stuber, "Overview of radiolocation in CDMA cellular system," *IEEE Communications Magazine*, Vol. **36**, No. 4, pp. 38–45, April 1998.

105 R. Harle, "A survey of indoor inertial positioning systems for pedestrians," in *IEEE Communications Surveys & Tutorials*, Vol. **15**, No. 3, pp. 1281–1293, Third Quarter 2013.

106 H. WeinBerg, "AN-602: Using the ADXL202 in pedometer and personal navigation applications," *Analog Devices, Tech. Rep.*, 2002.

107 C. Randell, C. Djiallis, and H. Muller, "Personal position measurement using dead reckoning," in *Proceedings of IEEE International Symposium on Wearable Computers (ISWC)*, 2003.

108 C. Wu, Z. Yang, Y. Liu, and W. Xi, "WILL: Wireless indoor localization without site survey," *IEEE Transactions on Parallel and Distributed Systems*, Vol. **24**, No. 4, 839–848, 2013a.

109 A.R. Jimenez, F. Seco, C. Prieto, and J. Guevara, "A comparison of pedestrian dead-reckoning algorithms using a low-cost MEMS IMU," in *Proceedings of IEEE International Symposium on Intelligent Signal Processing (WISP)*, 2009.

110 Z. Yang, C. Wu, and Y. Liu, "Locating in fingerprint space: Wireless indoor localization with little human intervention," in *Proceedings of ACM International Conference on Mobile Computing and Networking (MobiCom)*, 2012.

111 H. Wang, S. Sen, A. Elgohary, M. Farid, M. Youssef, and R. Roy Choudhury, "No need to war-drive: Unsupervised indoor localization," in *Proceedings of ACM International Conference on Mobile Systems, Applications, and Services (MobiSys)*, 2012.

112 S. Sen, J. Lee, K.-H. Kim, and P. Congdon, "Back to the basics: Avoiding multipath to revive inbuilding WiFi localization," in *Proceedings of ACM International Conference on Mobile Systems, Applications, and Services (MobiSys)*, 2013.

113 P. Goyal, V.J. Ribeiro, H. Saran, and A. Kumar, "Strap-down Pedestrian Dead-Reckoning system," in *2011 International Conference on Indoor Positioning and Indoor Navigation*, IEEE, September 2011, pp. 1–7.

114 X. Zhu, Q. Li, and G. Chen, "APT: Accurate outdoor pedestrian tracking with smartphones," in *Proceedings of IEEE International Conference on Computer Communications (INFOCOM)*, 2013.

115 J. Saarinen, J. Suomela, S. Heikkila, M. Elomaa, and A. Halme, "Personal navigation system," *2004 IEEERSJ International Conference on Intelligent Robots and Systems*, IROS IEEE Cat No04CH37566, 2004, pp. 212–217.

116 Q. Wang, X. Zhang, X. Chen, R. Chen, W. Chen, and Y. Chen, "A novel pedestrian dead reckoning algorithm using wearable EMG sensors to measure walking strides," in *2010 Ubiquitous Positioning Indoor Navigation and Location Based Service*, IEEE, October 2010, pp. 1–8.

117 R. Margaria and R. Margaria, *Biomechanics and Energetics of Muscular Exercise*, Clarendon Press, Oxford, 1976.

118 Q. Ladetto, "On foot navigation: Continuous step calibration using both complementary recursive prediction and adaptive Kalman filtering," in *Proceedings of ION GPS*, 2000.

119 D. Gusenbauer, C. Isert, and J. Krosche, "Self-contained indoor positioning on off-the-shelf mobile devices," in *Proceedings of International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2010.

120 J.E.A. Bertram and A. Ruina, "Multiple walking speed–frequency relations are predicted by constrained optimization," *Elsevier Journal of Theoretical Biology*, Vol. **209**, No. 4, 445–453. 2001.

121 D.-K. Cho, M. Mun, U. Lee, W.J. Kaiser, and M. Gerla, "AutoGait: A mobile platform that accurately estimates the distance walked," in *Proceedings of IEEE International Conference on Pervasive Computing and Communications (PerCom)*, 2010.

122 G. Shen, Z. Chen, P. Zhang, T. Moscibroda, and Y. Zhang, "Walkie-Markie: Indoor pathway mapping made easy," in *Proceedings of USENIX Conference on Networked Systems Design and Implementation (NSDI)*, 2013.

123 E. Miluzzo, N.D. Lane, K. Fodor, R. Peterson, H. Lu, M. Musolesi, S.B. Eisenman, X. Zheng, and A.T. Campbell, "Sensing meets mobile social networks: The design, implementation and evaluation of the CenceMe application," in *Proceedings of ACM Conference on Embedded Networked Sensor Systems (SenSys)*, 2008.

124 T. Iso and K. Yamazaki, "Gait analyzer based on a cell phone with a single three-axis accelerometer," in *Proceedings of ACM Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI)*, 2006.

125 J.-G. Park, A. Patel, D. Curtis, S. Teller, and J. Ledlie, "Online pose classification and walking speed estimation using handheld devices," In *Proceedings of ACM International Conference on Ubiquitous Computing (UbiComp)*, 2012.

126 R. Harle, "A survey of indoor inertial positioning systems for pedestrians," *IEEE Communications Surveys & Tutorials*, Vol. **15**, No. 3, 1281–1293, 2013.

127 K. Park, H. Shin, and H. Cha, "Smartphone-based pedestrian tracking in indoor corridor environments," *Springer Personal Ubiquitous Computing*, Vol. **17**, No. 2, pp. 359–370, 2013.

128 M.H. Afzal, V. Renaudin, and G. Lachapelle, "Assessment of indoor magnetic field anomalies using multiple magnetometers," in *ION GNSS*, September 2010, pp. 21–24.

129 M. Youssef, M.A. Yosef, and M. El-Derini, "GAC: Energy-efficient hybrid GPS-accelerometer-compass GSM localization," in *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM)*, 2010.

130 J.W. Kim and C. Park, "A step, stride and heading determination for the pedestrian navigation system," *Technology*, Vol. **3**, No. 1, pp. 273–279, 2005.

131 U. Steinhoff and B. Schiele, "Dead reckoning from the pocket – an experimental study," in *Pervasive Computing and Communications (PerCom), 2010 IEEE International Conference on*, 29 March 2010–2 April 2010, pp. 162–170.

132 X. Zhu, Q. Li, and G. Chen, "APT: Accurate outdoor pedestrian tracking with smartphones," in *Proceedings of IEEE International Conference on Computer Communications (INFOCOM)*, 2013.

133 K.-C. Lan and W.-Y. Shih, "Using floor plan to calibrate sensor drift error for indoor localization," in *Proceedings of the Joint ERCIM eMobility and MobiSense Workshop (ERCIM)*, 2013.

134 D. Fox, S. Thrun, W. Burgard, and F. Dellaert, "Particle filters for mobile robot localization," in *Sequential Monte Carlo Methods in Practice*, pp. 401–428, Springer, 2001.

135 O. Woodman and R. Harle, "Pedestrian localisation for indoor environments," in *Proceedings of ACM International Conference on Ubiquitous Computing (UbiComp)*, 2008.

136 S. Beauregard, Widyawan, and M. Klepal, "Indoor PDR performance enhancement using minimal map information and particle filters," in *Proceedings of the 2012 IEEE/ION Position, Location and Navigation Symposium*, 2008, pp. 141–147.

137 Widyawan, M. Klepal, and D. Pesch, "A Bayesian Approach for RF-Based Indoor Localisation," in Proceedings of the 4th IEEE International Symposium on Wireless Communication Systems 2007 (ISWCS 2007), Trondheim, Norway, October 16 2007.

138 Z. Yang, C. Wu, Z. Zhou, X. Zhang, X. Wang, and Y. Liu, "Mobility increases localizability: A survey on wireless indoor localization using inertial sensors," *ACM Computing Surveys (CSUR)*, Vol. **47**, No. 3, 2015: 54.

139 S. Yoon, K. Lee, and I. Rhee, "FM-based indoor localization via automatic fingerprint DB construction and matching," in *Proceedings of ACM International Conference on Mobile Systems, Applications, and Services (MobiSys)*, 2013.

140 P. Robertson, M. Angermann, B. Krach, and M. Khider, "Slam dance: Inertial-based joint mapping and positioning for pedestrian navigation," in *Proceedings of Inside GNSS*, pp. 48–59, 2010.

141 M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit, "FastSLAM: A factored solution to the simultaneous localization and mapping problem," in *Proceedings of the AAAI National Conference on Artificial Intelligence*, Edmonton, Canada, 2002.

142 J. Seitz, T. Vaupel, J. Jahn, S. Meyer, J. G. Boronat, and J. Thielecke, "A hidden Markov model for urban navigation based on fingerprinting and pedestrian dead reckoning," in *Proceedings of the 13th International Conference on Information Fusion*, pp. 1–8, 2010.

143 L. Rabiner and B. Juang, "An introduction to hidden Markov models," *ASSP Magazine, IEEE*, Vol. **3**, pp. 4–16, January 1986.

144 H. Wang, A. Szabo, and J. Bamberger, "Performance comparison of nonlinear filters for indoor WLAN positioning," *The 11th International Conference on Information Fusion*, Vol. **1**, 2008.

145 O. Woodman and R. Harle, "Pedestrian localisation for indoor environments," *Proceedings of the 10th International Conference on Ubiquitous Computing*, ACM New York, 2008, pp. 114–123.

146 H. Wen, Z. Xiao, N. Trigoni, and P. Blunsom, "On assessing the accuracy of positioning systems in indoor environments," in *Proceedings of the 10th European Conference on Wireless Sensor Networks*, pp. 1–17, 2013.

147 L.R. Welch, "Hidden Markov models and the Baum-Welch algorithm," *IEEE Information Theory Society Newsletter*, Vol. **53**, No. 4, 2003.

148 S. Pasricha, V. Ugave, C.W. Anderson, and Q. Han, "LearnLoc: A framework for smart indoor localization with embedded mobile devices," in *Proceedings of the 10th International Conference on Hardware/Software Codesign and System Synthesis*, IEEE Press, pp. 37–44, October 2015.

149 A. Martin, C. Ionut, and C. Romit, "SurroundSense: Mobile phone localization via ambience fingerprinting," MobiCom, 2009.

150 L. Bitsch, G. Jo´, S. Paul, and W. Klaus, "FootPath: Accurate map-based indoor navigation using smartphones," IPIN. September 2011.

151 A. Mittal, S. Tiku, and S. Pasricha, "Adapting convolutional neural networks for indoor localization with smart mobile devices," *ACM Great Lakes Symposium on VLSI (GLSVLSI)*, May 2018.

152 R. Faragher, C. Sarno, and M. Newman, "Opportunistic radio slam for indoor navigation using smartphone sensors," in *Position Location and Navigation Symposium (PLANS)*, **2012** IEEE/ION, April 2012, pp. 120–128.

153 B. Ferris, D. Fox, and N. Lawrence, "Wifi-SLAM using Gaussian process latent variable models," in *Proceedings of the 20th International Joint Conference on Artificial Intelligence, Ser. IJCAI'07*, San Francisco, California: Morgan Kaufmann Publishers Inc., 2007, pp. 2480–2485.

154 P. Mirowski, T.K. Ho, S. Yi, and M. MacDonald, "SignalSLAM: Simultaneous localization and mapping with mixed WiFi, Bluetooth, LTE and magnetic signals," in *Indoor Positioning and Indoor Navigation (IPIN), 2013 International Conference on*, IEEE, October 2013, pp. 1–10.

155 G. Grisetti, R. Kummerle, C. Stachniss, and W. Burgard, "A tutorial on graph-based SLAM," *IEEE Intelligent Transportation Systems Magazine*, Vol. **2**, No. 4, 2010.

156 P. Lazik, N. Rajagopal, O. Shih, B. Sinopoli, and A. Rowe, "Alps: A Bluetooth and ultrasound platform for mapping and localization," In *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems*, ACM, November 2015, pp. 73–84.

157 C. Brignone, T. Connors, G. Lyon, and S. Pradhan, "SmartLOCUS: An autonomous, self-assembling sensor network for indoor asset and systems management," Mobile Media Syst. Lab., HP Laboratories, Palo Alto, California, Tech. *Rep*, 41, 2003.

158 A. Smith, H. Balakrishnan, M. Goraczko, and N. Priyantha, "Tracking moving devices with the cricket location system," *Proceedings of the 2nd USENIX/ACM MOBISYS Conference*, Boston, MA, June 2004.

159 Radianse, Inc. Radianse Indoor Positioning. http://www.radianse.com

160 Versus Technology, http://www.versustech.com

161 EIRIS System, http://www.elcomel.com.ar/english/eiris.htm

162 S. Sen, D. Kim, S. Laroche, K.H. Kim, and J. Lee, "Bringing CUPID indoor positioning system to practice," in *Proceedings of the 24th International Conference on World Wide Web*, ACM, May 2015, pp. 938–948.

163 Q. Xu, R. Zheng, and S. Hranilovic, "Idyll: Indoor localization using inertial and light sensors on smartphones," in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* ACM2015, September 2015, pp. 307–318.

164 T. Liu, M. Carlberg, G. Chen, J. Chen, J. Kua, and A. Zakhor, "Indoor localization and visualization using a human-operated backpack system," *Proceedings of the 2010 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, September 15–17, 2010 Campus Science City, ETH Zurich, Switzerland, 2010, pp. 890–899.

165 D. Lymberopoulos, J. Liu, X. Yang, R.R. Choudhury, S. Sen, and V. Handziski, "Microsoft indoor localization competition: Experiences and lessons learned," *GetMobile: Mobile Computing and Communications*, **18** (4), pp. 24–31, 2015.

166 M. Romanovas, V. Goridko, A. Al-Jawad, M. Schwaab, L. Klingbeil, M. Traechtler, and Y. Manoli, "A study on indoor pedestrian localization algorithms with foot-mounted sensors," *International Conference on Indoor Positioning and Indoor Navigation*, 2012.

167 B. Donohoo, C. Ohlsen, and S. Pasricha, "A middleware framework for application-aware and user-specific energy optimization in smart mobile devices," *Journal of Pervasive and Mobile Computing*, Vol. **20**, pp. 47–63, July 2015.

168 B. Donohoo, C. Ohlsen, S. Pasricha, C. Anderson, and Y. Xiang, "Context-aware energy enhancements for smart mobile devices," *IEEE Transactions on Mobile Computing (TMC)*, Vol. **13**, No. 8, August 2014, pp. 1720–1732.

169 B. Donohoo, C. Ohlsen, S. Pasricha, and C. Anderson, "Exploiting spatiotemporal and device contexts for energy-efficient mobile embedded systems," *IEEE/ACM Design Automation Conference (DAC 2012)*, July 2012.

170 B. Donohoo, C. Ohlsen, and S. Pasricha, "AURA: An application and user interaction aware middleware framework for energy optimization in mobile devices," *IEEE International Conference on Computer Design (ICCD 2011)*, October 2011.

171 S. Tiku and S. Pasricha, "PortLoc: A portable data-driven indoor localization framework for smartphones," *IEEE Design and Test*, 2019.

172 S. Tiku, S. Pasricha, B. Notaros, and Q. Han, "SHERPA: A lightweight smartphone heterogeneity resilient portable indoor localization framework," *IEEE International Conference on Embedded Software and Systems (ICESS)*, Las Vegas, Nevada, June 2019.

173 L. Wirola, T. Laine, and J. Syrjärinne, "Mass market considerations for indoor positioning and navigation," *Proceedings of the 2010 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, September 15–17, 2010 Campus Science City, ETH Zurich, Switzerland, 2010.

174 O. Schneider, "Requirements for positioning and navigation in underground constructions," *Proceedings of the 2010 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, September 15–17, 2010 Campus Science City, ETH Zurich, Switzerland, 2010.

175 J. Rantakokko, P. Händel, M. Fredholm, and F. Marsten-Eklöf, "User requirements for localization and tracking technology: A survey of mission-specific needs and constraints," *Proceedings of the 2010 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, September 15–17, 2010 Campus Science City, ETH Zurich, Switzerland, 2010.

38

Navigation with Cellular Signals of Opportunity

Zaher (Zak) M. Kassas

University of California Irvine, United States

38.1 Introduction

Among the different types of signals of opportunity, cellular signals are particularly attractive for positioning, navigation, and timing (PNT) due to their inherently attractive characteristics:

Abundance
: Cellular base transceiver stations (BTSs) are plentiful due to the ubiquity of cellular and smartphones and tablets. The number of BTSs is bound to increase dramatically with the introduction of small cells to support fifth-generation (5G) wireless systems.

Geometric diversity
: The cell configuration by construction yields favorable BTS geometry, unlike certain terrestrial transmitters, which tend to be colocated (e.g. digital television). Such geometric diversity yields low geometric dilution of precision (GDOP) factors, which results in a precise PNT solution.

High carrier frequency
: The current cellular carrier frequency ranges between 800 MHz and 1900 MHz, which yields precise carrier phase navigation observables. Future 5G networks will tap into frequencies between 30 and 300 GHz.

Large bandwidth
: Cellular signals have a large bandwidth, which yields accurate time-of-arrival (TOA) estimation (e.g. the bandwidth of certain cellular long-term evolution (LTE) reference signals is up to 20 MHz).

High transmitted power
: Cellular signals are often available and usable in environments where global navigation satellite system (GNSS) signals are challenged (e.g. indoors and in deep urban canyons). The received carrier-to-noise ratio, C/N_0, from nearby cellular BTSs is more than 20 dB-Hz than that received from GPS space vehicles (SVs).

Free to use
: There is no deployment cost associated with using cellular signals for PNT – the signals are practically free to use. Specifically, the user equipment (UE) could "eavesdrop" on the transmitted cellular signals without communicating with the BTS, extract necessary PNT information from received signals, and calculate the navigation solution locally. While other navigation approaches requiring two-way communication between the UE and BTS (i.e. network-based) exist, this chapter focuses on explaining how UE-based navigation could be achieved.

Regardless of whether GNSS signals are available or not, cellular signals of opportunity could be used to produce or improve the navigation solution. In the absence of GNSS signals, cellular signals could be used to produce a navigation solution in a stand-alone fashion or to aid the inertial navigation system (INS) [1–6]. When GNSS signals are available, cellular signals could be fused with GNSS signals, yielding a navigation solution that is superior to a stand-alone GNSS solution, particularly in the vertical direction [7, 8].

Cellular signals are not intended for PNT. Therefore, to use these signals for such purpose, several challenges must be addressed. This has been the subject of extensive research over the past few years. These challenges and potential remedies are summarized next.

- Cellular signals are modulated and subsequently transmitted for non-PNT purposes. These signals are much more complicated than GNSS signals, and extracting relevant PNT information from them is not straightforward. Recent research has focused on deriving appropriate

Position, Navigation, and Timing Technologies in the 21st Century: Integrated Satellite Navigation, Sensor Systems, and Civil Applications, Volume 2, First Edition. Edited by Y. T. Jade Morton, Frank van Diggelen, James J. Spilker Jr., and Bradford W. Parkinson.

low-level models to optimally extract states and parameters of interest for PNT from received cellular signals. The effect of different propagation channels on such signals is an ongoing area of research [9–15].
- GNSS receivers are commercially available, and there is a rich body of literature on GNSS receiver design. This is not the case for cellular navigation receivers. The recent literature has published specialized receiver designs for producing navigation observables from received cellular signals (e.g. code phase, carrier phase, and Doppler frequency) [16–19].
- GNSS SVs are equipped with atomic oscillators and are tightly synchronized. However, cellular towers are equipped with less stable oscillators, typically oven-controlled crystal oscillators (OCXOs), and are less tightly synchronized. This is because communication synchronization requirements are less stringent than PNT synchronization requirements. Timing errors arising due to this somewhat loose synchronization could introduce tens of meters of localization error. Researchers have been modeling such errors and synthesizing PNT estimators that compensate for them [20–25].
- GNSS SVs transmit all necessary states and parameters to the receiver in the navigation message (e.g. SV position, clock bias, ionospheric model parameters, etc.). In contrast, cellular BTSs do not transmit such information. Therefore, navigation frameworks must be developed to estimate the states and parameters of cellular BTSs (position, clock bias, clock drift, frequency stability, etc.), which are not necessarily known a priori. Several navigation frameworks have been proposed. One such framework is to have a dedicated station that acts as a mapper, which knows its states (from GNSS signals, for instance), is estimating the unknown states of cellular BTSs, and is sharing such estimates with navigating receivers. Another framework is to simultaneously estimate the states of the receiver and cellular BTSs in a radio simultaneous localization and mapping (radio SLAM) manner [26–29].

This chapter discusses how cellular signals could be used for PNT by presenting relevant signal models, receiver architectures, PNT sources of error and corresponding models, navigation frameworks, and experimental results. The remainder of this chapter is organized as follows. Section 38.2 gives a brief overview of the evolution of cellular systems. Section 38.3 discusses modeling the clock error dynamics to facilitate estimating the unknown BTSs' clock error states. Section 38.4 describes two frameworks for navigation in cellular environments. Sections 38.5 and 38.6 discuss how to navigate with cellular code-division multiple access (CDMA) and LTE signals, respectively. Section 38.7 discusses a timing error that arises in cellular networks: clock bias discrepancy between different sectors of a BTS cell. Section 38.8 highlights the achieved navigation solution improvement upon fusing cellular signals with GNSS signals. Section 38.9 describes how cellular signals could be used to aid an INS.

Throughout this chapter, italic small bold letters (e.g. $\boldsymbol{x}$) represent vectors in the time domain, italic capital bold letters (e.g. $\boldsymbol{X}$) represent vectors in the frequency domain, and capital bold letters represent matrices (e.g. $\mathbf{X}$).

38.2 Overview of Cellular Systems

Cellular systems have evolved significantly since the first handheld cell phone was demonstrated by John F. Mitchell and Martin Cooper of Motorola in 1973. The first commercially automated cellular network was launched in Japan by Nippon Telegraph and Telephone (NTT) in 1979. This first generation (1G) was analog and used frequency division multiple access (FDMA). The second generation (2G) transitioned to digital and mostly used time-division multiple access (TDMA), which later evolved into 2.5G: General Packet Radio Service (GPRS) and 2.75G: Enhanced Data Rates for GSM Evolution (EDGE). The third generation (3G) upgraded 2G networks for faster Internet speed and used CDMA. The fourth generation (4G), commonly referred to as LTE, was introduced to allow for even faster data rates. LTE used orthogonal frequency division multiple access (OFDMA) and featured multiple-input multiple-output (MIMO), that is, antenna arrays. Figure 38.1 summarizes the existing cellular generations and their corresponding predominant modulation schemes.

This chapter focuses on using cellular CDMA and LTE signals for PNT. Table 38.1 compares the main characteristics of (i) GPS coarse/acquisition (C/A) code, (ii) CDMA

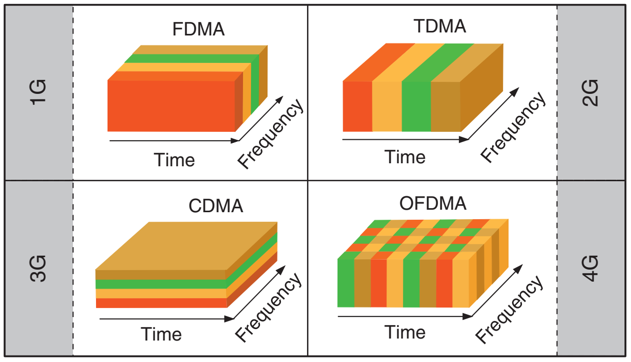

Figure 38.1 Cellular systems generations. *Source:* Adapted from A. Elnashar, "Wireless Broadband Evolution," http://www.slideshare.net/aelnashar/ayman-el-nashar, June 2011, accessed on: June 2019.

Table 38.1 GPS versus cellular CDMA and LTE comparison

Standard	Signal	Possible number of sequences	Bandwidth (MHz)	Code period (ms)	Expected ranging precision (m)*
GPS	**C/A code**	63	1.023	1	2.93
CDMA	**Pilot**	512	1.2288	26.67	2.44
LTE	**PSS**	3	0.93	10	3.22
	SSS	168	0.93	10	3.22
	CRS	504	up to 20	0.067	0.15

* 1% of chip width

pilot signal, and (ii) three LTE reference signals: primary synchronization signal (PSS), secondary synchronization signal (SSS), and cell-specific reference signal (CRS).

In 2012, the International Telecommunication Union Radiocommunication (ITU-R) sector started a program to develop an international mobile telecommunication (IMT) system for 2020 and beyond. This program set the stage for 5G research activities. The main goals of 5G compared to 4G include (i) higher density of mobile users; (ii) supporting device-to-device, ultra-reliable, and massive machine communications; (iii) lower latency; and (iv) lower battery consumption. To achieve these goals, millimeter wave bands were added to the current frequency bands for data transmission. Other salient features of 5G include millimeter waves, small cells, massive MIMO, beamforming, and full duplex [30, 31].

38.3 Clock Error Dynamics Modeling

GNSS SVs are equipped with atomic clocks, are synchronized, and their clock errors are transmitted in the navigation message along with the SVs' orbital elements. In contrast, cellular BTSs are equipped with less stable oscillators (typically OCXOs), are roughly synchronized to GNSS, and their clock error states (bias and drift) and positions are typically unknown. As such, the cellular BTSs' clock errors and positions must be estimated. Therefore, it is important to model the clock error state dynamics. To this end, a typical model for the dynamics of the clock error states is the so-called two-state model, composed of the clock bias δt and clock drift $\dot{\delta t}$, as depicted in Figure 38.2.

The clock error states evolve according to

$$\dot{\boldsymbol{x}}_{\text{clk}}(t) = \mathbf{A}_{\text{clk}}\boldsymbol{x}_{\text{clk}}(t) + \tilde{\boldsymbol{w}}_{\text{clk}}(t),$$

$$\boldsymbol{x}_{\text{clk}} = \begin{bmatrix} \delta t \\ \dot{\delta t} \end{bmatrix}, \quad \tilde{\boldsymbol{w}}_{\text{clk}} = \begin{bmatrix} \tilde{w}_{\delta t} \\ \tilde{w}_{\dot{\delta t}} \end{bmatrix}, \quad \mathbf{A}_{\text{clk}} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \tag{38.1}$$

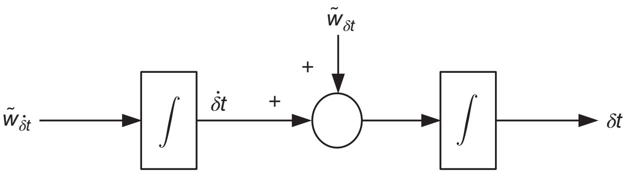

Figure 38.2 Clock error states dynamics model. *Source:* Z. Kassas, Analysis and synthesis of collaborative opportunistic navigation systems, Ph.D. Dissertation, The University of Texas at Austin, USA, May 2014. Reproduced with permission of Z. Kassas (University of Texas at Austin).

where the elements of $\tilde{\boldsymbol{w}}_{\text{clk}}$ are modeled as zero-mean, mutually independent white noise processes, and the power spectral density of $\tilde{\boldsymbol{w}}_{\text{clk}}$ is given by $\tilde{\mathbf{Q}}_{\text{clk}} = \text{diag}\left[S_{\tilde{w}_{\delta t}}, S_{\tilde{w}_{\dot{\delta t}}}\right]$. The power spectra $S_{\tilde{w}_{\delta t}}$ and $S_{\tilde{w}_{\dot{\delta t}}}$ can be related to the power-law coefficients $\{h_\alpha\}_{\alpha=-2}^{2}$, which have been shown through laboratory experiments to be adequate to characterize the power spectral density of the fractional frequency deviation $y(t)$ of an oscillator from the nominal frequency, which takes the form $S_y(f) = \sum_{\alpha=-2}^{2} h_\alpha f^\alpha$ [32, 33]. It is common to approximate the clock error dynamics by considering only the frequency random walk coefficient h_{-2} and the white frequency coefficient h_0, which lead to $S_{\tilde{w}_{\delta t}} \approx \frac{h_0}{2}$ and $S_{\tilde{w}_{\dot{\delta t}}} \approx 2\pi^2 h_{-2}$ [34, 35]. Typical OCXO values for h_0 and h_{-2} are given in Table 38.2.

Discretizing the dynamics (Eq. (38.1)) at a sampling interval T yields the discrete-time-equivalent model

$$\boldsymbol{x}_{\text{clk}}(k+1) = \mathbf{F}_{\text{clk}}\boldsymbol{x}_{\text{clk}}(k) + \boldsymbol{w}_{\text{clk}}(k),$$

where $\boldsymbol{w}_{\text{clk}}$ is a discrete-time zero-mean white noise sequence with covariance $\mathbf{Q}_{\text{clk}}$, and

$$\mathbf{F}_{\text{clk}} = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix}, \quad \mathbf{Q}_{\text{clk}} = \begin{bmatrix} S_{\tilde{w}_{\delta t}} T + S_{\tilde{w}_{\dot{\delta t}}} \frac{T^3}{3} & S_{\tilde{w}_{\dot{\delta t}}} \frac{T^2}{2} \\ S_{\tilde{w}_{\dot{\delta t}}} \frac{T^2}{2} & S_{\tilde{w}_{\dot{\delta t}}} T \end{bmatrix}. \tag{38.2}$$

Table 38.2 Typical h_0 and h_{-2} values for different OCXOs [36]

h_0	h_{-2}
2.6×10^{-22}	4.0×10^{-26}
8.0×10^{-20}	4.0×10^{-23}
3.4×10^{-22}	1.3×10^{-24}

Source: J. Curran, G. Lachapelle, and C. Murphy, "Digital GNSS PLL design conditioned on thermal and oscillator phase noise," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 1, pp. 180–196, January 2012.

38.4 Navigation Frameworks in Cellular Environments

BTS positions can be readily obtained via several methods, for example, (i) from cellular BTS databases (if available) or (ii) by deploying multiple mapping receivers with knowledge of their own states, estimating the position states of the BTSs for a sufficiently long period of time [37–39]. These estimates are physically verifiable via surveying or satellite images. Unlike BTS positions, which are static, the clock error states are stochastic and dynamic, as discussed in Section 38.3, and are difficult to verify.

Estimating the BTSs' states can be achieved via two frameworks:

Mapper/Navigator
: This framework comprises (i) receiver(s) with knowledge of their own states, referred to as mapper(s), making measurements on ambient BTSs (e.g. pseudorange and carrier phase). The mappers' role is to estimate the cellular BTSs' states. (ii) A receiver with no knowledge of its own states, referred to as the navigator, making measurements on the same ambient BTSs to estimate its own states, while receiving estimates of the BTSs' states from the mappers.

Radio SLAM
: In this framework, the receiver maps the BTSs simultaneously with localizing itself in the radio environment.

To make the estimation problems associated with the above frameworks observable, certain a priori knowledge about the BTSs' or receiver's states must be satisfied [27, 40–42]. For simplicity, a planar environment will be assumed, with the receiver and BTS three-dimensional (3D) position states appropriately projected onto such a planar environment. The state of the receiver is defined as $\boldsymbol{x}_r \triangleq \left[\boldsymbol{r}_r^{\mathrm{T}}, c\delta t_r\right]^{\mathrm{T}}$, where $\boldsymbol{r}_r = [x_r, y_r]^{\mathrm{T}}$ is the position vector of the receiver, δt_r is the receiver's clock bias, and c is the speed of light. Similarly, the state of the i-th BTS is defined as $\boldsymbol{x}_{s_i} \triangleq \left[\boldsymbol{r}_{s_i}^{\mathrm{T}}, c\delta t_{s_i}\right]^{\mathrm{T}}$, where $\boldsymbol{r}_{s_i} = \left[x_{s_i}, y_{s_i}\right]^{\mathrm{T}}$ is the position vector of the i-th BTS, and δt_{s_i} is its clock bias. The pseudorange measurement to the i-th BTS, ρ_i, can be expressed as

$$\rho_i = h_i(\boldsymbol{x}_r, \boldsymbol{x}_{s_i}) + v_i, \tag{38.3}$$

where $h_i(\boldsymbol{x}_r, \boldsymbol{x}_{s_i}) \triangleq \|\boldsymbol{r}_r - \boldsymbol{r}_{s_i}\|_2 + c \cdot [\delta t_r - \delta t_{s_i}]$ and v_i is the measurement noise, which is modeled as a zero-mean Gaussian random variable with variance σ_i^2 [27]. The following sections outline the calculations associated with each navigation framework assuming pseudorange measurements from cellular towers. Frameworks with carrier phase measurements are discussed in [43].

38.4.1 Mapper/Navigator Framework

Assuming that the receiver is drawing pseudoranges from $N \geq 3$ BTSs with *known* states, the receiver's state can be estimated from (Eq. (38.3)) by solving a weighted nonlinear least-squares (WNLS) problem. However, in practice, the BTSs' states are *unknown*, in which case the mapper/navigator framework can be employed [18, 25].

Consider a mapper with knowledge of its own state vector (by having access to GNSS signals, for example) to be present in the navigator's environment as depicted in Figure 38.3.

The mapper's objective is to estimate the BTSs' position and clock bias states and share these estimates with the navigator through a central database. For simplicity, assume the position states of the BTSs to be known and stored in a database. In the sequel, it is assumed that the mapper is producing an estimate $\hat{\delta t}_{s_i}$ and an associated estimation error variance $\sigma^2_{\hat{\delta t}_{s_i}}$ for each of the BTSs.

Consider M mappers and N BTSs. Denote the state vector of the j-th mapper by $\boldsymbol{x}_{r_j}$, the pseudorange measurement by

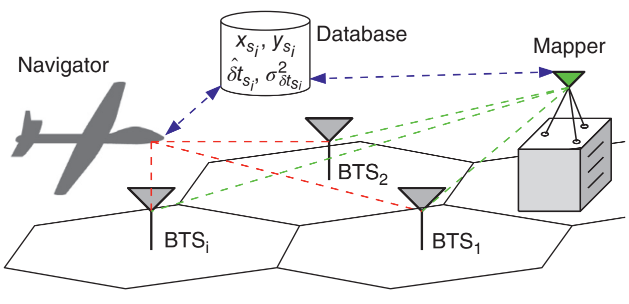

Figure 38.3 Mapper and navigator in a cellular environment (Khalife et al. [18]; Khalife and Kassas [25]). *Source:* Reproduced with permission of IEEE, ION.

the j-th mapper on the i-th BTS by $\rho_i^{(j)}$, and the corresponding measurement noise by $v_i^{(j)}$. Assume $v_i^{(j)}$ to be independent for all i and j with a corresponding variance $\sigma_i^{(j)2}$. Define the set of measurements made by all mappers on the i-th BTS as

$$\boldsymbol{z}_i \triangleq \begin{bmatrix} \|\boldsymbol{r}_{r_1} - \boldsymbol{r}_{s_i}\| + c\delta t_{r_1} - \rho_i^{(1)} \\ \vdots \\ \|\boldsymbol{r}_{r_M} - \boldsymbol{r}_{s_i}\| + c\delta t_{r_M} - \rho_i^{(M)} \end{bmatrix} = \begin{bmatrix} c\delta t_{s_i} - v_i^{(1)} \\ \vdots \\ c\delta t_{s_i} - v_i^{(M)} \end{bmatrix}$$
$$= c\delta t_{s_i}\boldsymbol{1}_M + \boldsymbol{v}_i,$$

where $\boldsymbol{1}_M \triangleq [1, ..., 1]^{\mathrm{T}}$ and $\boldsymbol{v}_i \triangleq -\left[v_i^{(1)}, ..., v_i^{(M)}\right]^{\mathrm{T}}$. The clock bias δt_{s_i} is estimated by solving a weighted least-squares (WLS) problem, resulting in the estimate

$$\hat{\delta t}_{s_i} = \frac{1}{c}\left(\boldsymbol{1}_M^{\mathrm{T}}\mathbf{W}\boldsymbol{1}_M\right)^{-1}\boldsymbol{1}_M^{\mathrm{T}}\mathbf{W}\boldsymbol{z},$$
$$\mathbf{W} = \mathrm{diag}\left[\frac{1}{\sigma_i^{(1)2}}, ..., \frac{1}{\sigma_i^{(M)2}}\right]$$

and the associated estimation error variance $\sigma^2_{\delta t_{s_i}} = \frac{1}{c^2}\left(\boldsymbol{1}_M^{\mathrm{T}}\mathbf{W}\boldsymbol{1}_M\right)^{-1}$, where $\mathbf{W}$ is the weighting matrix. The true clock bias of the i-th BTS can now be expressed as $\delta t_{s_i} = \hat{\delta t}_{s_i} + w_i$, where w_i is a zero-mean Gaussian random variable with variance $\sigma^2_{\delta t_{s_i}}$.

Since the navigating receiver is using the estimate of the BTS clock bias, which is produced by the mapping receiver, the pseudorange measurement made by the navigating receiver on the i-th BTS becomes

$$\rho_i = h_i(\boldsymbol{x}_r, \hat{\boldsymbol{x}}_{s_i}) + \eta_i,$$

where $\hat{\boldsymbol{x}}_{s_i} = \left[\boldsymbol{r}_{s_i}^{\mathrm{T}}, c\hat{\delta t}_{s_i}\right]^{\mathrm{T}}$ and $\eta_i \triangleq v_i - w_i$ models the overall uncertainty in the pseudorange measurement. Hence, the vector $\boldsymbol{\eta} \triangleq [\eta_1, ..., \eta_N]^{\mathrm{T}}$ is a zero-mean Gaussian random vector with a covariance matrix $\boldsymbol{\Sigma} = \mathbf{C} + \mathbf{R}$, where $\mathbf{C} = c^2 \cdot \mathrm{diag}\left[\sigma^2_{\delta t_{s_1}}, ..., \sigma^2_{\delta t_{s_N}}\right]$ is the covariance matrix of $\boldsymbol{w} \triangleq [w_1, ..., w_N]^{\mathrm{T}}$ and $\mathbf{R} = \mathrm{diag}\left[\sigma_1^2, ..., \sigma_N^2\right]$ is the covariance of the measurement noise vector $\boldsymbol{v} = [v_1, ..., v_N]^{\mathrm{T}}$. The Jacobian matrix $\mathbf{H}$ of the nonlinear measurements $\boldsymbol{h} \triangleq \left[h_1(\boldsymbol{x}_r, \hat{\boldsymbol{x}}_{s_1}), ..., h_N(\boldsymbol{x}_r, \hat{\boldsymbol{x}}_{s_N})\right]^{\mathrm{T}}$ with respect to $\boldsymbol{x}_r$ is given by $\mathbf{H} = [\mathbf{G} \quad \boldsymbol{1}_N]$, where

$$\mathbf{G} \triangleq \begin{bmatrix} \frac{x_r - x_{s_1}}{\|\boldsymbol{r}_r - \boldsymbol{r}_{s_1}\|} & \frac{y_r - y_{s_1}}{\|\boldsymbol{r}_r - \boldsymbol{r}_{s_1}\|} \\ \vdots & \vdots \\ \frac{x_r - x_{s_N}}{\|\boldsymbol{r}_r - \boldsymbol{r}_{s_N}\|} & \frac{y_r - y_{s_N}}{\|\boldsymbol{r}_r - \boldsymbol{r}_{s_N}\|} \end{bmatrix}.$$

The navigating receiver's state can now be estimated by solving a WNLS problem. The WNLS equations are given by

$$\hat{\boldsymbol{x}}_r^{(l+1)} = \hat{\boldsymbol{x}}_r^{(l)} + \left(\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\left(\boldsymbol{\rho} - \hat{\boldsymbol{\rho}}^{(l)}\right)$$
$$\mathbf{P}^{(l)} = \left(\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\right)^{-1},$$

where l is the iteration number, and $\hat{\boldsymbol{\rho}}^{(l)}$ denotes the nonlinear measurements $\boldsymbol{h}$ evaluated at the current estimate $\hat{\boldsymbol{x}}_r^{(l)}$.

38.4.2 Radio SLAM Framework

A dynamic estimator, such as an extended Kalman filter (EKF), can be used in the radio SLAM framework for stand-alone receiver navigation (i.e. without a mapper). Certain a priori knowledge about the BTSs' and/or receiver's states must be satisfied to make the radio SLAM estimation problem observable [27, 40–42].

To demonstrate a particular formulation of the radio SLAM framework, consider the simple case where the BTSs' positions are known. Also, assume the receiver's *initial* state vector to be known (e.g. from a GNSS navigation solution). Using the pseudoranges (Eq. (38.3)), the EKF will estimate the state vector composed of the receiver's position $\boldsymbol{r}_r$ and velocity $\dot{\boldsymbol{r}}_r$ as well as the difference between the receiver's clock bias and each BTS and the difference between the receiver's clock drift and each BTS, specifically

$$\boldsymbol{x} = \left[\boldsymbol{r}_r^{\mathrm{T}}, \dot{\boldsymbol{r}}_r^{\mathrm{T}}, \boldsymbol{x}_{\mathrm{clk}_1}^{\mathrm{T}}, ..., \boldsymbol{x}_{\mathrm{clk}_N}^{\mathrm{T}}\right]^{\mathrm{T}},$$

where $\boldsymbol{x}_{\mathrm{clk}_i} \triangleq \left[(\delta t_r - \delta t_{s_i}), (\dot{\delta t}_r - \dot{\delta t}_{s_i})\right]^{\mathrm{T}}$; δt_r and δt_{s_i} are the receiver's and the i-th BTS clock bias, respectively; and $\dot{\delta t}_r$ and $\dot{\delta t}_{s_i}$ are the receiver's and the i-th BTS clock drift, respectively.

Assuming the receiver to be moving with velocity random walk dynamics, the system's dynamics after discretization at a uniform sampling period T can be modeled as

$$\boldsymbol{x}(k+1) = \mathbf{F}\boldsymbol{x}(k) + \boldsymbol{w}(k), \tag{38.4}$$

$$\mathbf{F} = \begin{bmatrix} \mathbf{F}_{\mathrm{pv}} & \mathbf{0}_{4\times 2N} \\ \mathbf{0}_{2N\times 4} & \mathbf{F}_{\mathrm{clk}} \end{bmatrix}, \qquad \mathbf{F}_{\mathrm{clk}_i} = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix},$$
$$\mathbf{F}_{\mathrm{clk}} = \mathrm{diag}\left[\mathbf{F}_{\mathrm{clk}_1}, ..., \mathbf{F}_{\mathrm{clk}_N}\right], \qquad \mathbf{F}_{\mathrm{pv}} = \begin{bmatrix} \mathbf{I}_{2\times 2} & T\mathbf{I}_{2\times 2} \\ \mathbf{0}_{2\times 2} & \mathbf{I}_{2\times 2} \end{bmatrix},$$

where $\boldsymbol{w}(k)$ is a discrete-time zero-mean white noise sequence with covariance $\mathbf{Q} = \mathrm{diag}\left[\mathbf{Q}_{\mathrm{pv}}, \mathbf{Q}_{\mathrm{clk}}\right]$. Defining $\tilde{q}_x$ and $\tilde{q}_y$ to be the power spectral densities of the acceleration in the x– and y– directions, $\mathbf{Q}_{\mathrm{pv}}$ and $\mathbf{Q}_{\mathrm{clk}}$ are given by

$$\mathbf{Q}_{\mathrm{pv}} = \begin{bmatrix} \tilde{q}_x \frac{T^3}{3} & 0 & \tilde{q}_x \frac{T^2}{2} & 0 \\ 0 & \tilde{q}_y \frac{T^3}{3} & 0 & \tilde{q}_y \frac{T^2}{2} \\ \tilde{q}_x \frac{T^2}{2} & 0 & \tilde{q}_x T & 0 \\ 0 & \tilde{q}_y \frac{T^2}{2} & 0 & \tilde{q}_y T \end{bmatrix},$$

$$\mathbf{Q}_{\mathrm{clk}} = \begin{bmatrix} \mathbf{Q}_{\mathrm{clk}_r} + \mathbf{Q}_{\mathrm{clk}_{s_1}} & \mathbf{Q}_{\mathrm{clk}_r} & \cdots & \mathbf{Q}_{\mathrm{clk}_r} \\ \mathbf{Q}_{\mathrm{clk}_r} & \mathbf{Q}_{\mathrm{clk}_r} + \mathbf{Q}_{\mathrm{clk}_{s_2}} & \cdots & \mathbf{Q}_{\mathrm{clk}_r} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{Q}_{\mathrm{clk}_r} & \mathbf{Q}_{\mathrm{clk}_r} & \cdots & \mathbf{Q}_{\mathrm{clk}_r} + \mathbf{Q}_{\mathrm{clk}_{s_N}} \end{bmatrix},$$

where $\mathbf{Q}_{\mathrm{clk}_r}$ and $\mathbf{Q}_{\mathrm{clk}_{s_i}}$ correspond to the receiver's and the i-th BTS clock process noise covariances, respectively, specified in Eq. (38.2). Formulations of other more sophisticated radio SLAM scenarios are discussed in [27, 29, 41].

Note that in many practical situations, the receiver is coupled with an inertial measurement unit (IMU), which can be used instead of the statistical model to propagate the estimator's state between measurement updates from BTSs [44, 45]. This is discussed in more detail in Section 38.9.

38.5 Navigation with Cellular CDMA Signals

To establish and maintain a connection between cellular CDMA BTSs and the UE, each BTS broadcasts comprehensive timing and identification information. Such information could be utilized for PNT. The sequences transmitted on the forward link channel, that is, from BTS to UE, are known. Therefore, by correlating the received cellular CDMA signal with a locally generated sequence, the receiver can estimate the TOA and produce a pseudorange measurement. This technique is used in GPS. With enough pseudorange measurements and knowing the states of the BTSs, the receiver can localize itself within the cellular CDMA environment.

This section is organized as follows. Section 38.5.1 provides an overview of the modulation process of the forward link channel. Section 38.5.2 presents a receiver architecture for producing navigation observables from received cellular CDMA signals. Section 38.5.3 analyzes the precision of the cellular CDMA pseudorange observable. Section 38.5.4 shows experimental results for ground and aerial vehicles navigating with cellular CDMA signals.

38.5.1 Forward Link Signal Structure

Cellular CDMA networks employ orthogonal and maximal-length pseudorandom noise (PN) sequences in order to enable multiplexing over the same channel. In a cellular CDMA communication system, 64 logical channels are multiplexed on the forward link channel: a pilot channel, a sync channel, 7 paging channels, and 55 traffic channels [46]. The following sections discuss the modulation process of the forward link and give an overview of the pilot, sync, and paging channels from which timing and positioning information can be extracted. Models of the transmitted and received signals are also given.

38.5.1.1 Modulation of Forward Link CDMA Signals

The data transmitted on the forward link channel in cellular CDMA systems is modulated through quadrature phase shift keying (QPSK) and then spread using direct-sequence CDMA (DS-CDMA). However, for the channels of interest from which positioning and timing information is extracted, the in-phase and quadrature components, I and Q, respectively, carry the same message $m(t)$ as shown in Figure 38.4. The spreading sequences c_I and c_Q, called the short code, are maximal-length PN sequences that are generated using 15 linear feedback shift registers (LFSRs). Hence, the length of c_I and c_Q is $2^{15} - 1 = 32,767$ chips at a chipping rate of 1.2288 Mcps [47]. The characteristic polynomials of the short code I and Q components, $P_I(D)$ and $P_Q(D)$, are given by

$$P_I(D) = D^{15} + D^{13} + D^9 + D^8 + D^7 + D^5 + 1$$
$$P_Q(D) = D^{15} + D^{12} + D^{11} + D^{10} + D^6 + D^5 + D^4 + D^3 + 1,$$

where D is the delay operator. It is worth noting that an extra zero is added after the occurrence of 14 consecutive zeros to make the length of the short code a power of two.

In order to distinguish the received data from different BTSs, each station uses a shifted version of the PN codes. This shift is an integer multiple of 64 chips, and this integer multiple, which is unique for each BTS, is known as the pilot offset. The cross-correlation of the same PN sequence with different pilot offsets can be shown to be negligible [46]. Each individual logical channel is spread by a unique

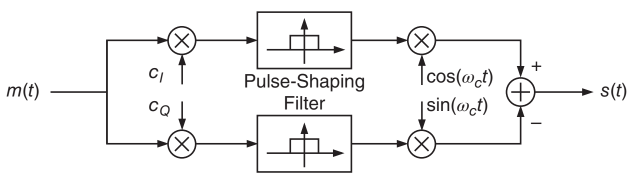

Figure 38.4 Forward link modulator (Khalife et al. [18]). *Source:* Reproduced with permission of IEEE.

64-chip Walsh code [48]. Therefore, at most 64 logical channels can be multiplexed at each BTS. Spreading by the short code enables multiple access for BTSs over the same carrier frequency, while orthogonal spreading by the Walsh codes enables multiple access for users over the same BTS. The CDMA signal is then filtered using a digital pulse shaping filter that limits the bandwidth of the transmitted CDMA signal according to the cdma2000 standard. The signal is finally modulated by the carrier frequency ω_c to produce $s(t)$.

38.5.1.2 Pilot Channel

The message transmitted by the pilot channel is a constant stream of binary zeros and is spread by Walsh code zero, which also consists of 64 binary zeros. Therefore, the modulated pilot signal is nothing but the short code, which can be utilized by the receiver to detect the presence of a CDMA signal and then track it. The fact that the pilot signal is data-less allows for longer integration time. The receiver can differentiate between the BTSs based on their pilot offsets.

38.5.1.3 Sync Channel

The sync channel is used to provide time and frame synchronization to the receiver. Cellular CDMA networks typically use GPS as the reference timing source, and the BTS sends the system time to the receiver over the sync channel [49]. Other information such as the pilot PN offset and the long code state are also provided on the sync channel [47]. The long code is a PN sequence and is used to spread the reverse link signal (i.e. UE to BTS) and the paging channel message. The long code has a chip rate of 1.2288 Mcps and is generated using 42 LFSRs. The outputs of the registers are masked and modulo-two added together to form the long code. The latter has a period of more than 41 days; hence, the states of the 42 LFSRs and the mask are transmitted to the receiver so that it can readily achieve long code synchronization. The sync message encoding before transmission is shown in Figure 38.5.

The initial message, which is at 1.2 Ksps, is convolutionally encoded at a rate $r = (1/2)$ with generator functions $g_0 = 753$ (octal) and $g_1 = 561$ (octal) [48]. The state of the encoder is not reset during the transmission of a message capsule. The resulting symbols are repeated twice, and the resulting frames, which are 128 symbols long, are block-interleaved using the bit reversal method [47]. The modulated symbols, which have a rate of 4.8 Ksps, are spread with Walsh code 32. The sync message is divided into 80 ms superframes, and each superframe is divided into three frames. The first bit of each frame is called the start of message (SOM). The beginning of the sync message is set to be on the first frame of each superframe, and the SOM of this frame is set to one. The BTS sets the other SOMs to zero. The sync channel message capsule is composed of the message length, the message body, cyclic redundancy check (CRC), and zero padding. The length of the zero padding is such that the message capsule extends up to the start of the next superframe. A 30-bit CRC is computed for each sync channel message with the generator polynomial

$$g(x) = x^{30} + x^{29} + x^{21} + x^{20} + x^{15} + x^{13} + x^{12} + x^{11} + x^{8} + x^{7} + x^{6} + x^{2} + x + 1.$$

The SOM bits are dropped by the receiver, and the frames bodies are combined to form a sync channel capsule. The sync message structure is summarized in Figure 38.6.

38.5.1.4 Paging Channel

The paging channel transmits all the necessary overhead parameters for the UE to register into the network [46]. Some mobile operators also transmit the BTS latitude and longitude on the paging channel, which can be exploited for navigation. The major cellular CDMA providers in the United States, Sprint and Verizon, do not transmit the BTS latitude and longitude. US Cellular used to transmit the BTS latitude and longitude, but this provider does not operate anymore. The Base Station ID (BID) is also transmitted in the paging channel, which is important to decode for data association purposes. The paging channel message encoding before transmission is illustrated in Figure 38.7.

The initial bit rate of the paging channel message is either 9.6 Kbps or 4.8 Kbps and is provided in the sync channel message. Next, the data is convolutionally encoded in the same way as that of the sync channel data. The output symbols are repeated twice only if the bit rate is less than 9.6

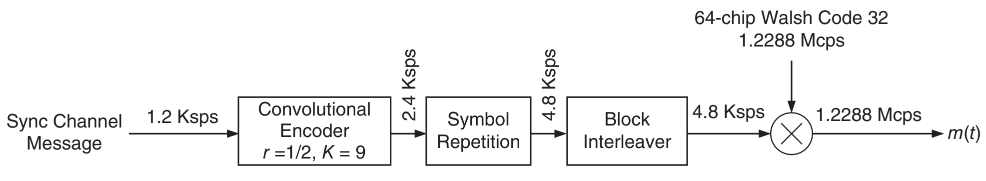

Figure 38.5 Forward link sync channel encoder (Khalife et al. [18]; 3GPP2 [50]). *Source:* Reproduced with permission of IEEE.

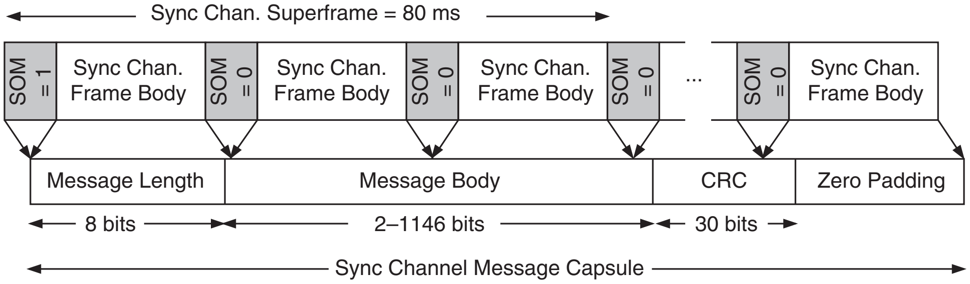

Figure 38.6 Sync channel message structure (Khalife et al. [18]; 3GPP2 [50]). *Source:* Reproduced with permission of IEEE.

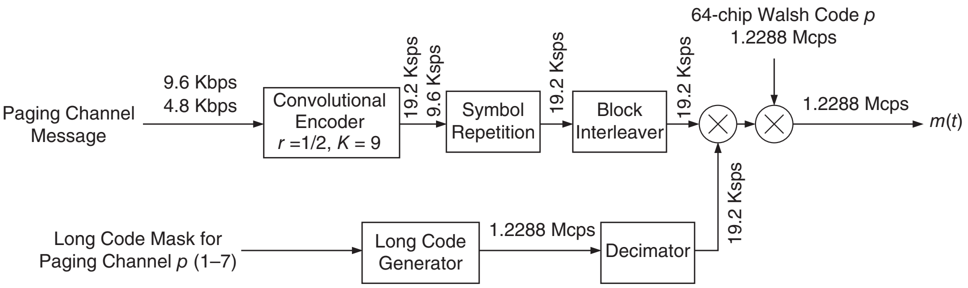

Figure 38.7 Forward link paging channel encoder (Khalife et al. [18]; 3GPP2 [50]). *Source:* Reproduced with permission of IEEE.

Kbps. After symbol repetition, the resulting frames, which are 384 symbols long, are block-interleaved one frame at a time. The interleaver is different from the one used for the sync channel because it operates on 384 symbols instead of 128 symbols. However, both interleavers use the bit reversal method. Finally, the paging channel message is scrambled by modulo-two addition with the long code sequence.

The paging channel message is divided into 80 ms time slots, where each slot is composed of eight half-frames. All the half-frames start with a synchronized capsule indicator (SCI) bit. A message capsule can be transmitted in both a synchronized and an unsynchronized manner. A synchronized message capsule starts exactly after the SCI. In this case, the BTS sets the value of the first SCI to one and the rest of the SCIs to zero. If by the end of the paging message capsule there remains fewer than 8 bits before the next SCI, the message is zero-padded to the next SCI. Otherwise, an unsynchronized message capsule is sent immediately after the end of the previous message [46]. The paging channel structure is summarized in Figure 38.8.

38.5.1.5 Transmitted Signal Model

The pilot signal, which is purely the PN sequence, is used to acquire and track a cellular CDMA signal. The acquisition and tracking will be discussed in Section 38.5.2. Demodulating the other channels becomes an open-loop problem, since no feedback is taken from the sync, paging, or any of the other channels for tracking. Since all the other channels are synchronized to the pilot, only the pilot needs to be tracked. In fact, it is required by the cdma2000 specification that all the coded channels be synchronized with the pilot to within ±50 ns [50]. Although signals from multiple BTSs could be received simultaneously, a UE can associate each individual signal with the corresponding BTS, since the offsets between the transmitted PN sequences are much larger than one chip. This is because the autocorrelation function has negligible values for delays greater than one chip. Therefore, the PN offsets, which are much larger than one chip delay, guarantee that no significant interference is introduced (the autocorrelation function is discussed in Section 38.5.2.3 and is shown in Figure 38.13).

The normalized transmitted pilot signal $s(t)$ by a particular BTS can be expressed as

$$\begin{aligned} s(t) &= \sqrt{C}\left\{c_I'[t-\Delta(t)]\cos(\omega_c t) - c_Q'[t-\Delta(t)]\sin(\omega_c t)\right\} \\ &= \Re\left\{\sqrt{C}\left[c_I'[t-\Delta(t)] + jc_Q'[t-\Delta(t)]\right]\cdot e^{j\omega_c t}\right\} \\ &= \frac{\sqrt{C}}{2}\left\{c_I'[t-\Delta(t)] + jc_Q'[t-\Delta(t)]\right\}\cdot e^{j\omega_c t} \\ &\quad + \frac{\sqrt{C}}{2}\left\{c_I'[t-\Delta(t)] - jc_Q'[t-\Delta(t)]\right\}\cdot e^{-j\omega_c t}, \end{aligned}$$

where $\Re\{\cdot\}$ denotes the real part; C is the total power of the transmitted signal; $c'_I(t) = c_I(t) * h(t)$ and $c'_Q(t) = c_Q(t) * h(t)$; h is the continuous-time impulse response of the pulse shaping filter; c_I and c_Q are the in-phase and quadrature

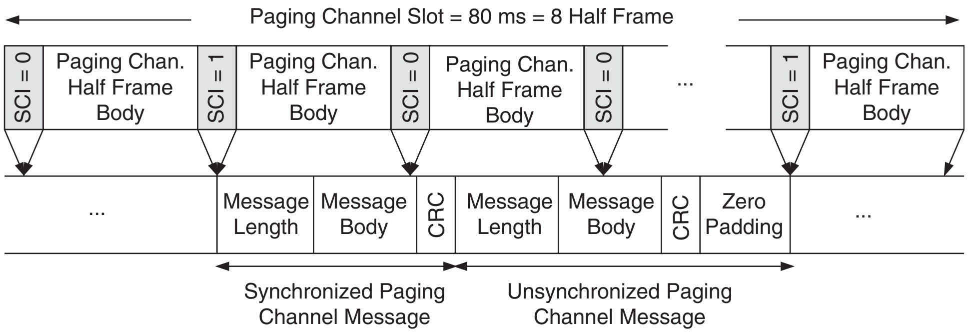

Figure 38.8 Paging channel message structure (Khalife et al. [18]; 3GPP2 [50]). *Source:* Reproduced with permission of IEEE.

PN sequences, respectively; $\omega_c = 2\pi f_c$, where f_c is the carrier frequency; and Δ is the absolute clock bias of the BTS from GPS time. The total clock bias Δ is defined as

$$\Delta(t) = 64 \cdot (PN_{\text{offset}} T_c) + \delta t_s(t),$$

where PN_{offset} is the PN offset of the BTS, $T_c = \dfrac{1 \times 10^{-6}}{1.2288}$ s is the chip interval, and δt_s is the BTS clock bias. Since the chip interval is known and the PN offset can be decoded by the receiver, only δt_s needs to be estimated.

It is worth noting that the cdma2000 standard requires the BTS's clock to be synchronized with GPS to within 10 μs, which translates to a range of approximately 3 km (the average cell size) [51]. Note that a PN offset of 1 (i.e. 64 chips) is enough to prevent significant interference from different BTSs. This translates to more than 15 km between BTSs. Subtracting 6 km from this value due to worst-case synchronization with GPS (i.e. 3 km for each BTS), BTSs at 9 km or more from the serving BTS could cause interference (assuming all BTSs suffer from the worst-case synchronizations). But 9 km is larger than the maximum distance for receiving cellular CDMA signals for ground receivers. Therefore, this synchronization requirement is enough to prevent severe interference between the short codes transmitted from different BTSs and maintains the CDMA system's capability to perform soft hand-offs [47]. The clock bias of the BTS can therefore be neglected for communication purposes. However, ignoring δt_s in navigation applications can be disastrous, and it is therefore crucial for the receiver to know the BTS's clock bias. The estimation of δt_s can be accomplished via the frameworks discussed in Section 38.4.

38.5.1.6 Received Signal Model

Assuming the transmitted signal to have propagated through an additive white Gaussian noise channel with a power spectral density of $\frac{N_0}{2}$, a model of the received discrete-time signal $r[m]$ after radio frequency (RF) front-end processing: down-mixing, a quadrature approach to bandpass sampling [52], and quantization can be expressed as

$$r[m] = \frac{\sqrt{C}}{2}\left\{c_I'[t_m - t_s(t_m)] - jc_Q'[t_m - t_s(t_m)]\right\} \cdot e^{j\theta(t_m)} + n[m], \tag{38.5}$$

where $t_s(t_m) \triangleq \delta t_{\text{TOF}} + \Delta(t_k - \delta t_{\text{TOF}})$ is the PN code phase of the BTS, $t_m = mT_s$ is the sample time expressed in receiver time, T_s is the sampling period, δt_{TOF} is the time of flight (TOF) from the BTS to the receiver, θ is the beat carrier phase of the received signal, and $n[m] = n_I[m] + jn_Q[m]$ with n_I and n_Q being independent and identically distributed Gaussian random sequences with zero mean and variance $\frac{N_0}{2T_s}$. The receiver presented in Section 38.5.2 will operate on the samples of $r[m]$ in Eq. (38.5).

38.5.2 CDMA Receiver Architecture

This section details the architecture of a cellular CDMA navigation receiver, which consists of three main stages: signal acquisition, tracking, and message decoding [18]. The receiver utilizes the pilot signal to detect the presence of a CDMA signal and then tracks it. Section 38.5.2.1 describes the correlation process in the receiver. Sections 38.5.2.2 and 38.5.2.3 discuss the acquisition and tracking stages, respectively. Section 38.5.2.4 details decoding the sync and paging channel messages.

38.5.2.1 Correlation Function

Given samples of the baseband signal exiting the RF front-end defined in Eq. (38.5), the cellular CDMA receiver first wipes off the residual carrier phase and match-filters the resulting signal. The output of the matched filter can be expressed as

$$x[m] = \left[r[m] \cdot e^{-j\hat{\theta}(t_m)}\right] * h[-m], \tag{38.6}$$

Table 38.3 FIR of the pulse shaping filter used in cdma2000 [50]

m'	$h[m']$	m'	$h[m']$	m'	$h[m']$
0, 47	−0.02528832	8, 39	0.03707116	16, 31	−0.01283966
1, 46	−0.03416793	9, 38	−0.02199807	17, 30	−0.14347703
2, 45	−0.03575232	10, 37	−0.06071628	18, 29	−0.21182909
3, 44	−0.01673370	11, 36	−0.05117866	19, 28	−0.14051313
4, 43	0.02160251	12, 35	0.00787453	20, 27	0.09460192
5, 42	0.06493849	13, 34	0.08436873	21, 26	0.44138714
6, 41	0.09100214	14, 33	0.12686931	22, 25	0.78587564
7, 40	0.08189497	15, 32	0.09452834	23, 24	1.0

Source: 3GPP2, "Physical layer standard for cdma2000 spread spectrum systems (C.S0002-E)," 3rd Generation Partnership Project 2 (3GPP2), TS C. S0002-E, June 2011. Reproduced with permission of 3CPP2.

where $\hat{\theta}$ is the beat carrier phase estimate, and h is a pulse shaping filter, which is a discrete-time version of the one used to shape the spectrum of the transmitted signal, with a finite-impulse response (FIR) given in Table 38.3. The samples m' of the FIR in Table 38.3 are spaced by $\frac{T_c}{4}$.

Next, $x[m]$ is correlated with a local replica of the spreading PN sequence. In a digital receiver, the correlation operation is expressed as

$$Z_k = \frac{1}{N_s} \sum_{m=k}^{k+N_s-1} x[k]\left\{c_I\left[t_m - \hat{t}_s(t_m)\right] + jc_Q\left[t_m - \hat{t}_s(t_m)\right]\right\} \tag{38.7}$$

$$\triangleq I_k + jQ_k,$$

where Z_k is the k-th subaccumulation, N_s is the number of samples per subaccumulation, and $\hat{t}_s(t_m)$ is the code start time estimate over the k-th subaccumulation. The code phase can be assumed to be approximately constant over a short subaccumulation interval $T_{\text{sub}} = N_s T_s$; hence, $\hat{t}_s(t_m) \approx \hat{t}_{s_k}$. It is worth mentioning that theoretically, T_{sub} can be made arbitrarily large since no data is transmitted on the pilot channel. Practically, T_{sub} is mainly limited by the stability of the BTS and receiver oscillators. In the following, T_{sub} is set to one PN code period. The carrier phase estimate is modeled as $\hat{\theta}(t_m) = 2\pi \hat{f}_{D_k} t_m + \theta_0$, where $\hat{f}_{D_k}$ is the apparent Doppler frequency estimate over the i-th subaccumulation, and θ_0 is the initial beat carrier phase of the received signal. As in a GPS receiver, the value of θ_0 is set to zero in the acquisition stage and is subsequently updated in the tracking stage. The apparent Doppler frequency is assumed to be constant over a short T_{sub}. Substituting for $r[m]$ and $x[m]$, defined in Eqs. (38.5)–(38.6), into Eq. (38.7), it can be shown that

$$Z_k = \sqrt{C} R_c(\Delta t_k)\left[\frac{1}{N_s} \sum_{m=k}^{k+N_s-1} e^{j\Delta\theta(t_m)}\right] + n_k, \tag{38.8}$$

where R_c is the autocorrelation function of the PN sequences c'_I, and c'_Q, $\Delta t_k \triangleq \hat{t}_{s_k} - t_{s_k}$ is the code phase error, $\Delta\theta(t_m) \triangleq \theta(t_m) - \hat{\theta}(t_m)$ is the carrier phase error, and $n_k \triangleq n_{I_k} + jn_{Q_k}$ with n_{I_k} and n_{Q_k} being independent and identically distributed Gaussian random sequences with zero mean and variance $\frac{N_0}{2T_s N_s} = \frac{N_0}{2T_{\text{sub}}}$.

The expression for Z_k in Eq. (38.8) assumes that the locally generated c_I and c_Q have the same code phase. To ensure this, both sequences must begin with the first binary "1" that occurs after 15 consecutive zeros; otherwise, $|Z_k|$ will be halved. Figure 38.9 shows $|Z_k|^2$ for unsynchronized and synchronized c_I and c_Q code phases (i.e. shifted by 34 chips). The correlation peak of the synchronized codes is four times the peak of the unsynchronized case.

The carrier wipe-off and correlation stages are illustrated in Figure 38.10.

38.5.2.2 Acquisition

The objective of this stage is to determine which BTSs are in the receiver's proximity and to obtain a coarse estimate of their corresponding code start times and Doppler frequencies. For a particular PN offset, a search over the code start time and Doppler frequency is performed to detect the presence of a signal. To determine the range of Doppler frequencies to search over, one must consider the relative motion between the receiver and the BTS and the stability of the receiver's oscillator. For instance, a Doppler shift of 122 Hz will be observed for a cellular CDMA carrier frequency of 882.75 MHz at a mobile receiver with a

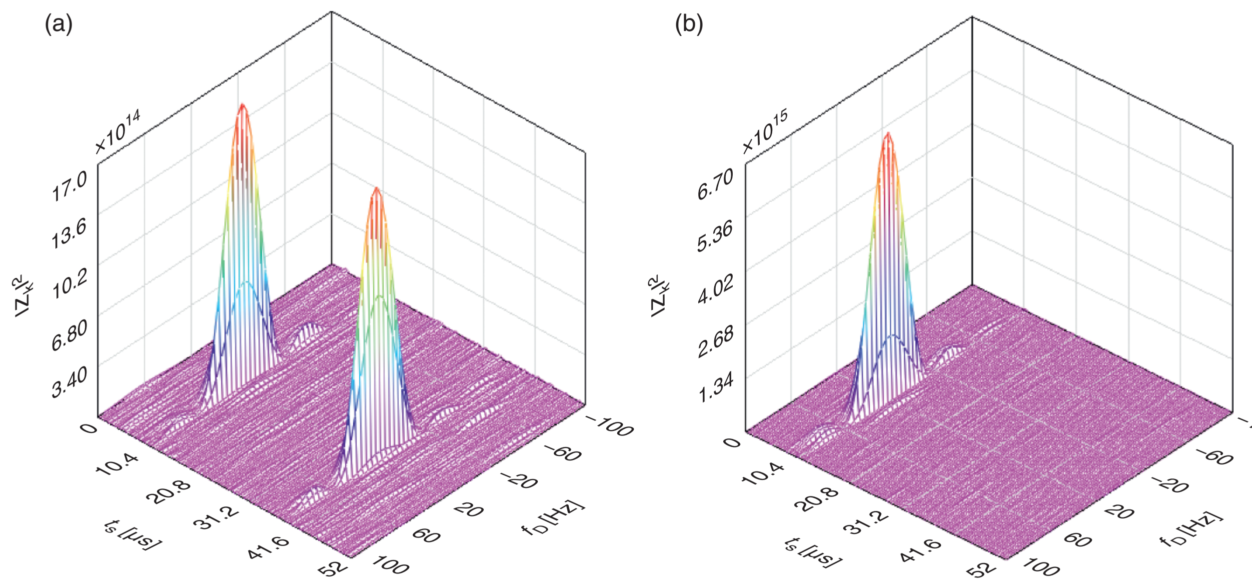

Figure 38.9 $|Z_k|^2$ for (a) unsynchronized and (b) synchronized c_I and c_Q codes (Khalife et al. [18]). *Source:* Reproduced with permission of IEEE.

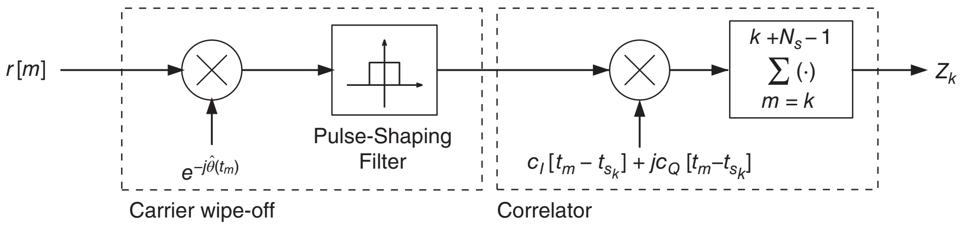

Figure 38.10 Carrier wipe-off and correlator. Thick lines indicate a complex-valued variable (Khalife et al. [18]). *Source:* Reproduced with permission of IEEE.

receiver-to-BTS line-of-sight velocity of 150 km/h. Therefore, to account for this Doppler (at a carrier frequency of 882.75 MHz) as well as oscillator-induced Doppler, the Doppler frequency search window is chosen to lie between −500 and 500 Hz. The frequency spacing Δf_D must be a fraction of $1/T_{\text{sub}}$, which implies that $\Delta f_D \ll 37.5$ Hz, if T_{sub} is assumed to be one PN code period (e.g. a Δf_D between 8 and 12 Hz can be chosen). The code start time search window is naturally chosen to be one PN code interval with a delay spacing of one sample.

Similar to GPS signal acquisition, the search could be implemented either serially or in parallel, which in turn could be performed over the code phase or the Doppler frequency. The receiver presented here performs a parallel code phase search by exploiting the optimized efficiency of the fast Fourier transform (FFT) [53]. If a signal is present, a plot of $|Z_k|^2$ will show a high peak at the corresponding code start time and Doppler frequency estimates. A hypothesis test could be performed to decide whether the peak corresponds to a desired signal or noise. Since there is only one PN sequence, the search needs to be performed once. Then, the resulting surface is subdivided in the time axis into intervals of 64 chips, each division corresponding to a particular PN offset. The PN sequences for the pilot, sync, and paging channels could be generated off-line and stored in a binary file to speed up the processing. Figure 38.11 depicts the acquisition stage of a cellular CDMA signal with a software-defined receiver (SDR) developed in LabVIEW, showing $|Z_k|^2$ along with $\hat{t}_{s_k}$, $\hat{f}_{D_k}$, the PN offset, and the carrier-to-noise ratio C/N_0 for a particular BTS [18].

38.5.2.3 Tracking

After obtaining an initial coarse estimate of the code start time $\hat{t}_{s_k}$ and Doppler frequency $\hat{f}_{D_k}$, the receiver refines and maintains these estimates via tracking loops. A phase-locked loop (PLL) or a frequency-locked loop

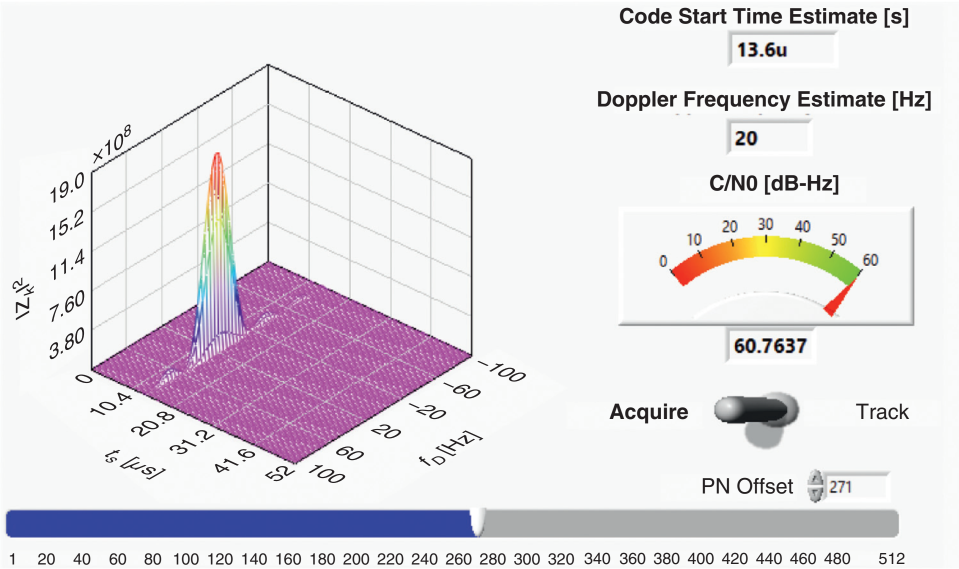

Figure 38.11 Cellular CDMA signal acquisition front panel showing $|Z_k|^2$ along with $\hat{t}_{s_k}$, $\hat{f}_{D_k}$, PN offset, and C/N_0 for a particular BTS (Khalife et al. [18]). *Source:* Reproduced with permission of IEEE.

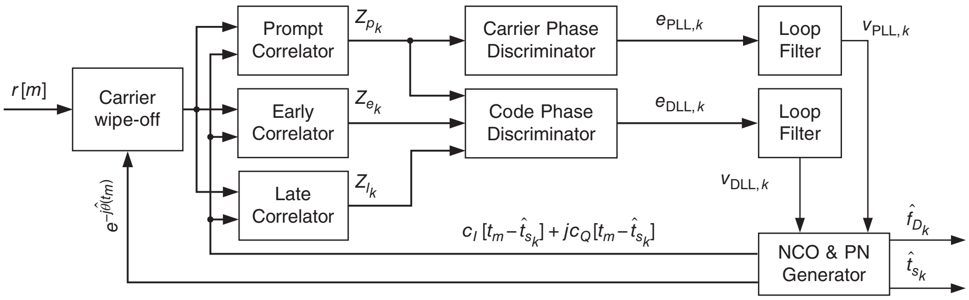

Figure 38.12 Tracking loops in a navigation cellular CDMA receiver. Thick lines represent complex quantities (Khalife et al. [18]). *Source:* Reproduced with permission of IEEE.

(FLL) can be employed to track the carrier phase, and a carrier-aided delay-locked loop (DLL) can be used to track the code phase. FLLs are generally more robust than PLLs, are useful when transitioning from acquisition to tracking, and can track in more challenging environments [54, 55]. Figure 38.12 depicts a block diagram of a PLL-aided DLL tracking loop [12, 18]. The PLL and DLL are discussed in detail next.

PLL: The PLL consists of a phase discriminator, a loop filter, and a numerically controlled oscillator (NCO). Since the receiver is tracking the data-less pilot channel, an `atan2` discriminator can be used, given by

$$e_{\mathrm{PLL},k} = \mathrm{atan2}\left(Q_{p_k}, I_{p_k}\right),$$

where $Z_{p_k} = I_{p_k} + jQ_{p_k}$ is the prompt correlation. The `atan2` discriminator remains linear over the full input error range of $\pm\pi$ and could be used without the risk of introducing phase ambiguities. In contrast, a GPS receiver cannot use this discriminator unless the transmitted data bit values of the navigation message are known [54]. Furthermore, while GPS receivers require second- or higher-order PLLs due to the high dynamics of GPS SVs, lower-order PLLs could be used in cellular CDMA navigation receivers. It was found that the receiver could easily track the carrier phase with a second-order PLL with a loop filter transfer function given by

$$F_{\mathrm{PLL}}(s) = \frac{2\zeta\omega_n s + \omega_n^2}{s}, \tag{38.9}$$

where $\zeta \equiv \frac{1}{\sqrt{2}}$ is the damping ratio, and ω_n is the undamped natural frequency, which can be related to the PLL noise-equivalent bandwidth $B_{n,\mathrm{PLL}}$ by $B_{n,\mathrm{PLL}} = \frac{\omega_n}{8\zeta}(4\zeta^2 + 1)$ [55]. The output of the loop filter $v_{\mathrm{PLL},\,k}$ is the rate of change of the carrier phase error, expressed in rad/s. The Doppler frequency is deduced by dividing $\nu_{\mathrm{PLL},\,k}$ by 2π. The loop filter transfer function in Eq. (38.9) is discretized and realized in state space. The noise-equivalent bandwidth is chosen to range between 4 and 8 Hz.

DLL: The carrier-aided DLL employs a non-coherent dot-product discriminator given by

$$e_{\mathrm{DLL},k} = \Lambda\left[(I_{e_k} - I_{l_k})I_{p_k} + \left(Q_{e_k} - Q_{l_k}\right)Q_{p_k}\right],$$

where Λ is a normalization constant given by $\Lambda = T_c/2C$; C is the carrier power, which can be estimated from the prompt correlation; and $Z_{p_k} = I_{p_k} + jQ_{p_k}$, $Z_{e_k} = I_{e_k} + jQ_{e_k}$, and $Z_{l_k} = I_{l_k} + jQ_{l_k}$ are the prompt, early, and late correlations, respectively. The prompt correlation was described in Section 38.5.2.1. The early and late correlations are calculated by correlating the received signal with an early and a delayed version of the prompt PN sequence, respectively. The time shift between Z_{e_k} and Z_{l_k} is defined by an early-minus-late time t_{eml}, expressed in chips. Since the autocorrelation function of the transmitted cellular CDMA pulses is not triangular as in the case of GPS, a wider t_{eml} is preferable in order to have a significant difference between Z_{p_k}, Z_{e_k}, and Z_{l_k}. Figure 38.13 shows the autocorrelation function of the cellular CDMA PN code as specified by the cdma2000 standard and that of the C/A code in GPS. It can be seen from Figure 38.13 that for $t_{\mathrm{eml}} \le 0.5$ chips, $R_c(\tau)$ in the cdma2000 standard has an approximately constant value, which is not desirable for precise tracking. A good rule of thumb is to choose $1 \le t_{\mathrm{eml}} \le 1.2$ chips.

The DLL loop filter is a simple gain K, with a noise-equivalent bandwidth $B_{n,\mathrm{DLL}} = \frac{K}{4} \equiv 0.5$ Hz. The output of the DLL loop filter $\nu_{\mathrm{DLL},\,k}$ is the rate of change of the code phase, expressed in s/s. Assuming low-side mixing, the code start time is updated according to

$$\hat{t}_{s_{k+1}} = \hat{t}_{s_k} - \left(\nu_{\mathrm{DLL},k} + \hat{f}_{D_k}/f_c\right) \cdot N_s T_s.$$

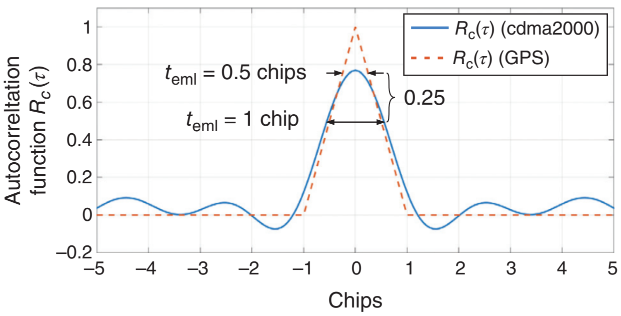

Figure 38.13 Autocorrelation function of GPS C/A code and cellular CDMA PN sequence according to the cdma2000 standard (Khalife et al. [12]). *Source:* Reproduced with permission of IEEE.

In a GPS receiver, the pseudorange is calculated based on the time a navigation message subframe begins, which eliminates ambiguities due to the relative distance between GPS SVs [55]. This necessitates decoding the navigation message in order to detect the start of a subframe. These ambiguities do not exist in a cellular CDMA system. This follows from the fact that a PN offset of one translates to a distance greater than 15 km between BTSs, which is beyond the size of a typical cell [56].

Finally, the pseudorange estimate ρ can be deduced by multiplying the code start time by the speed of light c; that is,

$$\rho(k) = c \cdot \hat{t}_{s_k}. \tag{38.10}$$

Figure 38.14 shows the intermediate signals produced within the tracking loops of the cellular CDMA navigation receiver: code error; phase error; Doppler frequency; early, prompt, and late correlations; pseudorange; and in-phase and quadrature components of the correlation.

38.5.2.4 Message Decoding

Demodulating the sync and paging channel signals is performed similarly to the pilot signal but with two major differences: (i) the locally generated PN sequence is furthermore spread by the corresponding Walsh code and (ii) the subaccumulation period is bounded by the data symbol interval. In contrast to GPS signals, in which a data bit stretches over 20 C/A codes, a sync data symbol comprises only 256 PN chips, and a paging channel data symbol comprises 128 chips. After carrier wipe-off, the sync and paging signals are processed in the reverse order of the steps illustrated in Figures 38.5 and 38.7, respectively. It is worth noting that the start of the sync message always coincides with the start of the PN code, and the corresponding paging channel message starts after 320 ms minus the PN offset (expressed in seconds), as shown in Figure 38.15. Recall that the long code is also used to spread the paging message in the downlink (see Figure 38.7). The long code state decoded from a sync message is valid at the beginning of the corresponding paging channel message.

The long code is generated by masking the outputs of the 42 registers and computing the modulo-two sum of the resulting bits. In contrast to the short code generator in cellular CDMA and the C/A code generator in GPS, the 42 long code generator registers are configured to satisfy a linear recursion given by

$$\begin{aligned} p(x) = {} & x^{42} + x^{35} + x^{33} + x^{31} + x^{27} + x^{25} + x^{22} + x^{21} \\ & + x^{19} + x^{18} + x^{17} + x^{16} + x^{10} + x^{7} \\ & + x^{6} + x^{5} + x^{3} + x^{2} + x + 1. \end{aligned}$$

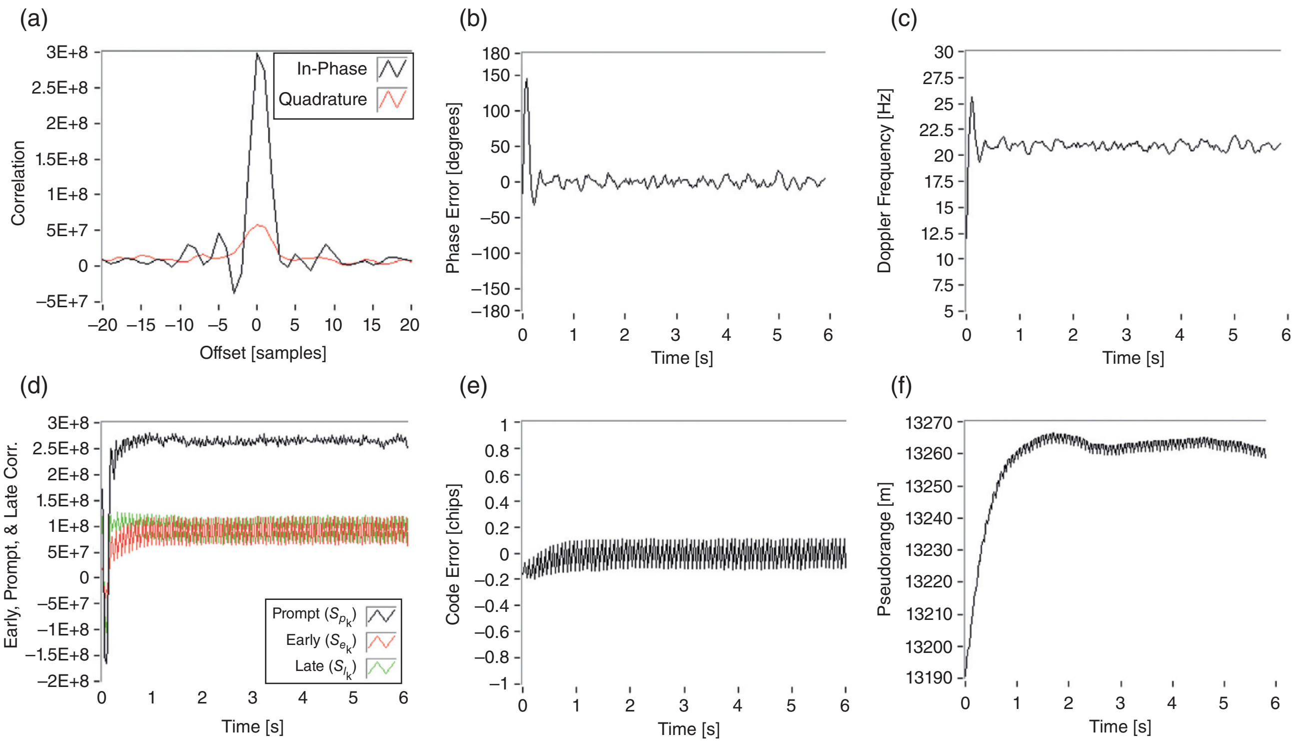

Figure 38.14 Cellular CDMA signal tracking: (a) code phase error (chips), (b) carrier phase error (degrees), (c) Doppler frequency estimate (hertz), (d) prompt (black), early (red), and late (green) correlation, (e) measured pseudorange (m), and (f) correlation function (Khalife et al. [18]). *Source:* Reproduced with permission of IEEE.

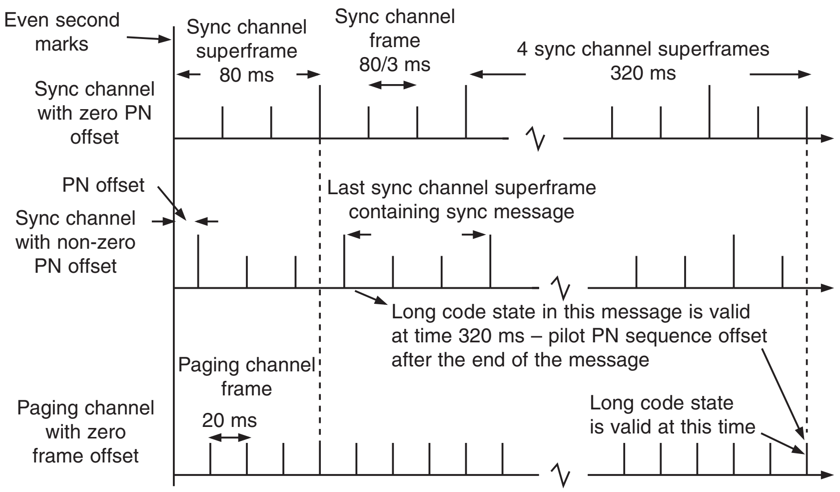

Figure 38.15 Sync and paging channel timing (Khalife et al. [18]; 3GPP2 [50]). *Source:* Reproduced with permission of IEEE.

The long code mask is obtained by combining the PN offset and the paging channel number p as shown in Figure 38.16.

Subsequently, the sync message is decoded first, and the PN offset, the paging channel number, and the long code state are then used to descramble and decode the paging message. It is important to note that the long code is first decimated at a rate of 1/64 to match the paging channel symbol rate. More details are specified in [47]. Figure 38.17 shows the demodulated sync signal as well as the final information decoded from the sync and paging channels. Note that the shown signal corresponds to the US cellular provider Verizon, which does not broadcast its BTS position information (latitude and longitude). Moreover,