Angle-of-Arrival Based Ultrasonic 3-D Location for Ubiquitous Computing

by

Tarig Ballal Khidir Ahmed

A Thesis submitted to
University College Dublin
for the degree of Doctor of Philosophy
in the College of Engineering, Mathematical & Physical Sciences
School of Computer Sciences & Informatics

July 2010

Department of Computer Science
Prof. Joe Carthy (Head of Department)
Under the supervision of
Dr. Chris Bleakley
## Contents

Acknowledgements xv

Abstract xviii

1 Introduction 1
  1.1 Motivation ........................................ 1
  1.2 Problem Statement ................................ 3
  1.3 Approach ........................................ 4
    1.3.1 Signal Design .................................. 5
    1.3.2 Acquisition and Synchronization ............... 5
    1.3.3 AOA Estimation ................................ 5
    1.3.4 Location and Orientation Estimation .......... 5
  1.4 Thesis Contributions ............................. 6
  1.5 Structure of the Thesis .......................... 6

2 Background and Related Work 9
  2.1 Introduction .................................... 9
  2.2 Acoustic Wave Propagation ....................... 10
    2.2.1 Attenuation .................................. 11
    2.2.2 Noise ........................................ 12
    2.2.3 Effect of Obstacles on Acoustic Wave Propagation 12
  2.3 Angle-of Arrival Estimation ..................... 13
    2.3.1 Array Methods ................................ 14
    2.3.2 Time-Delay Methods ........................... 17
  2.4 Sparse Arrays .................................. 19
  2.5 Phase-Difference Ambiguity ...................... 21
  2.6 Spread Spectrum ................................. 22
    2.6.1 Direct Sequence Spread Spectrum .............. 23
    2.6.2 Frequency-Hopping Spread Spectrum .......... 24
  2.7 Location Techniques ............................. 25
    2.7.1 Related systems .............................. 27
  2.8 Summary ........................................ 27
### 3 Signal Design

- **3.1 Introduction** .................................................. 29
- **3.2 The Orthogonality of Signals** ...................................... 31
- **3.3 FHSS and Multiple Access** ........................................ 32
- **3.4 Multipath Interference** ........................................... 33
  - 3.4.1 Early Multipath: The Concept of Autocontamination ............. 33
  - 3.4.2 Late Multipath .................................................... 35
- **3.5 Design for Limited Bandwidth** ...................................... 35
  - 3.5.1 Requirements ..................................................... 35
  - 3.5.2 Design Trade-offs ............................................... 37
- **3.6 Summary** ............................................................ 39

### 4 Acquisition and Synchronization

- **4.1 Introduction** .................................................... 41
- **4.2 All-digital Receiver Implementation** .................................. 45
- **4.3 Modulation Scheme** .............................................. 46
  - 4.3.1 Differential Binary Phase-Shift Keying (DBPSK) Modulation .......... 47
- **4.4 The Proposed Acquisition and Synchronization System** .................. 53
  - 4.4.1 Preamble Detection and Coarse Synchronization ...................... 54
  - 4.4.2 Preamble Decoding and Fine Synchronization ....................... 57
  - 4.4.3 Multiple Transmitters ......................................... 60
  - 4.4.4 Payload Synchronization ........................................ 75
- **4.5 Practical System and Experiments** .................................. 83
  - 4.5.1 Experimental Setup ............................................ 84
- **4.6 Summary** ............................................................ 85

### 5 Spatial-Diversity-Based Phase-Difference Ambiguity Resolution

- **5.1 Introduction** .................................................... 89
- **5.2 The Phase-Difference Ambiguity Problem** .............................. 91
- **5.3 Condition for the Occurrence of Ambiguity** .......................... 93
- **5.4 General Formulation of the Disambiguation Method** ................. 95
- **5.5 Disambiguation in the Ideal Far-Field** .................................. 99
- **5.6 Disambiguation under Non-Ideal Far-Field Conditions** ................ 100
  - 5.6.1 Disambiguation in the GAR ..................................... 102
  - 5.6.2 Disambiguation in the GIR ..................................... 106
  - 5.6.3 Summary of the Disambiguation Algorithm ............................ 107
- **5.7 Error Analysis** .................................................... 107
  - 5.7.1 Error Analysis for MAV Step .................................... 109
  - 5.7.2 Error Analysis for the RO Step .................................. 110
  - 5.7.3 Noise-Free Success/Failure ..................................... 111
  - 5.7.4 Success/Failure under Noise .................................... 112
  - 5.7.5 Performance Versus SNR ......................................... 112
  - 5.7.6 Simulation and Analytical Results .................................. 118
- **5.8 Experimental Results** ............................................ 120
- **5.9 Summary** ............................................................ 123
# CONTENTS

6 Frequency-Diversity-Based Phase-Difference Ambiguity Resolution 125
   6.1 Introduction ................................................................. 125
   6.2 Phase Unwrapping for Two Frequency Components ............... 127
      6.2.1 The Noise-Free Case ............................................... 127
      6.2.2 The Noisy Case ................................................... 130
      6.2.3 Summary of the Disambiguation Algorithm ................... 131
   6.3 Noise Effect ............................................................. 132
      6.3.1 Noise Effect on the MAV Step ................................ 132
      6.3.2 Noise Effect on the RO Step ................................... 133
      6.3.3 Performance Versus SNR ......................................... 135
      6.3.4 Simulation and Analytical Results ............................ 136
   6.4 Experimental Results .................................................. 140
   6.5 Summary ................................................................. 141

7 Angle-of-Arrival Estimation ............................................ 143
   7.1 Introduction ............................................................. 143
   7.2 AOA Estimation: the Single-Source Case .......................... 144
      7.2.1 AOA Estimation with Phase Unwrapping ...................... 146
      7.2.2 Parameter Consideration ....................................... 147
   7.3 AOA Estimation: the Multi-Source Case ............................ 149
   7.4 Search-Based AOA Estimation Methods .............................. 151
      7.4.1 Spatial Diversity Based ....................................... 151
      7.4.2 Frequency Diversity Based ................................... 153
   7.5 The Effect of Cycle Slipping ........................................ 153
   7.6 Experimental Results .................................................. 156
      7.6.1 Single-Source ..................................................... 157
      7.6.2 Multi-Source ...................................................... 161
   7.7 Summary ................................................................. 162

8 Three-Dimensional Location and Orientation Estimation .............. 163
   8.1 Introduction ............................................................. 163
   8.2 Problem Description ................................................... 164
   8.3 Proposed Solutions ..................................................... 164
      8.3.1 Solutions Using Four Sensors ................................. 165
   8.4 Simulation Results ..................................................... 171
   8.5 Summary ................................................................. 175

9 AOA-Based Ultrasonic Location-Orientation System .................. 177
   9.1 Introduction ............................................................. 177
   9.2 The Proposed FHSS Design ........................................... 178
   9.3 AOA Estimation .......................................................... 179
      9.3.1 Direction Vectors Estimation ................................. 180
   9.4 Experiments ............................................................. 182
      9.4.1 System Components ............................................. 182
      9.4.2 Experimental Setup ............................................. 183
      9.4.3 Results ........................................................... 185
   9.5 Summary ................................................................. 199
1.1 Ultrasonic 3-D location-orientation problem. 3
1.2 Schematic for the ultrasonic 3-D location-orientation problem solution. 4
2.1 Spherical coordinates (|r|, θ, φ) and the corresponding Cartesian coordinates (x, y, z). 11
2.2 A uniform circular array geometry. 16
2.3 A uniform linear array geometry. 16
2.4 Azimuthal angle for emitter in the far-field of a linear array. 16
2.5 Geometrical model for time-delay estimation under the far-field assumption. 19
2.6 Example array beam-patterns. 21
2.7 An example of a hopping pattern. 25
2.8 FHSS generation. 25
3.1 Multipath generation. 35
3.2 Hop separation: f1, f2 and f3 are hopping frequencies. 38
4.1 Matched-filter coarse synchronization system. 44
4.2 Serial-search coarse synchronization system. 45
4.3 Example for a) BPSK and b) DBPSK signals for a transmitted binary sequence: 0, 1, 1 and 0. 48
4.4 Block diagram of a DBPSK transmitter. 48
4.5 Block diagram of a general receiver configuration. 49
4.6 Block diagram of a DBPSK quadrature receiver. 51
4.7 Probability of bit error (Pe) for DBPSK modulation versus the transmitted energy per bit to noise power ratio (Eb/η0). 52
4.8 Block diagram of a DBPSK receiver alternative structure. 52
4.9 Block diagram of digital implementation of a DBPSK receiver. 52
4.10 Data frame format. 54
4.11 Preamble detection system. 55
4.12 Correlations of s01[n]. 61
4.13 Correlations of $s_{11}[n]$. .................................................. 62
4.14 Decoding of a preamble hop. ........................................... 63
4.15 Noise effect on $|\Delta E_k|$. .............................................. 64
4.16 RMSE for $|\Delta E_k|$ peak location versus $E_b/\eta_0$. .......... 64
4.17 The probability of slot collision ($P_{sc}$) versus the number of payload hops ($H$). The number of preamble hops is $RK = 8$. 69
4.18 The probability of at least one bit error ($P_{e,1}$) for DBPSK modulation versus the transmitted energy per bit to noise power ratio ($E_b/\eta_0$). The number of digits $RK = 8$ .................................. 69
4.19 Probability of acquisition failure ($P_f$) versus average preamble transmission rate ($p$) for different number of transmitters ($N_b$), different values of $E_b/\eta_0$, $RK = 8$, $m = 4$ and $H = 14$. ................................. 76
4.20 Probability of acquisition failure ($P_f$) versus average preamble transmission rate ($p$) for different payload lengths ($H$), different values of $E_b/\eta_0$, $RK = 8$, $m = 4$ and $N_b = 3$ ........................................ 77
4.21 Mean acquisition time ($\mu_{acq}$) versus average preamble transmission rate ($p$) for different number of transmitters ($N_b$), different values of $E_b/\eta_0$, $RK = 8$, $m = 4$ and $H = 14$ ............... 78
4.22 Mean acquisition time ($\mu_{acq}$) versus average preamble transmission rate ($p$) for different payload lengths ($H$), different values of $E_b/\eta_0$, $RK = 8$, $m = 4$ and $N_b = 3$ ............................ 79
4.23 Standard deviation of acquisition time ($\sigma_{acq}$) versus average preamble transmission rate ($p$) for different number of transmitters ($N_b$), different values of $E_b/\eta_0$, $RK = 8$, $m = 4$ and $H = 14$. .......................................................... 80
4.24 Standard deviation of acquisition time ($\sigma_{acq}$) versus average preamble transmission rate ($p$) for different payload lengths ($H$), different values of $E_b/\eta_0$, $RK = 8$, $m = 4$ and $N_b = 3$. ......................... 81
4.25 Minimum root mean squared value of acquisition time ($\zeta_{\text{min}}$) versus payload length ($H$) for different number of transmitters ($N_b$), different values of $E_b/\eta_0$, $RK = 8$ and $m = 4$. The locations of the minima of $\zeta_{\text{min}}$ are marked with `x'. ......................... 82
4.26 The mean ($\mu$), standard deviation ($\sigma$) and RMSV ($\zeta$) of acquisition time for $(p, H) = (0.90, 70)$. .................................................. 85
4.27 Experimental Setup. ....................................................... 86
5.1 The proposed sensor configuration: The far-field model. ...... 90
5.2 The proposed sensor configuration: The general model. ....... 91
5.3 Deviations (in degrees) of AOA calculated based on Eq. (5.7) from the true emitter angle viewed from a) sensor 1; b) sensor 2; c) the center of the two sensor baseline. Sensor separation is 20 mm with the sensor baseline aligned along the x-axis and centered at the origin. ................. 94
5.4 The values of $\varphi$ for $f = 40$ kHz at a propagation speed of 343 m/s and $d_{12}$ set equal to: a) $1.01\lambda/2$; b) $2\lambda/2$; c) $6\lambda/2$; and d) $20\lambda/2$; $\Delta_d \approx 0.7\lambda/2$. The sensor baseline is aligned along the x-axis with $d_{12}$ centered at the origin. Resolution is 50 mm. .................................................. 100

5.5 The values of $\varphi$ at $y = 5$ m for different values of $d_{12}$ and $\Delta_d \approx 0.7\lambda/2$. Sensor 1 and sensor 2 baseline is aligned along the x-axis and centered at the origin. Resolution is 1 mm. .................................................. 101

5.6 The GAR (red) and the GIR (blue) for $f = 40$ kHz at a propagation speed of 343 m/s and $d_{12}$ set equal to: a) $1.01\lambda/2$; b) $2\lambda/2$; c) $6\lambda/2$; and d) $20\lambda/2$; $\Delta_d \approx 0.7\lambda/2$. The sensor baseline is aligned along the x-axis with $d_{12}$ centered at the origin. .................................................. 102

5.7 The wrapping (blue) and wrapping-free (red) regions for $f = 40$ kHz at a propagation speed of 343 m/s and $d_{12}$ set equal to: a) $1.01\lambda/2$; b) $2\lambda/2$; c) $6\lambda/2$; and d) $20\lambda/2$; $\Delta_d \approx 0.7\lambda/2$. The sensor baseline is aligned along the x-axis and centered at the origin. .................................................. 103

5.8 The values of $k$ for $f = 40$ kHz at a propagation speed of 343 m/s and $d_{12}$ set equal to: a) $1.01\lambda/2$; b) $2\lambda/2$; c) $6\lambda/2$; and d) $20\lambda/2$; $\Delta_d \approx 0.7\lambda/2$. The sensor baseline is aligned along the x-axis with $d_{12}$ centered at the origin. .................................................. 104

5.9 The success (red) and failure (blue) of the MAV algorithm under no noise for $f = 40$ kHz at a propagation speed of 343 m/s, and $d_{12}$ set equal to: a) $1.01\lambda/2$; b) $2\lambda/2$; c) $6\lambda/2$; and d) $20\lambda/2$; $\Delta_d \approx 0.7\lambda/2$. The sensor baseline is aligned along the x-axis with $d_{12}$ centered at the origin. .................................................. 113

5.10 The success (red) and failure (blue) of the RO step under no noise for $f = 40$ kHz at a propagation speed of 343 m/s and $d_{12}$ set equal to: a) $1.01\lambda/2$; b) $2\lambda/2$; c) $6\lambda/2$; and d) $20\lambda/2$; $\Delta_d \approx 0.7\lambda/2$. The sensor baseline is aligned along the x-axis with $d_{12}$ centered at the origin. .................................................. 114

5.11 The probability of success for the MAV algorithm for $f = 40$ kHz at a propagation speed of 343 m/s and $d_{12}$ set equal to $20\lambda/2$; $\Delta_d \approx 0.7\lambda/2$: a) $\sigma_p^2=-150$ dB; b) $\sigma_p^2=-130$ dB; c) $\sigma_p^2=-110$ dB. The sensor baseline is aligned along the x-axis with $d_{12}$ centered at the origin. .................................................. 115

5.12 The probability of success for the MAR algorithm for $f = 40$ kHz at a propagation speed of 343 m/s and $d_{12}$ set equal to $20\lambda/2$; $\Delta_d \approx 0.7\lambda/2$: a) $\sigma_p^2=-150$ dB; b) $\sigma_p^2=-130$ dB; c) $\sigma_p^2=-110$ dB. The sensor baseline is aligned along the x-axis with $d_{12}$ centered at the origin. .................................................. 116

5.13 The MSE for time-delay estimation: a) The MAV step b) the whole disambiguation algorithm. .................................................. 120

5.14 The (analytical) probability of success of the MAV step and that of the whole proposed disambiguation algorithm corresponding to Fig. 5.13. .................................................. 121
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.15</td>
<td>The MSE for time-delay estimation for different values of $\Delta d$, fixed $D_{12} = 5$ and emitter at $[1166, 1166, 0]^T$.</td>
</tr>
<tr>
<td>5.16</td>
<td>The MSE for time-delay estimation for different values of $D_{12}$, fixed $\Delta d = 1$ and emitter at $[1166, 1166, 0]^T$.</td>
</tr>
<tr>
<td>5.17</td>
<td>The MSE for time-delay estimation for different emitter distances/locations, fixed $D_{12} = 20$ and fixed $\Delta d = 1$.</td>
</tr>
<tr>
<td>5.18</td>
<td>The test room with the locations of the transmitter (Tx) and the receiver (Rx) marked. Heights (hgt) are quoted between brackets.</td>
</tr>
<tr>
<td>6.1</td>
<td>The MSE in the estimation of the delay corresponding to the unambiguous phase-difference obtained from a) the MAV step and b) the whole disambiguation algorithm.</td>
</tr>
<tr>
<td>6.2</td>
<td>The (analytical) probability of success of the MAV step and that of the whole proposed disambiguation algorithm corresponding to Fig. 6.1.</td>
</tr>
<tr>
<td>6.3</td>
<td>The MSE in the estimation of the delay corresponding to the unambiguous phase-difference obtained from the proposed disambiguation algorithm, for different values of frequency separation and $\Delta w$ ($2\pi$ k rad).</td>
</tr>
<tr>
<td>6.4</td>
<td>The MSE in the estimation of the delay corresponding to the unambiguous phase-difference obtained from the proposed disambiguation algorithm, for different values of sensor spacing and $d$ ($\lambda/2$).</td>
</tr>
<tr>
<td>6.5</td>
<td>The test room with the locations of the transmitter (Tx) and the receiver (Rx) marked. Heights (hgt) are quoted between brackets.</td>
</tr>
<tr>
<td>7.1</td>
<td>The error function $J(\theta)$ (normalized) for an AOA equal to $0^\circ$ for selected inter-sensor spacings (given in half-wavelengths). Far-field is assumed.</td>
</tr>
<tr>
<td>7.2</td>
<td>The probability of cycle slipping in space for four different frequencies. The distance between the two sensors is $d_{12} = 2$ cm. The sensor baseline is aligned along the x-axis, with the center of the baseline positioned at the origin. Resolution is 1 cm.</td>
</tr>
<tr>
<td>7.3</td>
<td>Calibration error effect.</td>
</tr>
<tr>
<td>7.4</td>
<td>The test room with the locations of the transmitter (Tx) and the receiver (Rx) marked. Heights (hgt) are quoted between brackets.</td>
</tr>
<tr>
<td>8.1</td>
<td>Sensor and beacon configurations.</td>
</tr>
<tr>
<td>8.2</td>
<td>Limitation of the angle $\gamma$. Plain arrows point to the direction of LOS while the arrows with squares point to the direction of NLOS (Non LOS).</td>
</tr>
<tr>
<td>8.3</td>
<td>3-D location error surface for solution 1.</td>
</tr>
<tr>
<td>8.4</td>
<td>3-D location error surface for solution 2.</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>----------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>8.5</td>
<td>Orientation error surface for solution 1: a) x-axis b) y-axis c) z-axis.</td>
</tr>
<tr>
<td>8.6</td>
<td>Orientation error surface for solution 2: a) x-axis b) y-axis c) z-axis.</td>
</tr>
<tr>
<td>9.1</td>
<td>The proposed sensor configuration for 2-D AOA estimation.</td>
</tr>
<tr>
<td>9.2</td>
<td>Experimental setup: the grid indicates the locations of the receiver.</td>
</tr>
<tr>
<td>9.3</td>
<td>System layout.</td>
</tr>
<tr>
<td>9.4</td>
<td>AOA estimation performance: source 1.</td>
</tr>
<tr>
<td>9.5</td>
<td>AOA estimation performance: source 2.</td>
</tr>
<tr>
<td>9.6</td>
<td>AOA estimation performance: source 3.</td>
</tr>
<tr>
<td>9.7</td>
<td>Direction vector estimation performance: source 1.</td>
</tr>
<tr>
<td>9.8</td>
<td>Direction vector estimation performance: source 2.</td>
</tr>
<tr>
<td>9.9</td>
<td>Direction vector estimation performance: source 3.</td>
</tr>
<tr>
<td>9.10</td>
<td>Inter-vector angle estimation performance.</td>
</tr>
<tr>
<td>9.11</td>
<td>Location estimation performance: solution 1.</td>
</tr>
<tr>
<td>9.12</td>
<td>Location estimation performance: solution 2.</td>
</tr>
<tr>
<td>9.13</td>
<td>Orientation estimation performance: solution 1.</td>
</tr>
<tr>
<td>9.14</td>
<td>Orientation estimation performance: solution 2.</td>
</tr>
<tr>
<td>9.15</td>
<td>Convergence: solution 1 and solution 2.</td>
</tr>
<tr>
<td>Table</td>
<td>Description</td>
</tr>
<tr>
<td>-------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>3.1</td>
<td>Examples of FHSS design for $B = 15$ kHz, $f_s = 200$ Hz and $RT = 125$ ms.</td>
</tr>
<tr>
<td>4.1</td>
<td>The number of occurrences of different number of bit errors in 15 tests per preamble when the receiver was close to the wall.</td>
</tr>
<tr>
<td>5.1</td>
<td>Sample experimental results for emitter at $60^\circ$.</td>
</tr>
<tr>
<td>6.1</td>
<td>Sample experimental results for emitter at $60^\circ$.</td>
</tr>
<tr>
<td>7.1</td>
<td>Cycle slipping angles for different acoustic frequencies and $d_{12} = 2$ cm.</td>
</tr>
<tr>
<td>7.2</td>
<td>Experimental results: The RMSE (in degrees) for $\theta = 0^\circ$.</td>
</tr>
<tr>
<td>7.3</td>
<td>Experimental results: The RMSE (in degrees) for $\theta = 20^\circ$.</td>
</tr>
<tr>
<td>7.4</td>
<td>Experimental results: The RMSE (in degrees) for $\theta = 45^\circ$.</td>
</tr>
<tr>
<td>7.5</td>
<td>Experimental results: The RMSE (in degrees) for $\theta = 60^\circ$.</td>
</tr>
<tr>
<td>7.6</td>
<td>Experimental results: The RMSE (in degrees) for 3 sources at $\theta = -43^\circ$, $\theta = -47^\circ$ and $\theta = 31^\circ$ (corresponding to location Tx1, Tx2 and Tx3 in Fig. 7.4).</td>
</tr>
<tr>
<td>8.1</td>
<td>Mean 3-D location and orientation errors.</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENTS

All praise and thanks are due to Allah, The One, The Only and The Indivisible creator and sustainer of the worlds. To Him belongs what is in the heavens and what is on the earth, and what is in between them, and what is beneath the soil. I wish to thank Him for His innumerable favors and bounties and all that He has gifted us with, although indeed, He can never be praised or thanked enough.

I owe my deepest gratitude to my supervisor, Dr. Chris Bleakley. First, for offering me the opportunity to study for Ph.D. in UCD. Second, for his guidance, support and invaluable advices during the course of this work. Thanks, Chris, for being there all the time, for the long, yet interesting and inspiring, discussions that lasted for hours. Thanks are also due to Science Foundation Ireland (SFI) for funding this work.

My sincerest thanks are offered to the postgrads and postdocs in Dr. Chris Bleakley’s research group. Special thanks to Juan, Jong Chern, Kunjan, Mohammed, Nagesh and Séamus. Thanks to all of you for being ready to help all the time. My thanks also go to the staff of The Electronics Workshop for their help and advice in preparing part of the hardware.

A number of people helped to make my life, especially during my first weeks in Dublin, easy and helped me settle down. In this regard, I want to say thanks to Dr. Muawya Hamza and to Mohammed Awadelkarim for their sincere support.

Last but not least, I wish to thank my extended family in Sudan and in the Middle East for their support throughout my educational journey.

Tarig Ballal
Dublin, July 2010
This thesis addresses the problem of simplifying indoor ultrasonic location systems through the disposal of the radio synchronization subsystem. The location approach proposed herein is completely based on angle-of-arrival (AOA) estimation of ultrasonic frequency-hopping spread spectrum (FHSS) signals transmitted by fixed beacons with known positions. The proposed system is privacy-oriented, that is the device to be located only receives and does not transmit signals.

In the proposed location system, the beacons transmit their IDs using ultrasonic signals. The receiver acquires these IDs and determines its room-level position by table lookup. Following beacon identification, the receiver exploits a priori knowledge of the hopping patterns associated with the beacons and a sensor array to determine the AOA of the signal from each beacon. Next, the AOA information is used to determine the receiver’s 3-D location using location algorithms proposed herein. In addition to location, these algorithms yield the receiver orientation.

AOA estimation is achieved by exploiting the phase of the cross-power spectra of the received signals. Special attention is paid to the phase-difference ambiguity problem. Novel disambiguation methods are proposed that facilitate AOA estimation in the ultrasonic case.

The proposed location system was tested experimentally in a prototype. The results confirm the effectiveness of the approach. Accuracy of about 30 cm for location, 5° for the x-axis orientation, 10° for the y-axis orientation and 15° for the z-axis orientation, at an 11 Hz update rate, was obtained. The experiments also reveal practical limitations leading to recommendations for future work.

The contributions of this thesis include novel methods for phase-difference ambiguity resolution exploiting either spatial diversity or frequency diversity. These methods lead to novel AOA estimation techniques for the widely-spaced sensor case. Other contributions are novel algorithms for receiver’s location and orientation estimation using only AOA information. Also, a novel system for binary data transmission using ultrasonic signals, and a novel method for FHSS acquisition and synchronization, are proposed.
1.1 Motivation

Recently, an increasing demand for location information has emerged in the home, office and outdoors. This demand has been accompanied by considerable developments in location technologies. The most notable of these technologies, the Global Positioning System (GPS), is now being used effectively by navigators, at sea and in the air, as well as on land, to find their locations. GPS receivers have also been embedded in different consumer hand-held devices, including mobile phones.

However, the application of GPS is limited to outdoor sites. This limitation is imposed by the physical characteristics of radio waves—they do not penetrate solid objects such as walls. This limitation makes GPS unusable for indoor positioning where the demand for location information has increased substantially in the last two decades. An important emerging area is the evolution of so-called location-aware applications, in which the context of a human or device is needed in order to adapt and personalize the information provided to the user [66].

Various technologies have been proposed for indoor location estimation. These technologies include radio frequency (RF) (e.g., [61, 62]), infra-red (IR) (e.g., [63, 64]) and ultrasonic (e.g., [65, 66, 67, 68, 69]) based systems. However, ultrasonic based location systems have an advantage due to the relatively low speed of sound in air. This allows for accurate computation of signal time-of-flight (TOF), a popular parameter that has been utilized extensively for location estimation.

In ultrasonic 3-D location systems, the 3-D location of a mobile device (MD) is estimated utilizing signals received from or by a number of fixed base stations (BSs). In general, ultrasonic location systems can be classified into two broad
classes according to the roles of the device and the BSs. Systems that have a transmitting MD and receiving BSs are referred to as *infrastructure-based systems* (e.g., [66, 67]). In this class, a central processor is responsible for determining the MD’s location by processing signals received by the BSs. On the other hand, systems that have transmitting BSs (or *beacons*) and a receiving MD are referred to as *privacy-oriented systems* (e.g., [65, 68, 69]). In this class, the MD is responsible for processing the received signals to determine its own location. The latter systems are useful when privacy is required at the application level, such that the MD controls dissemination of its location information.

The ultrasonic location problem has been addressed using various signal properties including TOF (e.g., [65, 66, 67, 68, 69]), difference of time-of-flight (e.g., [70]), and reception pattern change (e.g., [86, 87, 88]). Hybrid systems that combine more than one technique have also emerged (e.g., [56, 58]). The complexity of these systems varies depending on different factors, the most important of which is the need for synchronization between the individual system units. Normally, ultrasonic location systems use a second medium for synchronization. The transmitter typically emits an RF signal to indicate the start of the transmission of the ultrasonic signal. Also, almost all existing systems require some (wired or wireless) means of synchronization between individual beacons. The synchronization subsystem in ultrasonic location systems increases both the complexity and the cost of the system.

Coupled with location, another useful piece of information is the MD’s orientation. For many applications (e.g., [25]), orientation is an important element of the user/device context without which location information might be useless. For example, if the MD is “at the door”, one device orientation could mean that the device is “entering the room”; while the opposite orientation, at the same location, could mean that the MD is “leaving the room”. This is a crucial difference in context.

This work addresses the problem by developing a privacy-oriented ultrasonic system for estimating both the device’s location and orientation. The system is *time-reference-free* in the sense that a separate synchronization signal is not needed. As a result, the system is low-cost, easy-to-deploy and capable of handling multiple-room scenarios.
1.2 Problem Statement

The problem of interest for this thesis can be summarized as follows. Given a number of fixed beacons \( N_b \) that transmit ultrasonic signals with a specific modulation, a receiving MD is required to estimate its own 3-D location and orientation based on these signals. Beacons are located in 3-D space at known coordinates relative to a fixed global coordinate system. The MD contains a number, \( N_s \), of co-planar ultrasonic sensors. The MD has its own (local) coordinate system. The directions of the local coordinate axes in 3-D space represent the MD’s unknown orientation as depicted in Fig. 1.1. A reference point is designated as the origin of the MD’s local coordinate system. The coordinates \((x, y, z)\) of this reference point (given in terms of the global coordinate system) correspond to the 3-D location of the MD that is to be determined together with the device orientation.

Beacons transmit codes or patterns that uniquely identify each individual beacon. The MD is required to extract any identification information from the received signals. The MD is pre-programmed and provided with the IDs of the beacons and their corresponding (room-level) locations in the building. The beacons may be distributed throughout the building in different rooms. At startup, the MD does not know which room it is located in and needs to determine which beacons are in range. Following the identification stage, the MD needs to determine its location and orientation in terms of the global coordinate system.
1.3 Approach

The location approach presented in this thesis is based purely on angle-of-arrival (AOA) estimation at the MD. The selection of an AOA-based approach is motivated by the fact that AOA information is required, anyway, for orientation estimation. In this thesis, it will be shown that AOA information is sufficient to accurately estimate both the location and orientation of a receiver in a way that ensures privacy. In addition, the proposed approach utilizes the transmission of frequency hopping spread spectrum (FHSS) signals. The use of FHSS, in particular, is motivated by its good performance in indoor environment as demonstrated in [58].

The system consists of fixed beacons with known locations transmitting FHSS signals. The signals are designed such that part of the signaling carries information about beacons’ IDs, and part is used for AOA estimation. These signals are transmitted over different frequency bands. At startup, the MD listens to the frequency band devoted to transmission of the beacons’ IDs. When a beacon is detected, the MD decodes the received signal and acquires the received beacon’s ID. The MD keeps listening until it acquires the IDs of at least three beacons. Next, the MD synchronizes to theses beacons and estimates the AOA corresponding to each beacon. AOA estimation is carried out using an array of sensors that is contained on the MD. AOA information is further translated into location and orientation information using the relevant algorithms.

The proposed approach consists of the solution of four sub-problems. Each of these sub-problems is addressed separately but with consideration of the impact of the solution of the individual sub-problems on one another. These sub-problems are outlined in the following subsections.
1.3 APPROACH

1.3.1 Signal Design

Signal design is an important topic since it determines the performance of the system. It is well known that signals can be designed for optimal performance. The design can be aimed at overcoming various difficulties imposed by transmission impairments that exist naturally in any environment. The major signal impairments for indoor systems are noise and multipath (also referred to as reverberation). Signal design should consider the purpose of the transmission of the signals. In this work, signals are primarily transmitted for localization. In addition, signals are also transmitted for communication. Having identified the goals and the obstacles, the design should consider the resources available for carrying out the task. In this regard, the bandwidth of the system is a key system parameter. The design should be tailored to fit into the available system bandwidth.

1.3.2 Acquisition and Synchronization

Acquisition and synchronization are crucial for the operation of the whole system. At startup the MD needs to identify the beacons in range. Thereafter, the MD needs to track the timing corresponding to the intervals at which the transmission from each beacon starts. This step can be facilitated by proper signal design.

1.3.3 AOA Estimation

Different methods exist for this task, however, the suitability of the existing methods to ultrasonic signals is questionable. Study of the applicability of the existing AOA methods and development of novel AOA methods is an essential element of this thesis. AOA should be estimated to the best possible accuracy to ensure good location and orientation estimation performance. Signal design plays a significant role in the attainable performance.

1.3.4 Location and Orientation Estimation

Algorithms that transform the AOA information associated with the correctly identified beacons into location and orientation information are required. Basically, this consists of solving a system of (mostly nonlinear) equations. The
algorithms should respect the low computational cost requirement, as dictated by the application of interest.

The solutions for these four problems constitute the proposed approach as depicted in Fig. 1.2. The figure also shows the operation of the proposed system as indicated by the arrows.

1.4 Thesis Contributions

This thesis presents a novel ultrasonic system for receiver’s location and orientation estimation using only AOA information. To the best of the author’s knowledge, no such system has been reported in the literature. Solutions to several problems in the design of the system have lead to a number of novel methods that have not been reported previously. These contributions are summarized in the following points:

- A novel method for phase-difference ambiguity resolution exploiting spatial diversity.
- A novel method for phase-difference ambiguity resolution exploiting frequency diversity.
- Novel AOA estimation methods using widely-spaced sensors.
- A novel system for binary data transmission using ultrasonic signals.
- A novel method for FHSS acquisition and synchronization.
- Novel methods for receiver’s location and orientation estimation using AOA.

In most cases, the applicability of these contributions is not limited to the ultrasonic location-orientation problem. Most contributions provide generic approaches that can be of use in different areas.

1.5 Structure of the Thesis

The chapters of this thesis describe various aspects of the proposed system. The following chapter is meant to furnish the reader with the relevant background required to understand the approach presented in this thesis. The
chapter also gives a brief summary of the most relevant related works. Like the thesis itself, the chapter covers a range of topics.

In Chapter 3, a FHSS signal design approach is presented. The chapter also discusses the most serious types of signal impairment. The signal design is aimed at mitigating the effect of these impairments.

Chapter 4 focuses on the acquisition and synchronization problem. The chapter discusses issues related to signal modulation and demodulation, and efficient digital receiver implementation. A preamble based acquisition and synchronization approach is proposed. Special attention is paid to the problem of preamble collision avoidance. In this regard, an analytical approach for selecting the design parameters together with the relevant analytical results is presented. The chapter also presents experimental results that demonstrate the effectiveness of the proposed acquisition and synchronization approach.

Chapter 5 and Chapter 6 address the problem of phase-difference ambiguity in two different ways. In Chapter 5, spatial diversity is exploited while in Chapter 6, frequency diversity is proposed as an alternative that provides a similar solution. Despite being based on different concepts, the two chapters share the same spirit. In these chapters, the mathematical formulation of the problem is presented and solutions are proposed. In both chapters, analytical performance is derived and used to evaluate the quality of the solutions.

In Chapter 7, the developments in Chapter 5 and Chapter 6 are utilized for AOA estimation. A number of methods are proposed and their performance is evaluated experimentally. The chapter identifies the AOA methods to be adopted in the final system.

Chapter 8 provides solutions to the AOA-based 3-D location and orientation estimation problem. First, the problem is formulated and different solutions are considered. The chapter stresses the difficulty of the problem. Two approaches are presented that are capable of handling the problem. The performance of the two proposed approaches is evaluated and compared in simulation.

In Chapter 9 the subsystems are integrated to form the final system. A sensor configuration with a minimal number of sensors is proposed for AOA estimation and the final system parameters are given. Since experimental results for acquisition and synchronization are discussed in Chapter 4, Chapter 9 focuses on experimental evaluation of location-orientation estimation performance. The results pertaining to different stages of the system are presented and dis-
cussed; and the relationships between the performance of different stages are established.

Finally, Chapter 10 closes the thesis by giving remarks on the effectiveness of the proposed location-orientation estimation approach, together with suggestions on how to improve it.
This chapter covers the background required for the developments that will be revealed in the following chapters of this thesis. In addition, related works are discussed briefly. The chapter, in general, presents an assortment of methods that belong to different areas.

2.1 Introduction

This chapter begins by examining the physical properties of acoustic waves. A propagation model is described and the most relevant sound properties are discussed in Section 2.2. Following the very low level developments in Section 2.2, Section 2.3 derives the signal model used by many array signal processing methods for angle-of-arrival (AOA) (also known as direction-of-arrival (DOA)) estimation. The derivation is followed by discussion of the general approaches that utilize the model. An alternative AOA approach based on time-delay estimation is also discussed. The section tries to identify the most suitable AOA techniques for this work. In Section 2.4, sparse arrays are introduced and their key properties are discussed. Section 2.5 discusses the phase-difference ambiguity problem pertaining to the time-delay estimation methods of interest. In Section 2.6, spread spectrum techniques are discussed and their most important properties are highlighted. Finally, in Section 2.7, a literature review on location technologies is presented. Different classes of location systems are identified and the most relevant systems are highlighted.
2.2 Acoustic Wave Propagation

The propagation of acoustic waves in air takes place in the form of a displacement of the molecules in the direction of propagation. Such waves are referred to as compressional waves. Assuming that the medium is homogeneous and lossless, the wave equation for an acoustic wave can be expressed by the formula [1]

$$\nabla^2 p(t, r) = \frac{1}{c^2} \frac{\partial^2}{\partial t^2} p(t, r) \quad (2.1)$$

where $\nabla^2$ is the Laplacian operator; $c$ is the speed of wave propagation in the medium; $p(t, r)$ is the acoustic pressure at a point whose Cartesian coordinates are given by $r = [x, y, z]^T$, and elapsed time $t$. Note that the origin for all coordinates is the point of the wave source.

Assuming an isotropic (Isotropic refers to uniform propagation in all directions) point source, an acoustic wave propagates from its source in a spherical shape. At a distance that is sufficiently far from the source (normally several times greater than the wavelength), the spherical wave model can be accurately approximated by a plane wave model. The latter situation coincides with the so-called far-field [2]. On the other hand, when the plane wave model is not sufficiently accurate, the waves should be treated as spherical waves (which is what they actually are), a situation that is commonly referred to as the near-field [2]. Normally, the near-field coincides with the so-called Fresnel zone, whereas the far-field is usually associated with what is referred to as Fraunhofer zone [1].

Assuming a monochromatic plane wave (a wave with only one frequency component), the solution of the wave equation is given by [1]

$$p(t, r) = Ae^{i(\omega t - k\cdot r)} \quad (2.2)$$

where $\omega = 2\pi f$ is the angular (or radian) frequency, and $k$ is the wavenumber vector given by

$$k = \frac{2\pi}{\lambda} [\sin(\phi) \cos(\theta), \sin(\phi) \sin(\theta), \cos(\phi)]^T \quad (2.3)$$

where $\lambda$ is the wavelength; and $\theta$ and $\phi$ are the corresponding spherical coordinates of $r$, as in Fig. 2.1. Eq. (2.2) can simply be generalized to the polychromatic case by realizing that polychromatic wave can be represented as a linear combination of monochromatic waves. The case where more than one acoustic wave is present can be handled by exploiting the superposition principle [1, 2],
2.2 ACOUSTIC WAVE PROPAGATION

Figure 2.1: Spherical coordinates \((|r|, \theta, \phi)\) and the corresponding Cartesian coordinates \((x, y, z)\).

which assumes a linear model for the medium.

From the solution of the wave equation (Eq. 2.2), it can be seen that spatial information, such as the AOA of the signal at the point of interest, is contained in the acoustic wave. Such information can be retrieved, for example, by appropriate spatial and temporal sampling of the wave.

In the above model, a constant speed of sound is assumed. However in practice, the speed of sound varies depending on different factors including air temperature [3]. Herein, this variation will be ignored on the grounds that temperature variation is limited in indoor environments. Therefore, \(c = 343 \text{ m/s}\) (corresponding to a temperature of 20° Centigrade), will be adopted as the nominal speed of sound.

2.2.1 Attenuation

Eq. (2.2) was obtained assuming a lossless medium since the interest was in the cyclic aspect of the wave. In practice, acoustic waves are attenuated over distance and the amplitude \(A\) changes accordingly. Sound is mainly attenuated due to two types of loss; geometric spreading loss [4] and absorption loss [5].

Geometric spreading is the spreading of the acoustic energy as a result of wavefront expansion. Geometric spreading is frequency independent. In the case of an omnidirectional point source, the geometric spreading loss follows the well-known inverse-square law, which dictates a loss of 6 dB for each doubling of the distance.

Unlike geometric spreading, the absorption of sound in air depends, in addition to distance, on sound frequency. It also varies with temperature, relative humidity and atmospheric pressure [5]. Absorption increases sharply as fre-
frequency increases. This is an important phenomenon since it determines, together with geometric spreading, the range of ultrasonic frequencies that are suitable for indoor positioning. For example, a 170 kHz ultrasonic wave is attenuated by up to 10 dB/m in air, whereas the maximum attenuation of a 40 kHz signal is around 1 dB/m [6]. Therefore, ultrasonic location systems are restricted to use only the frequencies below 75 kHz [7].

Geometric spreading loss and absorption loss determines the maximum range for the wave to be received with adequate energy. Although one can adjust the source wave amplitude to compensate for any loss of energy, this approach may not be practical in many cases.

2.2.2 Noise

Acoustic noise exists in both the audible and ultrasonic frequency ranges. The noise is the combination of a large diversity of unwanted or disturbing acoustic waves introduced by man-made and natural sources. This type of noise is normally additive in nature, and in many situations, it is assumed to be white. Nevertheless, depending on the degree of correlation between noise signals at distinct spatial locations, different categories of noise fields can be defined [1].

Noise is an important factor since it interferes with the desired acoustic waves. The ratio of the power contained in the desired acoustic signal to that contained in the noise field is referred to as the signal-to-noise ratio (SNR). The SNR degrades as the distance from the emission source increases. This affects the range of an ultrasonic system. Considering the achievable SNR, ultrasonic air systems are limited to a maximum range of approximately 10 m, with variations among systems [7].

2.2.3 Effect of Obstacles on Acoustic Wave Propagation

Obstacles in the path of an acoustic wave affect the propagation of the wave in one or more ways. The effect depends on the nature of the object. In general, some of the sound is reflected back from the front surface of the object and some passes into the obstacle material, where it is absorbed or transmitted through the material [4].

Herein, the interest is in the amount of reflected sound since this is transmitted back as an interfering wave. The amount of reflected or absorbed sound
differs from material to material. Normally, the reflection is parameterized by
the material reflection coefficient—a value that falls between 0 and 1, with zero
representing no reflection at all [4]. In indoor systems, normally, the walls re-
present a good reflecting surface [8]. Also, the existence of materials with high
sound absorption and reflection properties in the environment may result in
sound occlusion [9].

Other effects are diffraction and scattering, but these are not as serious as re-
lection. Reflection, diffraction and scattering are all frequency dependent phe-
nomena [4].

2.3 Angle-of Arrival Estimation

AOA estimation techniques can be grouped into two types of approaches;
those based on a general array signal processing framework, and those based
on estimating time-delay. We will simply refer to the former as array methods
and the latter as time-delay methods. Techniques belonging to these two catego-
ries will be discussed in the subsequent sections.

It has already been shown in Section 2.2 that spatial information such as the si-
gnal AOA is contained in an acoustic wave (same applies for electromagnetic
waves). In array signal processing, estimation of AOA and other signal pa-
rameters, is achieved through the fusion of spatial and temporal information
captured by sampling the wavefield with a set of judiciously placed sensors
[2]. Different sensor arrangements can be exploited to achieve this goal. The
most frequently encountered arrangements are the uniform linear array (ULA)
and uniform circular array (UCA) depicted in Fig. 2.2 and Fig. 2.3, respectively.
A linear array is only capable of resolving one AOA, e.g., the azimuthal (θ),
whereas, a nonlinear array of coplanar sensors can resolve both the azimuthal
and elevational angles (θ and φ).

In the following discussion, the far-field model will be assumed (e.g., see
Fig. 2.4). For a sensor array, the far-field/near-field is determined by the aper-
ture of the array (da), the emitter range (r) and the wavelength of the signal
of interest (λ). The aperture of the array refers to the largest distance between
any pair of sensors in the array. If r \approx 2da/λ, the source is located close to the
near-field border of the array. For r > 2da/λ, and especially for r >> 2da/λ,
the source is situated in the array far-field. As a matter of fact, sources of range
r \approx 2da/λ or r > 2da/λ, but not r >> 2da/λ, are usually regarded as being lo-
cate in the near far-field of an array, for which the spherical wave propagation model has to be utilized [10].

2.3.1 Array Methods

Herein, we will focus on the single-source data model given in Eq. (2.2). First of all, the model has been derived for the monochromatic case, which is not the case of interest in practice. The model can be extended to the polychromatic case assuming that $|\mathbf{r}| << c/B$, where $B$ is the bandwidth of the signal denoted as $s(t)$. Based on this assumption, Eq. (2.2) can be written as [2]

$$p(t, \mathbf{r}) = s(t)e^{j(\omega t - \mathbf{k}^T\mathbf{r})}. \quad (2.4)$$

As previously alluded to, this model assumes a far-field situation and can easily be extended to the multi-source case through the superposition principle.

To illustrate the idea of AOA and its relation to sensor arrays, consider the scenario depicted in Fig. 2.4 where the sensor array and the emitter are assumed to be on the same plane. In such a case, the wave impinging on the array will have an elevational angle, $\phi = \pi/2$, for all array locations on the plane. The interest is however in the azimuthal angle, $\theta$. Without loss of generality, assume that the array baseline is parallel to the coordinates x-axis (indicated by the dotted line labeled “X” in Fig. 2.4). Denote the location of each sensor by $\mathbf{r}_l = [x_l, y_l, z_l]^T, l = 1, .., L$. Assuming that the array size is small compared to the distance to the emitter (the far-field assumption), the paths of propagation of the wave to each sensor can all be assumed to be parallel, as shown in Fig. 2.4. Substituting $\phi = \pi/2$ in (2.4) and manipulating, yields

$$p(t, \mathbf{r}) = s(t)e^{j\omega t - j\frac{2\pi}{\lambda}[x_l \cos(\theta) + y_l \sin(\theta)]}. \quad (2.5)$$

The acoustic pressure in Eq. (2.5) produces a response at sensor $l$ the sampled version of which can be written in the form

$$x_l(t) = g_l(\theta)e^{-j\frac{2\pi}{\lambda}[x_l \cos(\theta) + y_l \sin(\theta)]}s(t) = a_l(\theta)s(t) \quad (2.6)$$

where $g_l(\theta)$ represents the frequency response of sensor $l$. Eq. (2.6) requires that the array aperture (the physical size in wavelengths) be much less than the inverse bandwidth, $f/B$, which is frequently referred to as the narrowband assumption [2].
2.3 ANGLE-OF-ARRIVAL ESTIMATION

Considering the contributions from all the $L$ sensors at time instant $t$, the data model is commonly written in the form

$$x(t) = a(\theta) s(t)$$  \hspace{1cm} (2.7)

where $x(t) \triangleq [x_1(t), ..., x_L(t)]$ and $a(\theta) \triangleq [a_1(\theta), ..., a_L(\theta)]$ is the steering vector of the array.

For the more general multi-source (with $M$ sources) and noisy case, the model in Eq. (2.7) can be extended to have the form

$$x(t) = A(\theta)s(t) + w(t)$$  \hspace{1cm} (2.8)

where $s(t) \triangleq [s_1(t), ..., s_M(t)]$, $A(\theta) \triangleq [a(\theta_1), ..., a(\theta_M)]$ and $w(t)$ is the vector of noise samples.

Based on the model in Eq. (2.8) (mostly for $M < L$), AOA estimation methods can be classified into two main categories; spectral-based and parametric approaches [2]. These will be discussed subsequently. The discussion is based on the excellent review in [2].

2.3.1.1 Spectral-based Methods

In spectral-based methods, one forms a spectrum-like function of the parameter of interest (AOA in this case). The locations of the highest (separated) peaks of this function are recorded as the AOA estimates. There are two main classes of spectral-based methods; beamforming techniques and subspace-based methods.

The idea of beamforming techniques is to “steer” the array in one direction at a time and measure the output power. The directions that yield maximum power are the AOA estimates. The array response is steered by forming a linear combination of the sensor outputs.

Subspace-based methods rely on the spectral decomposition of a covariance matrix to carry out the analysis. The overwhelming interest in the subspace approach is chiefly due to the introduction of the MUltiple SIgnal Classification (MUSIC) algorithm [37], which offers an improved performance compared to the other subspace methods. The performance measure of interest is usually the mean-squared error (MSE). Following the success of MUSIC, several variants of the algorithm that improve the resolution of MUSIC were introduced (e.g.,
Subspace methods normally offer better resolution than beamforming techniques.

### 2.3.1.2 Parametric Methods

Although computationally attractive, spectral-based methods may not deliver sufficient accuracy in some cases, e.g., when the scenario involves coherent signals. Parametric, or model-based, approaches are capable of circumventing such difficulties, albeit at the price of an increased computational complexity. Typically, multidimensional search is required to find the estimates, making this set of methods computationally quite complex.

Parametric approaches require an assumed statistical data model. The most frequently used model-based technique in signal processing is probably the
2.3 ANGLE-OF ARRIVAL ESTIMATION

Maximum likelihood (ML) technique [39].

Motivated by the success of the MUSIC method, parametric subspace methods have been developed (e.g., [40, 41]). These have the same statistical performance as the ML method, and are amenable for computationally efficient implementations with ULAs.

2.3.1.3 The Wideband Case

The methods discussed so far are essentially limited to processing narrowband data. A natural extension to the wideband case is to employ narrowband filtering, for example, using the Fast Fourier Transform (FFT) [42]. Then the processing can be carried out independently for each individual frequency bin. Thereafter, the results need to be combined in some appropriate way.

In general, array methods are not suitable for indoor applications such as the location system of interest [43]. The main issue is that the presence of coherent sources poses difficulty for these methods [2]. In spite of the fact that some enhanced methods have been demonstrated to cope with the presence of coherent sources (e.g., parametric methods), this still requires large array structures that would not suit small indoor devices. The presence of coherent sources in indoor systems is inevitable due to reverberation [8]. The number of coherent sources can be quite large and absolutely unpredictable. Note that the coherent sources in this case represent nuisance or unwanted signals. To accommodate the overall number of sources in a reverberant environment requires huge array structures.

2.3.2 Time-Delay Methods

An alternative approach for estimating the AOA of a signal is to estimate the time-delay (also known as the time-difference of arrival (TDOA)) of the signal as received at two or more spatially separated sensors. This approach has been used in many fields including sonar [11], radar [12], acoustic localization [13] and wireless communications [14]. These methods were basically developed to handle the single-source case (e.g., [15, 16]). Extension to the multi-source case is difficult, excluding some particular situations (e.g., [17]).

Although estimation of AOA using time-delay can be motivated through the same data model obtained in Section 2.3.1, it is always easier to reach the same result from a much simpler (geometrical) model. In the simplest case, only one
pair of spatially separated sensors is required as depicted in Fig. 2.5. Assuming a far-field scenario, the relationship between time-delay and AOA can be obtained from geometry. Consider the two sensors in Fig. 2.5. The wave from the emitter in the far-field reaches each sensor via a different path (here considering only direct-paths). These two paths are generally different in length. The difference depends on two factors; the sensor separation and the AOA of the signal. From geometry, the following relationship holds for all sensor separations \(d_{12}\) and all AOA \(\theta\):

\[
\cos(\theta) = \frac{|r_1| - |r_2|}{d_{12}}
\]  

which straightforwardly gives

\[
\cos(\theta) = \frac{c \tau_{12}}{d_{12}}
\]  

where \(\tau_{12}\) is the signal time-difference of arrival between sensor 1 and sensor 2, or simply the time-delay. This delay can therefore be estimated and utilized to obtain the required spatial parameter—AOA. Normally, the process involves an estimation of time-delay to yield an estimate of AOA, or

\[
\hat{\theta} = \arccos \left( \frac{c \hat{\tau}_{12}}{d_{12}} \right).
\]

Time-delay estimation (TDE) is dominated by cross-correlation based methods. In particular, the generalized cross-correlation (GCC) methods [18, 19] have been extensively used. GCC-based TDE consists of weighting the cross-power spectra (CPS) of the observed signals, and transforming the resulting spectra to the time domain. In the time domain, the peak location is taken as the time-delay estimate, which is an integer number of samples. Sub-sample accuracy can be obtained through computationally intensive interpolation [20]. In addition to the resolution limitation, GCC-based methods are not capable of handling the multi-source case and are therefore not useful for the location system under consideration. The interest in the multi-source case stems from the fact that 3-D location estimation requires simultaneous angle measurements for at least three base stations.

An alternative approach for TDE is to estimate time-delay directly from the phase-difference of the received signals. For example, it is possible to estimate the phase-difference between two received signals from their CPS
2.4 Sparse Arrays

Traditionally, ULAs with inter-element spacing less than or equal to $\lambda/2$ (in the monochromatic case), have been used for AOA estimation [2]. The notion of

[13, 20, 21, 22, 16, 23]. The main advantage of this is that sub-sample time-delay estimates can be obtained, without resorting to interpolation. Other advantages include optimality and the ability to apply some sort of data filtering by selecting only adequate frequency bins [21, 22]. The approach is limited to single-source scenarios in the general case. However, it has been demonstrated that by exploiting some particular signal properties, extension to multi-source scenarios is possible. For example, in [17, 44], the sparsity and orthogonality of the sources in the time-frequency domain was exploited to achieve multi-source AOA estimation. The extended methods have the added advantage that multiple sources can be resolved using as few as two sensors [44]. This makes tackling reverberation much simpler by exploiting the underlying signal model and also via signal design (see Chapter 3). This makes it possible to instrument a system using a very compact sensor configuration (see Chapter 9).

The AOA approach used in this work (see Chapter 7) is inspired from the approach in [17, 44]. The latter approach is blind, however, herein, knowledge of the transmitted signals is exploited to identify the sources and associate them with their respective AOA estimates. In addition to the advantages listed above, the approach is also suitable for the wideband case as is demonstrated in [17, 44]. However, the main disadvantage of estimating time-delay directly from phase-difference is the requirement that phase has to be unwrapped before it can be used for TDE [21, 20]. Otherwise, the estimates will be ambiguous. This phase-difference ambiguity problem will be highlighted in the Section 2.5.

Figure 2.5: Geometrical model for time-delay estimation under the far-field assumption.
the $\lambda/2$ placement is that such receiver configuration is useful in eliminating ambiguities. In array signal processing, ambiguities arise as a manifestation of the fact that the array has identical responses to more than one AOA, such that the AOA estimation problem does not have a unique solution [45]. An important property of an array configuration is the array beam-pattern, which, for a source at $\theta = \theta_0$, is defined as [46]

$$G(\theta) \triangleq \frac{|a^H(\theta) a(\theta_0)|^2}{L^2}$$

(2.12)

where $(\cdot)^H$ indicates the Hermitian transpose operation. An array configuration is generally unambiguous if the maximum of $G(\theta)$ occurs only at $\theta = \theta_0$.

In practice, sparse arrays with spacings larger than $\lambda/2$ are required mainly to improve AOA estimation resolution by increasing the array aperture, without increasing system cost [46, 47, 48, 49]. However, the price paid is the introduction of so-called grating lobes in the array beam-pattern [46]. In sparse arrays, grating lobes can be of magnitudes that are comparable to that of the array main lobe such that ambiguities may arise when noise is present [46]. Depending on the sparse array configurations, grating lobes can be as high as the main lobe, resulting in ambiguities even in the noise-free case. To avoid such situation, traditionally, sparse arrays include at least a pair of elements that are placed at a distance of $\lambda/2$ or less from each other [47, 48]. An important category of such sparse arrays are the non-uniform linear arrays (NULAs) where different neighboring pairs of elements may have different spacings [47].

As an example, Fig. 2.6 plots the beam-patterns for three different arrays for a source at $\theta_0 = 0^\circ$. Each of the three arrays has an aperture of 10 (all distances are measured in $\lambda/2$ units). The three arrays are a ULA with $\lambda/2$ spacing and 11 elements; a sparse array (Sparse-1) with three elements located at 0, 1 and 10 from one edge of the array; and another sparse array (Sparse-2) with three elements situated at 0, 4 and 10 from the reference point. It can be seen that the ULA is characterized by low side-lobes and absence of grating lobes, which ensure low susceptibility to ambiguities. In contrast, the sparse arrays exhibit grating lobes that could result in unidentifiability. In particular, for the second sparse array, which does not have any $\lambda/2$ spacing, the grating lobes are exactly as high as the main lobe. This makes the latter array ambiguous in all cases.

The effect of grating lobes in preventing consistent AOA estimation is directly related to the array methods (see Section 2.3.1) and is hard to compensate for [45].
2.5 Phase-Difference Ambiguity

Phase-difference observations are naturally confined to the interval \([-\pi, \pi]\). This imposes the requirement that the inter-receiver separation must not exceed \(\lambda/2\), where \(\lambda\) is the wavelength of the received frequency. The penalty for violating this condition is that the phase-difference estimate —and hence the time-delay— will be ambiguous due to an integer number of phase cycles (integer multiple of \(2\pi\)) uncertainty [16, 24]. The phase-difference ambiguity effect pertaining to time-delay methods represents the counterpart of the grating lobes effect in sparse arrays. Both effects are the result of the receiver placement.

As explained in Section 2.4, sparse arrays generally utilize at least a single pair of sensors with \(\lambda/2\) spacing. However, in some cases (as in this work), \(\lambda/2\) spacing may not be possible—a wider spacing of receiver may be forced by physical constraints stemming from physical placement of the receivers due to, for example, their size [25, 26]. In some cases, increasing the minimum receiver separation over the \(\lambda/2\) limit is desirable to reduce the mutual coupling between the receivers (e.g, [27]), or to improve angular resolution (angular variance is inversely proportional to the square of the receiver separation [28]).

For some multi-frequency signals, when the \(\lambda/2\) condition is partially satisfied, estimates from the frequencies satisfying the condition (the lower frequencies) can be exploited to resolve the ambiguities associated with the estimates from the other frequencies. However, in some degenerate cases, the identifiability
condition is violated for all frequencies, for example, when the signals are narrowband (e.g., [25]).

The phase-difference ambiguity problem has received much attention in the GPS community. Various solutions have been proposed in the context of vehicle attitude determination (e.g., [24, 29, 30, 31]). In most cases, more than one received frequency is incorporated in order to resolve the ambiguity using data recorded over a long period of time. In some cases, receiver movement is required (e.g., [30]). In all cases, the solutions require computationally complex iterative search without guaranteed convergence.

In [25], the problem of phase-difference ambiguity resolution for a single-frequency signal is discussed. The proposed solution exploits the spatial diversity provided by a third (auxiliary) collinear receiver. The approach restricts the inter-receiver spacing in such a way that the triplet is only suitable for a single pre-defined frequency, which make it unusable for other frequencies, for example, when Doppler shifts of the operational frequency occur, or when the signal of interest comprises multiple frequencies.

Other methods are based on phase unwrapping techniques (e.g., [16, 33, 34, 35, 36]). In general, these methods rely on the fact that the signal spectrum is continuous, and that it starts at zero frequency (i.e., the wideband signal case) or some wrapping-free low frequency that can be used as the starting point of the unwrapping process. The phase is unwrapped progressively starting from the lowest frequency, leaving the success of the whole process dependent on the success of the unwrapping at the low frequencies. These method will not work when the lowest frequency is subject to phase wrapping. An example for this is a bandpass signal, received by receivers whose separation exceeds $\lambda_{\text{max}}/2$, half of the maximum wavelength. Conventional phase unwrapping methods cannot be used in this case, which is the case of interest for this work. This makes addressing the phase-difference ambiguity problem in a way that is capable of handling the bandpass signal case of central importance for this work.

### 2.6 Spread Spectrum

Spread spectrum is a signal transmission technique in which the transmitted data sequence occupies a bandwidth in excess of the minimum bandwidth necessary for transmitting the sequence. The spectrum spreading is achieved
2.6 SPREAD SPECTRUM

through the use of a code that is independent of the data sequence. The receiver, which operates in synchronism with the transmitter, uses the same code to despread the received signal and recover the transmitted data [50]. The principal benefits of using spread spectrum are [50, 51, 52]

- Anti-jamming and low probability of intercept.
- Providing multiple access, e.g., through what is known as code division multiple access (CDMA).
- Wideband signals can provide precise time-of-flight measurements for range and location determination.
- Spread-spectrum signals are resistant to multipath interference.

The use of spread-spectrum techniques in this work is motivated by the last three features. The multiple-access feature is achieved by employing proper modulation formats (spreading codes), so that we are able to achieve near orthogonality of waveforms despite the fact that many users are sharing the same spectrum. This orthogonality, if strict, would allow multiple users to coexist in a given frequency range without mutual interference. The advantage in range and location estimation, and the multipath resistance feature follow directly from the narrow autocorrelation responses associated with wideband signals [51].

There are two types of spread spectrum technique; direct-sequence spread spectrum (DSSS) and frequency-hopping spread spectrum (FHSS). For this work, FHSS is more suitable. In particular, its frequency domain properties that allows one to track individual transmissions is probably the most attractive feature for this work. This feature helps in non-blind AOA estimation in a simple manner as will be demonstrated later in this thesis. In addition, the results obtained by [57, 58] show better performance for FHSS versus DSSS in indoor environment, which is another driving factor. Moreover, FHSS is free from the near-far problem inherent in DSSS systems. The latter problem requires strict power management that can be an issue in applications such as location systems. In the following subsections, an overview of DSSS and FHSS is given.

2.6.1 Direct Sequence Spread Spectrum

DSSS signals are generated by the direct mixing of the data with a spreading waveform before the final carrier modulation. The spreading waveform is
constructed according to a spreading sequence that is known by both the transmitter and the receiver. To despread the received signal, the receiver multiplies it by the spreading waveform to return the data sequence to its original bandwidth [51].

### 2.6.2 Frequency-Hopping Spread Spectrum

In FHSS, digital modulation is performed in the normal way except that the carrier frequency is periodically changed (or *hopped*) [51]. The hops are taken from a set of frequencies denoted as the *hopping set*. The actual sequence of frequencies taken from the hopping set is known as the *hopping pattern* (HP), and the rate at which the carrier changes is called the *hopping rate*. Usually, the hopping pattern is determined based on a pseudorandom code [51, 52]. However, hopping patterns with some desirable properties are also used (e.g., see [59, 60]). Fig. 2.7 shows an example of a hopping pattern for a system that uses a hop duration $T_h$; and the bandwidths corresponding to a hop and that of the whole system are $B_h$ and $B$, respectively.

Fig. 2.8 shows a block diagram of a FHSS signal generation system. The code generator generates a pseudorandom code at a rate that is equal to the desired hopping rate. The frequency synthesizer takes this code as an input and uses it to generate the corresponding carrier frequency. Therefore, the input data sequence $x_n$ is modulated using a different carrier every time a new code is generated, resulting in an output signal, $s(t)$, that has much larger bandwidth compared to the data sequence. At the receiver side, a code generator that is synchronized to the transmitter’s code generator allows the right carrier frequencies to be generated at the receiver for the purpose of demodulating the received signal. Note that a pseudorandom code generator is not needed when the transmitter uses a predetermined HP. The latter approach is more appealing for a location system since it allows for the identification of beacons by their HPs.

Based on the hopping rate, two categories of FHSS can be recognized [52]:

1. **Fast FHSS**: The hopping rate exceeds the data symbol rate, which means that more than one hop occurs in a single symbol duration.

2. **Slow FHSS**: The hopping rate is at most equal to the data rate.

Slow FHSS is usually preferable because the transmitted waveform is much
more spectrally compact and the overhead cost of the switching time is reduced [52]. However, fast FHSS is useful in mitigating harmful effects caused by frequency-selective fading [51].

In this work, slow FHSS with equal data and hopping rates is used.

## 2.7 Location Techniques

Location techniques can be classified into different categories based on different aspects. In terms of the physical medium, different technologies have been used. These include RF (e.g., [61, 62]), infra-red (e.g., [63, 64]) and ultrasonic (e.g., [65, 66, 67, 68, 69]). Each technology has its advantages and drawbacks. Ultrasonic technology has been demonstrated to have good accuracy for indoor systems owing to the low speed of sound. On the other hand, short range and non-line-of-sight due to occlusion are the major downsides [71].

Another classification is based on the principle used for location estimation. In this regard, TOF (e.g., [65, 66, 67, 68, 69]), difference of time-of-flight (DTOF) (e.g.,
AOA (e.g., [78, 79, 80, 81, 82]) and received signal strength (RSS) (e.g., [83, 84, 85]) are the most notable principles. TOF has been used in many systems using different transmission media. The approach consists of estimating the time taken by the signal to travel the distance from the transmitter to the receiver, and exploiting knowledge of the speed of propagation of the signal to determine the distance. At least three different transmitter-receiver distances are determined and a trilateration [72] or multilateration [69] algorithm is used to compute 3-D location.

In DTOF systems, the difference in the time taken by the signal to reach a number of fixed receivers is utilized for localization. There are several formulations that can be used to obtain the transmitter’s location.

Similar to the TOF approach, AOA-based location estimation requires multiple AOA measurements to fixed beacons. The technique has been used in a number of systems in different contexts (see [78, 79, 80, 81, 82]).

The level of the received signal power can be used as an indicator of how far the receiver is from the transmitter. In RSS based systems, location is estimated based on the RSS from multiple nodes. A set of algorithms are available for this task. RSS has been widely used as a location technique in Wireless sensor network systems [83].

Another differentiator between location systems is whether the mobile device transmits or receives. We refer to those systems in which the mobile device (MD) receives signals as privacy-oriented systems (e.g., [65, 68, 69]), whereas those in which the MD transmits are referred to as infrastructure-based systems (e.g., [66, 67]). In privacy-oriented systems, the MD is capable of determining its own location from the signals it receives. The MD may however choose to send the information it derives from the received signals to a central processor where location is calculated. In infrastructure-based systems, the MD does not control knowledge of its location information since it is the transmitting side. A central processor is responsible for calculating the MD’s location based on the signals received by fixed beacons. The information may be sent back to the MD so that it knows its own location.

The accuracy of current location systems varies from a few millimeters (e.g., [89]) to several meters (e.g., [90]). Different applications have different accuracy and budget requirements.

Herein, the interest is in a privacy-oriented system that uses ultrasonic signals and AOA for determining location (and also orientation). The rationale for
such a system was explained in Chapter 1. The proposed system is significantly different from previous systems. In particular, a location estimation method is required that is capable of determining the location (and orientation) of a device with unknown orientation using only AOA information. The most relevant systems are discussed in the following subsection.

2.7.1 Related systems

In GPS based vehicle positioning and attitude (i.e., orientation) determination systems, the receiver’s location is determined exploiting signals originating from transmitters with known locations (satellites in this case). In these systems, location is determined using TOF and trilateration. Formulations for determining the receiver orientation have also been derived (e.g., [29, 30, 32]). The orientation determination systems use phase-differences observed by at least two receivers that are situated on the vehicle.

In [58], both TOF and AOA are combined for more accurate location applied to a receiver device with an unknown orientation. The system also uses ultrasonic FHSS signals. Knowledge of both TOF and AOA is exploited for determining the receiver orientation.

The cricket system [65] uses ultrasonic TOF for determining location in a privacy-oriented context. An attempt to extend the system for orientation determination is described in [25]. But no explicit formulae for orientation has been derived.

Like this work, the approach in [86, 87, 88] addresses the problem of RF-free ultrasonic location. These systems exploit a Doppler-like effect that is caused by change in the reception pattern due to the receiver movement. A Kalman or particle filter was used to model the receiver position according to the change of the received pattern. The approach is capable of estimating the location and the speed of a MD, however, no orientation information can be obtained. The approach is also not suitable for multi-room scenarios and is generally difficult to scale to a large number of beacons.

2.8 Summary

This chapter presented background material on a number of topics and identified related work. The physical properties of acoustic waves that are relevant
to this work were discussed. Based on these wave propagation properties, the model required by array signal processing methods to estimate AOA was derived. A brief review on AOA methods that utilize the derived model was given. AOA estimation through time-delay estimation was discussed. An introduction to sparse arrays and grating lobes theory was given followed by discussion of the phase-difference ambiguity problem attached to phase-based time-delay estimation methods. Another topic was spread spectrum techniques and their properties. At the end of the chapter, an overview of location techniques was presented highlighting the most relevant systems.
This chapter discusses the problem of designing a frequency-hopping spread spectrum (FHSS) signaling scheme for indoor positioning using angle-of-arrival (AOA). A generic approach is presented that aims at mitigating the effect of signal transmission impairments that commonly occur in the indoor environment. Various issues that affect the design process are discussed and their effects are highlighted.

### 3.1 Introduction

Signal design is an important issue in the design of a communication or positioning system. It is well known that the performance of such systems can be optimized through the signal design stage [54]. The need for good signal design is dictated by the signal impairments that naturally exist in the environment.

Different impairments occur to the signal while propagating in a transmission medium. Two important types of such impairments are noise and interference. FHSS does not provide protection against white noise [52]. However, it has good properties when it comes to interference, more specifically, narrowband interference [52]. The focus of this chapter is on designing FHSS signals in a way that reduces the effect of interference.

Interference can be attributed to two main types of sources; coherent and non-coherent. Non-coherent sources are signal sources that exist in the environment and that are independent of the source of the signal of interest. Coherent signal sources are those generated as images (delayed and scaled versions) of the original signal source (normally due to the existence of signal reflecting objects in the environment). Typically, coherent sources are referred to as reverberation [53] (in the acoustics domain) or multipath interference [54] (in the
In this work, signals are transmitted for two purposes. The main purpose is to utilize the signals received at multiple sensors for estimating the AOAs of these signals—and to use that to determine the location and orientation of the receiver. The second purpose is to communicate information such as the transmitter’s ID. It is clear that for localization purposes, only the signal direct path (from the transmitter to the receiver) should be engaged in the AOA estimation process. Otherwise, erroneous results may be obtained. For communicating information, however, multipath signals can also be utilized. Nominally, when it is possible to resolve signals received from different paths, diversity gain can be achieved by combining these components [55]. Herein, we will focus on designing signaling for the positioning purpose where only the signal direct-paths should be utilized for obtaining estimates of the required spatial parameters. In this case, no modulation is applied to the carrier signals, as will be assumed in the following discussions. A more specialized treatment of signal design for information communication will be presented in Chapter 4.

Non-coherent interference can be classified into two types; (narrowband) interference from an external source, and (broadband) multiple-access interference. In the case of external-source narrowband interference, FHSS naturally provides resistance to this type of impairment [52]. The effect of a narrowband interfering source is limited to the frequency band occupied by that source. The rest of the frequency bands will remain unaffected and can hence be used to extract any information the signal may carry. Therefore, it can be said that the effect of narrowband non-coherent sources on a FHSS system is limited. Broadband interference exists naturally in FHSS systems where FHSS is employed as a multiple-access technique. FHSS is primarily designed to handle this kind of interference by using orthogonal sets of sequences so-called hopping patterns (HPs). More discussion on this topic will follow later in this chapter.

Another type of interference that affects FHSS transmission is multipath interference. Multipath interference can completely destroy the received signal quality. The most serious known effect is the so-called frequency-selective fading, where the addition of multiple signals (i.e, multipath and direct-path) with different phases causes the signal amplitude at some frequencies to diminish [52]. This results in poor signal quality that renders any processing to be inadequate. However, even in less severe situations, the collision of the direct-path signal with a multipath signal, may preclude accurate estimation
of signal parameters such as AOA.

The design philosophy adopted herein is to reduce the effect of the collision of the signal with its own reflections by forcing a less destructive kind of collision. This is based on the notion that orthogonal signals add non-destructively. In other words, if two orthogonal signals are added, their individual properties will still be accessible. Therefore, any processing can be performed in the same way as when the two signals are separate. For this work, the most important property to be maintained while adding coherent sources, is phase. Maintaining phase guarantees that AOA estimation for an individual source is not affected by collision. Additionally, for this work, phase will be the signal property of choice to carry information transmitted for the purposes of acquisition and synchronization, as will be revealed in Chapter 4.

In Section 3.2, the concept of orthogonality in FHSS design is introduced. In Section 3.3 the use of FHSS as a multiple-access technique is discussed and ways of coping with multiple-access interference are briefly discussed. Section 3.4 discusses the effect of multipath interference and its effect on design. In Section 3.5, a design methodology that considers various system requirements is presented. Section 3.6 summarizes the chapter.

### 3.2 The Orthogonality of Signals

Two signals $s_1(t)$, and $s_2(t)$, where $t$ is time, are said to be *orthogonal* if and only if they satisfy [54]

$$\int_{-\infty}^{+\infty} s_1(t)s_2(t) = 0. \quad (3.1)$$

If $s_1(t)$ and $s_2(t)$ are *time-limited* signals, the integration in Eq. (3.1) will be carried over the time interval where the signals are defined. If the time variable, $t$, is discrete, the same rule in Eq. (3.1) applies by converting the integration into a summation over the discrete values of $t$. Orthogonal signals have the property that they can be resolved even when they are mixed, e.g., by exploiting their correlation properties [54].

From the properties of the Fourier transform [42], it is straightforward to prove that any two signals that occupy two non-overlapping frequency bands, are orthogonal and satisfy Eq. (3.1). It is this form of orthogonality that FHSS multiple-access is based on. Simply, by employing a set of carriers, so-called *hops*, at different frequencies, signals with good orthogonality properties can
be designed that can share the same channel bandwidth. Since these hops are time-limited, they occupy an infinite bandwidth in the frequency domain. However, sufficient orthogonality is guaranteed if the following condition is satisfied [54]:

$$\Delta f \geq \frac{1}{T_h}$$  \hspace{1cm} (3.2)

where $\Delta f$ is the separation between the center frequencies of any two hops and $T_h$ is the hop duration, assuming that all hops have the same duration.

Such orthogonality of signals makes it possible to estimate the parameters of different co-existing signals, either by exploiting the correlation properties of these signals (as manifested in Eq. (3.1)), or by obtaining the information directly from a frequency-domain representation of the signals where the signal are sufficiently disjoint.

### 3.3 FHSS and Multiple Access

FHSS can be used as a multiple-access technique to allow multiple users to share the same transmission channel [54]. Each user (beacon in our case) is provided with a unique hopping sequence or hopping pattern (HP) that differs from those of the other users. Normally, pseudo-random sequences are used as HPs [54]. However, several methods exists for designing HPs with good properties. The most desirable feature of a set of HPs is good orthogonality that is reflected by the sequence auto- and cross-ambiguities [91, 94]. Some good sequences are, for example, those based on Reed-Solomon codes [93, 92, 60], extended hyperbolic congruential codes [91] and Costas arrays [59, 94]. When a certain HP is used by a certain transmitter, that transmitter will periodically transmit the same HP. A course of a complete HP is referred to as a hopping cycle. Herein, it is assumed that the hopping cycle consists of unique hops; no hop is repeated more than once in a hopping cycle.

In the case of multiple access, frequency collisions (or hits) may occur in some situations. A frequency collision occurs when the signals received from more than one user happen to have the same hopping frequency at the same time, which violates the orthogonality principle of the system. HPs can be designed with a certain maximum number of hits in a hopping cycle (e.g., see [59, 60]), for all the shifts of the sequence. Ideally, this should be equal to one.

In addition to the interference between the direct-path signals, multipath from
other transmitters is another form of non-coherent interference. This may also cause frequency collisions and the total number of hits will increase as the number of transmitters increases. These effects resulting from multiple-access are unavoidable. However, the frequency diversity provided by FHSS can be sufficient to prevent the latter effects from overwhelming the system. It is obvious that employing long HPs together with short hop duration can help in decreasing collision rate. Also, employing HPs with good properties can help in reducing the total number of hits [59, 60]).

3.4 Multipath Interference

Here the focus is on multipath arising from the signal of interest (coherent multipath). Herein, we distinguish between two types of multipath:

- Early multipath: Multipath signals that are received within one hopping cycle duration from the reception of the direct-path signal
- Late multipath: Multipath that are received later than one hopping cycle duration from the reception of the direct-path signal.

3.4.1 Early Multipath: The Concept of Autocontamination

When early multipath is received, a hop on the multipath may be received sufficiently early such that it (partially) overlaps with its corresponding direct-path. This situation is extremely undesirable and, herein, is referred to as autocontamination. This has the effect that the phase information of the direct-path signal is likely to be corrupt.

Autocontamination occurs when the difference between the signal time of flight of the multipath signal and that of the direct-path signal, is less than a hop duration. To avoid autocontamination, the following condition must be satisfied

\[ T_h < \frac{R_{mp} - R_{dp}}{c} \]  (3.3)

where \( R_{dp} \) and \( R_{mp} \) are the lengths of, respectively, the direct signal path and the indirect signal path (multipath); and \( c \) is the speed of propagation. Eq. (3.3) has to be satisfied by all the paths of the signal.

The avoidance of autocontamination ensures that the direct-path signal properties such as phase and amplitude will not be affected by multipath. This
is because a received direct-path hop will always be added with a hop of different frequency, therefore preventing same frequency collision. The overlap of hops of different frequencies will not cause problems since these hops are orthogonal by design.

When Eq. (3.3) is not satisfied, the received signal will be autocontaminated. The level of autocontamination will however depend on the degree of overlap between same hops and the number of paths causing such overlap. If $R_{mp} - R_{dp} \in (0, T_h]$ for a certain multipath, the fraction of a hop that is autocontaminated due to that multipath will be given by $cT_h - (R_{mp} - R_{dp}) / cT_h$. If this fraction is sufficiently small, autocontamination will have no serious effect on AOA estimation, for example. However, the best case is always when Eq. (3.3) is perfectly satisfied.

### 3.4.1.1 Practical Considerations

To avoid signal autocontamination, one needs to make sure that no reflectors are located near the transmitter and/or the receiver at distances that would result in signal paths that do not satisfy Eq. (3.3). Since the main reflectors in an indoor environment are walls, it is inevitable that failure will occur in some regions of the space. To mitigate this effect, the signaling can be designed so that the failure zone close to the walls is minimized. The size of the failure zone near the walls (or any other reflectors) is determined by the hop duration, as can be inferred from Eq. (3.3). Assuming that the transmitters and the receiver are omnidirectional, or at least not very directional, a sufficient condition for avoiding autocontamination is that the transmitter and the receiver be located at distances larger than $cT_h / 2$ from the walls (or any other reflecting object). In other words, the hop duration, $T_h$, serves as a design parameter that determines the clearance distance, $d_{\text{clear}} = cT_h / 2$, that determines where (in space) the particular design should work. The distance $d_{\text{clear}}$ represents the worst case limit for the design. The design may still work at closer distances from an object depending on the reflection properties of that object.

Fig. 3.1 depicts a scenario where multipath is generated as reflections from walls. Multipath 1 is due to the transmitter’s location being close to the wall, while multipath 2 is due to the receiver’s location. Note that the occurrence of multipath 1 requires that the transmitter be omnidirectional, or at least not very directional such that it emits sufficient signal power towards the wall. Similarly, for multipath 2 to be received, receiver directivity that allows for re-
ception from the direction of the wall is required. The problem of multipath is greatly reduced when the transmitters and/or the receivers have strong directivity, and is completely eliminated when both of the transmission and reception sides are unidirectional (and are pointed at each other).

### 3.4.2 Late Multipath

In the case of late multipath, no autocontamination occurs. However, the multipath of a hop may overlap (or collide) with a hop of the same frequency in the next, or a later, hopping cycle. This causes the same effect of autocontamination, though not as severe since the multipath signals in this case will be more attenuated. To ensure no serious effects, it is desirable to design the system such that all multipath signals are sufficiently attenuated within the period of a hopping cycle. This can be guaranteed when the hopping cycle duration is greater than the multipath delay spread, which can be approximated by the (room) reverberation time. Therefore, it can be concluded that to avoid the effect of late multipath, a hopping cycle duration that is greater than the room reverberation time should be adopted.

### 3.5 Design for Limited Bandwidth

#### 3.5.1 Requirements

From the above discussions, it is clear that to reduce the effect of multipath, the following requirements need to be satisfied:

- Increasing the total number of hops, $N_h$: required for mitigating the effect of multiple-access interference.
• Reducing the hop duration, $T_h$: required for mitigating the effect of early multipath and multiple-access interference.

• Increasing the hopping cycle duration, $T_c = N_h T_h$: required for mitigating the effect of late multipath.

In theory, there is no limit for the number of hops required to improve performance against multiple-access interference—the larger $N_h$, the better. Therefore, in the following discussion, we will only consider obtaining the design value of $N_h$ that reduces the effect of (late) multipath interference. The last two requirements listed above are conflicting since reducing the hop duration will inevitably decrease the hopping cycle duration. Therefore, one may have to consider increasing the hopping cycle duration by only increasing the number of hops, while keeping the hop duration sufficiently small for the second requirement to be satisfied. This can unfortunately be prohibitive in practice, solely due to the limitation of the available bandwidth. The total bandwidth occupied by a FHSS signal is the sum of the bandwidths occupied by the individual hops. The bandwidth of each hop should satisfy Eq. (3.2) to obtain sufficient signal orthogonality.

The exact value of the bandwidth devoted for the transmission of each hop can be set to the lowest possible value, $1/T_h$, to conserve bandwidth. Nevertheless, practical considerations may require an increase in this value. Such considerations include frequency offsets due to Doppler shifts induced by the movement of the receiver, and frequency errors at the transmitter and the receiver. The latter effect is less significant so only the former effect will be considered. Doppler frequency shifts have no serious effect on signal orthogonality since the Doppler effect always acts consistently on different frequencies (either increasing or decreasing frequency). However, to facilitate hop detection in the frequency domain, redundancy is to be introduced. More specifically, for the system to tolerate Doppler shifts, the bandwidth devoted for a single hop transmission (which is equivalent to the frequency separation), should be increased. Here we propose increasing the bandwidth for each hop to

$$\Delta f = \frac{1}{T_h} + 2f_g$$

where $f_g$ is a guard frequency meant to help in accommodating Doppler shifts from $-f_g$ Hz to $+f_g$ Hz. This guard frequency is recommended to be at least equal to the maximum expected Doppler shift the system is expected to undergo at all frequencies. Fig. 3.2 illustrates the rationale for introducing redun-
dancy as per Eq. (3.4). It is clear from the figure that the introduction of such frequency gaps ensures that the frequency contents of each hop will always remain within a distinct portion of the bandwidth. Therefore, the presence of any hopping frequency can always be detected based on the power content of the corresponding frequency band. The redundancy can also be useful in improving signal orthogonality.

The total bandwidth is the sum of the bandwidths of the individual channels and can hence be written as

$$B = N_h \Delta f = N_h \left( \frac{1}{T_h} + 2f_g \right)$$

(3.5)

From Eq. (3.5), it is clear that for a fixed bandwidth and guard frequency, the number of hopping frequencies is determined by the hop duration and vice versa. Therefore, the designer has to choose between obtaining a target number of hops, or a hop duration that optimizes performance in the sense revealed in Section 3.4.1. To maintain both features, the choice may only be to increase the system bandwidth.

### 3.5.2 Design Trade-offs

Let us assume that an amount of bandwidth, $B$, is available for FHSS transmission. We need to design the signals for a satisfactory clearance distance and a hopping cycle that is at least equal to the room reverberation time, $RT$.

The problem can be formulated as an optimization problem to find the values of $N_h$ and $T_h$ that optimizes performance against reverberation and noise. The design is constrained by the available bandwidth, the value of the reverberation time to be accommodated and the amount of Doppler shift to be accommodated. The value of the optimal pair, $(N_h, T_h)$ depends on many factors, including the reverberation model of the space of interest. The problem of finding such an optimal pair can be mathematically complex, however, suboptimal values of $N_h$ and $T_h$ can be determined in a simple way as will be explained. Given the values of $B$ and $f_g$, pairs of $N_h$ and $T_h$ can be directly determined from Eq. (3.5). The pair that most closely satisfies the design goals is selected as the design of interest. A good measure of the quality of a design is the combination of the clearance distance; and the difference between the design hopping cycle duration and the reverberation time, denoted as $T_{silence}$.

Table 3.1 presents results for such a design approach for $B = 15$ kHz, $f_g =$
CHAPTER 3: SIGNAL DESIGN

Figure 3.2: Hop separation: $f_1$, $f_2$ and $f_3$ are hopping frequencies.

Table 3.1: Examples of FHSS design for $B = 15$ kHz, $f_s = 200$ Hz and $RT = 125$ ms.

<table>
<thead>
<tr>
<th>$T_h$ (ms)</th>
<th>$N_h$</th>
<th>$d_{\text{clear}}$ (m)</th>
<th>$T_{\text{silence}}$</th>
<th>$N_{h0}$</th>
<th>$B_0$ (kHz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>0.17</td>
<td>0.92</td>
<td>125</td>
<td>175</td>
</tr>
<tr>
<td>2</td>
<td>16</td>
<td>0.35</td>
<td>0.74</td>
<td>63</td>
<td>56</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>0.50</td>
<td>0.52</td>
<td>42</td>
<td>31</td>
</tr>
<tr>
<td>4</td>
<td>23</td>
<td>0.68</td>
<td>0.26</td>
<td>31</td>
<td>20</td>
</tr>
<tr>
<td>5</td>
<td>25</td>
<td>0.86</td>
<td>0.00</td>
<td>25</td>
<td>15</td>
</tr>
</tbody>
</table>

200 Hz (corresponding to a speed of approximately 1.7 m/s at a frequency of 40 kHz) and $RT = 125$ ms. The first pair of columns lists values of $T_h$ and $N_h$ that satisfy Eq. (3.5). The second pair of columns lists the corresponding values of the clearance distance and $T_{\text{silence}}$, given as a fraction from the design reverberation time. The third column lists the total number of hops, $N_{h0}$, required to accommodate the design reverberation time; and the corresponding total required bandwidth.

It can be seen that short hop durations give small clearance distances, but only accommodates a fraction of the design reverberation time. To accommodate the whole reverberation time, a substantial extra bandwidth is required. Increasing the hop duration, on the other hand, reduces the demand on bandwidth but increases the clearance distance. The last row of Table 3.1 represents the only case where the design exactly satisfies the reverberation time requirement. This should be the design of choice if the corresponding clearance distance is satisfactory. If not, the designer may have to choose one of the other designs. The designer may choose the design with the maximum acceptable clearance distance and compensate for the limitation in the number of hops by inserting a sufficiently long silence period at the end of the hopping cycle such that reverberation will die out before the start of the next cycle. Alternatively, the hopping cycle can be interleaved with smaller silence periods that are inserted between hops to achieve the same goal. This technique can be efficient, but it may not be capable of delivering the desired system update rate.
3.6 Summary

This chapter discussed various issues concerning the design of a FHSS signaling scheme for indoor positioning using AOA. The effects of different types of interference were considered, namely, the effect of multiple-access and multipath interference. The design criteria were focused on the effects of early and late multipath. Finally, an example that illustrated the design process was presented and suggestions to overcome certain design limitations were given.
In this chapter, the problem of acquiring the hopping pattern of a FHSS signal in the context of the location system is addressed. Another issue that is discussed is that of synchronizing the receiver to the received signals. The acquisition and synchronization problem is of central importance to the location system. Acquisition failure means that the system will fail to perform the main task it is designed for—localization. At the start of this chapter, the problem is defined and a generic discussion of possible solutions is given. Next, proposed solutions are described in detail. The chapter presents both analytical and experimental results to demonstrate the performance of the proposed solutions.

4.1 Introduction

The scenario under consideration can be described as follows. A mobile device (MD) (will also be referred to as receiver) is switched on and starts receiving ultrasonic FHSS signals from all the transmitters (or beacons) in range. The MD does not know the hopping patterns (HPs) of the transmitters that are in range. However, it knows the set of all HPs, for example, in a building. The set of all HPs in use together with the corresponding transmitters’ IDs are stored in the MD memory. It is also assumed that the MD knows information such as the hopping frequencies, hopping rate and the modulation scheme being used.

When the receiver starts up, it needs to find out the HPs in the transmissions being received and associate these HPs with their relevant transmitters’ IDs. Furthermore, the receiver needs to accurately acquire the timing information within each HP. By timing information we mean the start and end of each hop.
Typically, this process is performed in a two-step manner. In the first step, a rough estimate of the hop timing is obtained. In the second stage, this rough estimate is refined to a smaller amount of error. Thus, the problem of FHSS acquisition and synchronization can be summarized by the following three steps:

1. Step 1: determining the hopping patterns of the beacons in range (HP acquisition).
2. Step 2: determining the timing of each of the acquired HPs to within a fraction of a hop duration (coarse synchronization).
3. Step 3: refining the timing of each of the acquired HPs to the smallest required accuracy (fine synchronization).

In the spread spectrum literature, step 2 is normally referred to as acquisition and step 3 as tracking.

It should be noted here that, in most of the existing communications literature, synchronization is required to align a received HP with a version of it that is locally generated at the receiver. The goal is to facilitate the dehopping of the signal [52]. On the other hand, in this work, no explicit signal dehopping is required. In fact, synchronization is required to identify, within a particular data window, which hop frequency belongs to which transmitter, so that parameters (e.g., time-delay) estimated from the different frequencies can be associated with the correct transmitters. Additionally, acquiring timing information of a particular transmission can be utilized to precisely align the Discrete Fourier Transform (DFT) windows with the hops in order to maximize the fraction of power of a hop frequency in a window, which can improve the results of signal parameter estimation. Generally speaking, the synchronization accuracy required for the AOA based location system under consideration is far less strict than that is required for a communications system. In general, coarse synchronization can be sufficient.

Acquiring the received HPs is essentially a frequency detection problem [105] over multiple frequencies (from a specified frequency set). The (timely) ordered detection of these frequencies corresponds to the detection of a HP. Detecting more than one HP in a signal reception can be viewed as a multiuser version of the single HP detection problem and is hence more complicated.

For the scenario under consideration, acquiring HPs can be achieved by using blind and semi-blind methods [103, 104, 105, 106, 107, 108] that are capable
of estimating hop frequencies and hop-timings, and hence hopping patterns. These methods are designed for noncooperative scenarios where interception of FHSS signals, whose features are completely or partially unknown, is required. These methods are generally characterized by their high complexity and susceptibility to error and will hence not be considered herein. Instead, simpler and more robust cooperative approaches will be investigated.

A simpler approach for acquiring HPs is to encode the data required for determining the HP in the form of a message transmitted via a dedicated channel, as is done in [109, 110, 111, 112]. Normally, such data is transmitted in a preamble of the data frame. The preamble also serves as a marker that indicates the start of the hopping cycle and hence it facilitates synchronization. In [113, 114], for example, preambles are used solely for synchronization, i.e., HPs are assumed to be known a priori.

On decoding a preamble, the receiver can use the data contained therein as input to a HP generator, or as the entry to a lookup table that contains the set of all possible HPs. The decoding process can be aided by careful selection of the channels used for preamble transmission and the signal modulation.

Once the HP from a particular transmitter has been acquired, Step 2, coarse synchronization, amounts to listening to the channel and detecting the reception of a complete hopping cycle. For systems that use preambles, both step 2 and step 3 may be achieved as part of the preamble decoding process. For example, detection of the arrival of a preamble can be used as a form of coarse synchronization. Furthermore, preamble decoding may involve synchronization that is required to coherently demodulate the received data. In this case, assuming that the timing within each data frame is fixed and known by the receiver, the receiver can achieve fairly accurate synchronization by locking to the preamble section of the frame alone.

If preambles are not used, or if using preambles for synchronization is error prone, classical approaches for coarse synchronization (see [52]) can be used. An example of a case where relying on preambles for synchronization can be problematic, is when the preamble consists of the same frequency channel(s) for all users in a multiuser system. In such a case, collision may prevent correct reception of the preamble and therefore single-user tracking of the preamble will not be possible. Another example is when the assumption that the timing of the data frame relative to the preamble is known a priori is not accurate, e.g., due to Doppler effect, etc. Therefore, it is more convenient to detect the preamble of a particular user from the first collision-free reception after
CHAPTER 4: ACQUISITION AND SYNCHRONIZATION

Figure 4.1: Matched-filter coarse synchronization system.

the receiver starts up. Thereafter, the receiver may repetitively need to re-synchronize to the HP that has been acquired utilizing the preamble. Synchronization independent of the HP acquisition stage, may be required whether preambles are being used or not.

Two popular approaches for coarse synchronization (when HPs are known a priori) are the serial-search and the matched-filter approaches [52]. Fig. 4.1 shows an example of a programmable matched-filter coarse synchronization system. It is assumed that a single frequency channel is used during each hop interval. Fig. 4.2 illustrates a serial-search course synchronization system. A matched-filter coarse synchronization system performs a parallel search for the desired HP, while a serial-search system looks for the whole (or at least a
subset of the) HP structure at one time [52]. It is therefore not surprising that matched-filter coarse synchronization offers faster synchronization as compared to serial search synchronization. Full details of matched-filter and serial-search synchronization are discussed in the communications literature (e.g., [52, 54]).

Fine synchronization aims at reducing the residual misalignment after coarse synchronization. Delay-locked and tau-dither loops can be used for this purpose [52]. However, the prominent form of fine synchronization in FHSS is provided by the early-late-gate tracking loop [52].

The next section will discuss the issue of all-digital receiver implementation. In Section 4.3, the discussion will focus on issues related to signal modulation, demodulation and detection; and the chosen modulation scheme is discussed in detail. Section 4.4 discusses the proposed methods for acquisition and synchronization. Section 4.5 is concerned with the practical implementation and experimental testing of the proposed methods; experimental results are given at the end of the section. Finally, Section 4.6 summarizes this chapter and highlights the most important future work for this aspect of the system.

### 4.2 All-digital Receiver Implementation

In this work, an all-digital receiver implementation is desirable. Most of the techniques for acquisition and synchronization found in the literature, including the ones mentioned in the previous section use analogue circuitry. The
interest in this work is to feature all the system functions on a digital integrated

circuit. This is required for low-power low-cost commercial devices.

In this work, a preamble based acquisition and synchronization approach is

adopted. Preambles are used both for transmitting identification information

and for aiding synchronization. The main source of difficulty for a digital im-

plementation of a preamble transmission system is the signal demodulation

process. The most widely used modulation schemes with FHSS are frequency

shift keying (FSK) based (e.g., [104, 115] ) and phase shift keying (PSK) based

(e.g., [116, 117, 118]) modulations, though other modulations (e.g., see [119])

may also be used. Normally, coherent or non-coherent detectors that are im-

plemented using analogue hardware are used. Generally speaking, coherent

demodulation is found to be less adequate for digital implementation. For

example, the software implementation of PSK demodulation in [120] used ex-

cessive over-sampling that may not be practical for some applications. On the

other hand, non-coherent demodulation is found more attractive since it lends

itself well to digital implementation without extra requirements, as will be de-

monstrated in the next section.

Towards an all-digital implementation, a digitized version of the matched-

filter coarse synchronization in Fig. 4.1 is implemented by detecting the power

of each hop frequency in the DFT domain, as will be presented in Section 4.4.

The method is essentially implemented to track preambles.

Fine synchronization with preamble hops is also implemented in the digital

domain using simple cross-correlation of discrete-time signals. The process is

carried out as part of the preamble decoding operation and it gives a precision

in the order of a sample duration. Details are presented in Section 4.4.

4.3 Modulation Scheme

An important feature of a modulation scheme is the detection requirements of

that scheme at the receiver side. Detection methods can be classified into two

broad categories, namely, coherent and non-coherent detection. Coherent detec-

tion requires strict carrier synchronization at the receiver that involves both

carrier frequency and phase. Carrier synchronization generally requires com-

plex carrier tracking circuitry to be included as a part of the receiver. On the

other hand, non-coherent detection does not require carrier synchronization.

However, symbol synchronization is required in all cases [50].
In the case of FHSS transmission, the carrier is persistently changing resulting in difficulties when coherent detection is used [51]. Carrier tracking with FHSS is even more difficult in the multiuser case. Therefore, non-coherent modulation schemes are found to be prevalent in FHSS systems, specifically, frequency-shift keying (FSK) and differential phase-shift keying (DPSK) are used with FHSS to facilitate signal demodulation and detection. The price paid is, however, loss in performance compared to coherent systems [50].

In this work, DPSK, or more precisely, differential binary phase-shift keying (DBPSK) has been chosen as the modulation scheme for transmitting binary data over preambles. This choice is motivated by two factors. First, it is well known that DBPSK outperforms BFSK (binary FSK) in terms of bit error rate and robustness [50]. The second driving factor is that under certain practical conditions, the bandwidth required for BFSK exceeds that required for DBPSK (as will be presented in Section 4.4). This seems to contradict previous findings that the bandwidth required for both modulation schemes is nearly equivalent, or even BFSK signaling produces a more compact spectra compared to DBPSK [50].

### 4.3.1 Differential Binary Phase-Shift Keying (DBPSK) Modulation

In this section, we focus on binary transmission using DBPSK which is the non-coherent counterpart of binary phase-shift keying (BPSK). The main difference between BPSK and DBPSK modulations is that in BPSK modulation, information is encoded in the phase of each information symbol, while in DBPSK modulation, information is encoded in the phase-difference of each pair of consecutive symbols. For example, in DBPSK, binary 0 can be encoded as a phase-difference of 0 radians and binary 1 as a phase-difference of $\pi$ radians. Fig. 4.3 (a) and (b) shows an example that illustrates the difference between BPSK and DBPSK modulation schemes. This conceptual difference of the two modulation schemes does not induce any difference in the spectral properties of the two types [50]. However, the need for a start-up symbol that is naturally required for the differential encoding process, reduces (normally slightly) DBPSK modulation data rates [50]. Performance-wise, DBPSK exhibits slightly higher bit error rate than BPSK [50].
Figure 4.3: Example for a) BPSK and b) DBPSK signals for a transmitted binary sequence: 0, 1, 1 and 0.

Figure 4.4: Block diagram of a DBPSK transmitter.

4.3.1.1 DBPSK Transmitter

A block diagram of a DBPSK transmitter is depicted in Fig. 4.4. The transmitter consists of a logic network that is interconnected to a one-bit delay unit so as to convert the raw binary sequence \( b_k \) to a differentially encoded sequence \( d_k \). The latter sequence is amplitude-level encoded and then used to modulate a carrier wave of frequency \( f_c \) to produce the desired DBPSK signal. For more details, the reader is referred to [50]. An example of the output signal is shown in Fig. 4.3 (b).
4.3 MODULATION SCHEME

4.3.1.2 DBPSK Receiver

A block diagram of a general digital receiver is depicted in Fig. 4.5. The receiver consists of two parts, namely, a demodulator and a detector. The function of the demodulator is to convert the received waveform $x(t)$ into an $N$-dimensional vector $\mathbf{y} = [y_1, y_2, ..., y_N]$, where $N$ is the dimension of the transmitted signal waveform. Based on the vector $\mathbf{y}$, the detector decides which of an $M$ waveforms was transmitted [54]. The received signal of consideration has the form

$$x(t) = s_m(t) + v(t) \quad (4.1)$$

where $v(t)$ is a sample of a noise process, which is for simplicity assumed to be a additive white Gaussian noise (AWGN) process with a zero mean and power spectral density $\eta_0/2$; and $s_m(t), m = 0, ..M − 1$ represents the transmitted waveforms. For DBPSK, we have $M = 2$ and the waveforms $s_m(t)$ can be expressed as

$$s_0(t) = \sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_c t) g \left( t - \frac{T_b}{2} \right) + \sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_c t) g \left( t - \frac{3T_b}{2} \right)$$

$$s_1(t) = \sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_c t) g \left( t - \frac{T_b}{2} \right) - \sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_c t) g \left( t - \frac{3T_b}{2} \right) \quad (4.2)$$

where $f_c$ is the carrier frequency; $E_b$ is the symbol energy over the time duration $T_b$ and is assumed to be constant; and $g(t)$ is a baseband function that is defined as $g(t) = 1, -T_b/2 \leq t < T_b/2; g(t) = 0$, otherwise. Eq. (4.2) represents the waveform for binary 0 as a continuous waveform (with zero phase-shift) over a duration of $2T_b$, which is equivalent to two symbol durations in BPSK. The waveform for binary 1 is, on contrast, subject to a phase-shift of $\pi$ radians in the middle of the $2T_b$ waveform duration. The waveforms in Eq. (4.2) clearly satisfy $\sum_{t=0}^{2T_b} s_0(t)s_1(t) = 0$ (inner product is equal to zero), which indicates that they are orthogonal. It can therefore be concluded that DBPSK can be considered as a form of non-coherent orthogonal modulation when considered over a period of two-symbol duration ($2T_b$). Also, for
DBPSK, it can be shown that $N = 2$ and the waveforms can be decomposed into the basis functions

$$
\phi_{0c}(t) = \sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_c t) g(t - \frac{T_b}{2}) + \sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_c t) g(t - \frac{3T_b}{2})
$$

$$
\phi_{0s}(t) = \sqrt{\frac{2E_b}{T_b}} \sin(2\pi f_c t) g(t - \frac{T_b}{2}) + \sqrt{\frac{2E_b}{T_b}} \sin(2\pi f_c t) g(t - \frac{3T_b}{2})
$$

$$
\phi_{1c}(t) = \sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_c t) g(t - \frac{T_b}{2}) - \sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_c t) g(t - \frac{3T_b}{2})
$$

$$
\phi_{1s}(t) = \sqrt{\frac{2E_b}{T_b}} \sin(2\pi f_c t) g(t - \frac{T_b}{2}) - \sqrt{\frac{2E_b}{T_b}} \sin(2\pi f_c t) g(t - \frac{3T_b}{2})
$$

where $\phi_{mc}(t)$ and $\phi_{ms}(t)$ are normally referred to as the in-phase and quadrature components of $s_m(t)$.

A form of an optimum DBPSK receiver is depicted in Fig. 4.6, where a correlation demodulator is used. Generally, a correlation demodulator expands the received signal (together with the noise) into a series of linearly weighted orthonormal basis functions. For the receiver depicted in Fig. 4.6, the set of basis functions pertaining to $s_0(t)$ and $s_1(t)$ are given by Eq. 4.3. This kind of receiver is also referred to as quadrature receiver. An alternative form of the quadrature receiver may equivalently employ matched-filters in place of the correlators [50].

For a quadrature receiver, the $y$ vector that constitutes the input to the detector is constructed according to $y_m = \sqrt{x_{mc}^2 + y_{ms}^2}$, $m = 0, 1$. The detector, then compares the two variables $y_m$, which represent sufficient statistics for making the decision as to which waveforms was transmitted [54]. The decision is made as follows. If $y_0 > y_1$, $s_0(t)$ was transmitted and the output of the detector is set to binary 0. On the other hand, the detector yields binary 1 when $y_1 > y_0$. In the case where $y_0 = y_1$, the detector can remain undecided, or alternatively, a fair coin can be flipped if hard decision is required [50].

For the DBPSK receiver described above, the probability of error assuming equiprobable symbols with equal energy, is given by [50, 51, 54]

$$
P_e = \frac{1}{2} e^{-\frac{E_b}{2}}
$$

(4.4)

The probability $P_e$ represents the probability that the detector produces an erroneous output that is equal to binary 1 when $s_0(t)$ was transmitted, or equal
4.3 MODULATION SCHEME

$$\text{Figure 4.6: Block diagram of a DBPSK quadrature receiver.}$$

to binary 0 when $s_1(t)$ was transmitted. Fig. 4.7 plots $P_c$ versus $E_b/\eta_0$.

Fig. 4.8 depicts an alternative to the quadrature receiver, where the demodulator directly cross-correlates the received signal with each of the two possible transmitted signals. This is shown to be equivalent to the quadrature receiver in terms of performance and can hence be used in place of the latter [54].

4.3.1.3 Digital Implementation of The DBPSK Receiver

A digital version of the DBPSK receiver depicted in Fig. 4.8 is shown in Fig. 4.9. Analog correlators (integrations) are replaced by digital ones (summation) to cope with the input signals, which are discrete in this case. This receiver is obtained by replacing the continuous-time variable $t$ by $nT_s$, where $n = 0, 1, 2, ...$ and $T_s$ is the sampling interval. This results in the discrete waveforms, $s_{0k}[nT_s]$ and $s_{1k}[nT_s]$. For the orthogonality property to be maintained for these sampled signals, the sampling theorem [42] has to be respected. Namely, the sampling rate, $F_s = 1/T_s$, has to satisfy $F_s \geq 2W$, where $W$ is the highest frequency in the frequency band of interest. The maintenance of the orthogonality property can be directly implied from the relationship between the Fourier transforms of the continuous-time and the corresponding discrete-time signals.
Figure 4.7: Probability of bit error ($P_e$) for DBPSK modulation versus the transmitted energy per bit to noise power ratio ($E_b/\eta_0$).

Figure 4.8: Block diagram of a DBPSK receiver alternative structure.

Figure 4.9: Block diagram of digital implementation of a DBPSK receiver.
4.4 The Proposed Acquisition and Synchronization System

A preamble sequence is exploited to facilitate HP acquisition and synchronization. Fig. 4.10 shows the data frame format to be used. Each data frame is divided into two parts—a preamble and a payload. The preamble part uses $K$ channels to carry binary information. The $K$ channels are used for data transmission in sequence such that data transmission starts off at channel 1 followed by channel 2 until all $K$ channels are used. The transmitter repeats the sequence of channels a total of $R$ times (or cycles). Each channel is used to transmit a single binary digit per cycle using DBPSK (which is equivalent to two binary symbols), giving a total of $RK$ binary digits. These binary digits represent the transmitter’s ID.

The payload uses $H$ channels that are different from the channels used for the preamble. Since no actual data transmission takes place in the payload, only sinusoidal pulses (the unmodulated carriers) are transmitted. For convenience, these can be viewed as DBPSK with all the transmitted data equal to binary 0. To conserve bandwidth, phase continuity between different channels is maintained in the payload. A transmitter typically transmits the $H$ channels in a non-sequential manner determined by its hopping pattern (HP). The dwell time of a transmitter for each channel is fixed for both the preamble and the payload and is denoted as the hop duration. Note that the payload may also contain silence periods as recommended in Chapter 3. However, for the sake of simplicity, herein, it will be assumed that such silence periods do not occur in the payload.

Let us designate a hop duration as $T_h$ (seconds), then the symbol duration is $T_b = T_h/2$ sec. Considering the preamble, the bandwidth required for transmitting a symbol is approximately $1/T_b$ Hz. However, for the payload due to the continuity of phase within each hop, the bandwidth required for each hop will be equal to $1/T_h = 1/2T_b$ Hz. Assuming an expected frequency offset due to different factors (including offset at the transmitter, offset at the receiver, Doppler, etc.) of $f_g$ Hz, the minimum required frequency separation is $\Delta f = 2/T_h + 2f_g$ Hz for the preamble and $\Delta f = 1/T_h + 2f_g$ Hz for the payload. Note that the amount of Doppler shift is frequency dependent, which asserts that either the frequency separation should be different between different pairs of frequencies, or the maximum Doppler offset should be adopted. Here, we consider the latter approach.
The frequency separation $\Delta f$ determines the total bandwidth required for the communication system for different modulation schemes. Since the actual data transmission occurs in the preamble, the bandwidth required for preamble transmission will be used as a benchmark. Let us assume that a limited bandwidth of $W_p$ Hz is available for transmitting the preamble. The maximum size of the set of frequencies that are available for any modulation scheme is approximately given by $W_p = W_p/(2/T_h + 2f_g)$ for DBPSK. If BFSK is to be considered over the bandwidth $W_p$, the size of the frequency set will be $N_{FSK} = W_p/(1/T_h + 2f_g)$. It is noticed that for DBPSK and BFSK to result in the same data rate, $N_{FSK}$ must be twice as large as $N_{DPSK}$. If $N_{FSK}$ is less than twice $N_{DPSK}$, that means that DBPSK will be more efficient than BFSK as far as bandwidth is concerned. The following design criterion illustrates the grounds on which DBPSK has been favoured on BFSK. It is found that the minimum Doppler that should be factored into the design is at least $1/2T_h$. This asserts that $N_{FSK} = 1.5N_{DPSK}$, which is less than the required proportion for BFSK to be the attractive choice. It is noted that from the definition of $N_{FSK}$ and $N_{DPSK}$, the larger $f_g$ is, the less attractive BFSK becomes. The two modulation schemes become equivalent when $f_g$ is zero, i.e., when the system ideally does not suffer frequency offsets. In addition to the above criterion, another attractive feature of DBPSK is its lower bit error rate compared to BFSK in the case of an AWGN channel [50].

### 4.4.1 Preamble Detection and Coarse Synchronization

The first step for the receiver is to acquire the IDs of the transmitters in range. To do this the receiver must detect and decode the preamble sequences. We will start by assuming a single transmitter that transmits a preamble every $T_p$ seconds.

It is well known that a DFT decomposes a signal into separate frequency
Figure 4.11: Preamble detection system.
bands. Bandpass information can therefore be gained from the DFT of a signal if the frequency bins occupied by the hopping sequence are known. In our case, all of the hops and the associated bandwidth involved in preamble transmission are known a priori. Therefore, the corresponding frequency bins can be calculated for any DFT size (with uncertainty due to frequency offset). Note that the effect of frequency offsets has already been accounted for in the signal design stage such that a designated frequency range will almost always contain the contribution from the corresponding frequency channel. Due to the discrete nature of the DFT in frequency, the boundaries of the frequency bands associated with the hops may need to be moved up or down by an amount of $\pm \Delta f_{dft}/2$ Hz that is equivalent to half of the frequency resolution of the DFT used.

The preamble detection system is illustrated in Fig. 4.11. It can be viewed as a digital version of the matched-filter acquisition system illustrated in Fig. 4.1. It therefore inherits the fast acquisition property of the latter. In fact, the scheme is motivated by the fact that the DFTs of the received signals are also required for AOA estimation (as will be shown in the subsequent chapters). Therefore, the proposed preamble detection system does not require much extra computational cost.

The operation of the proposed preamble detection scheme (which will be shown to embody also acquisition and coarse synchronization) can be described as follows. First, a window of the received signal samples that correspond to one hop duration is collected and Fourier-transformed using an N-point DFT. The powers in the subbands pertaining to each frequency channel are calculated. A threshold detector compares the power of each subband to a predetermined threshold. If the threshold is exceeded, arrival of the designated hop is declared by generating a binary value of 1 at the end of the hop duration, otherwise, the output of the detector is set to binary 0. Each of the binary signals from the threshold detectors is delayed by a suitable number of hop durations. More specifically, the output of the threshold detector of band $k_i$ is delayed by $(K - k_i)T_h$ sec. The delayed outputs are added to form the signal $L(t)$. Next, $L(t)$ is passed to the threshold generator to produce the adaptive threshold $V(t)$, which is compared to a delayed version of $L(t)$ named $D(t)$. The comparator produces a binary output of 1 when $D(t) \geq V(t)$ and 0 otherwise. An effective choice of $V(t)$ is

$$V(t) = \min\{L(t) + l_0, K\}$$  (4.5)
where $l_0$ is a positive integer. This choice is aimed at protecting from false alarms due to continuous interference [52]. The comparator output $a_1(t)$ represents the binary decision of the detection of a cycle of the preamble sequence. This may be sufficient to decide on the reception of the whole preamble. However, it is more robust to detect reception of all $R$ cycles of the preamble. To achieve this, $a_1(t)$ is delayed $R - 1$ times, each time using a delay of $KT_h$ sec. This produces additional $R - 1$ binary signals, which are added to form the final output $y(t)$. The comparator detects the whole preamble sequence simply by checking whether $y(t)$ is equal to $R$ or not. When $y(t) = R$, this means that all the $R$ preamble cycles have successfully been detected.

It should be noted that due to misalignment, windowing can result in a portion of a hop being in one window and a portion in the next window. If the proportions are close to 0.5, this may lead to a duplication of the detection of possibly all hops and hence of the whole preamble sequence. Naturally, the receiver will consider the first detection.

### 4.4.2 Preamble Decoding and Fine Synchronization

Once the preamble has been detected, the receiver will be synchronized to the received preamble sequence to within a half of a hop duration, or $T_h/2$ seconds. For decoding to be carried out, the timing uncertainty should be significantly reduced. This can be achieved by exploiting two types of orthogonality in the system. First, the set of all hops (regardless of the data encoded therein) represents an orthogonal set of signals by virtue of design. Second, within each hop DBPSK modulation is used, where binary digits are represented by two orthogonal signals. These two forms of orthogonality can be used to synchronize the receiver to the received signal. The process is described in the following paragraphs.

The receiver starts by picking its local (sampled) version of the two binary waveforms $s_{01}[n]$ and $s_{11}[n]$ pertaining to hop 1. Note that hop 1 coincides with channel 1 and that there is a total of $RK$ hops in the preamble. Each of $s_{01}[n]$ and $s_{11}[n]$ has the same length as the window of the received signal. The receiver starts off at $N_h/2$ samples before the start of the preamble that is declared by coarse synchronization. The waveforms $s_{01}[n]$ and $s_{11}[n]$ are multiplied by the received signal window and the absolute difference of the absolute values of the two cross-correlations $(|\Delta E_1|)$ are determined. The received signal window is slid one sample at a time until a peak of $|\Delta E_1|$ is formed. At that par-
ticular moment, the receiver can be said to have locked to the hop frequency pertaining to hop 1 to within a few samples. The process is then repeated for the next hops in order. The cross-correlation of the binary symbol \( b \) at hop \( k \) with the received signal can be calculated as

\[
E_{bk}[l] = \sum_{n=n_{0k}-n_{uk}}^{n_{0k}+(N_h-n_{uk})-1} x[n+l]s_{bk}[n-(n_{0k}-n_{uk})], \quad l = 0, 1, \ldots, N_h
\]  

(4.6)

where \( s_{bk}[n], b = 0, 1 \) are the sampled waveforms for binary 0 and binary 1 pertaining to hop \( k \); \( n_{0k} \) is a start-up position, which for the first hop is determined from coarse synchronization, and for the remaining hops is determined from the previous hop timing; and \( n_{uk} \) are arbitrary numbers of samples that account for the uncertainty in the timing of the start up point for each frequency. For hop 1, \( n_{u1} = N_h/2 \) is used. For other hops, \( n_{uk} \) can be selected arbitrary and are recommended to be less than or equal to \( N_h/2 \). This is aimed at preventing errors from the previous stages from affecting the current stage.

The absolute difference of the absolute values of the two cross-correlations is defined as

\[
\Delta E_k[l] = |E_{0k}[l]| - |E_{1k}[l]|
\]  

(4.7)

and the fine timing acquisition process for hop \( k \) can be summarized as

\[
n_k = \arg \max_l |\Delta E_k[l]|.
\]  

(4.8)

The idea behind (4.8) can be described as follows. Without loss of generality, assume that binary 0 is being transmitted on hop \( k \), which uses a specific frequency channel. When the test signals \( s_{0k}[n] \) and \( s_{1k}[n] \) are not aligned at all with the received (noisy and possibly distorted) version of \( s_{0k}[n] \), \( |E_{0k}| \) and \( |E_{1k}| \) are zero due to the orthogonality of different hop frequencies. Consequently \( |\Delta E_k| \) will also be zero. When the test signals are partially aligned, but with less than 50% of alignment, \( |\Delta E_k| \) will again be zero due to \( |E_{0k}| \) and \( |E_{1k}| \) being equal. Nevertheless, the values \( |E_{0k}| \) and \( |E_{1k}| \) continue to increase as the overlap increases. When the overlap increases to more than 50%, \( |E_{0k}| \) will increase further, while \( |E_{1k}| \) decreases until the point of the best alignment of the test signals with the received hop of interest is reached. At that point, \( |E_{0k}| \) reaches its maximum and, due to perfect orthogonality, \( |E_{1k}| \) becomes zero. This results in \( |\Delta E_k| \) reaching its maximum. If the test signals are slid further, \( |\Delta E_k| \) will start to drop due to reduction in the correlation of the test signal \( s_{0k}[n] \) with the received signal.
4.4 THE PROPOSED ACQUISITION AND SYNCHRONIZATION SYSTEM

When the value of \( n_k \) is identified, hop \( k \) can be demodulated using the method described in Section 4.3.1. Equivalently, the quantities \( E_{0k} \) and \( E_{1k} \) can be used directly to decide which binary digit has been received in the same manner described in Section 4.3.1.

At the end of the process, the receiver will acquire the information pertaining to the transmitter’s ID. The receiver will also acquire timing information from different cycles. Due to noise, the values of \( n_k, \forall k = 1, ..., RK \) may result in different alignments. The receiver may compute the mean timing (of the start of the preamble) based on the observed values of \( n_k \) as

\[
\hat{n}_s = \frac{1}{RK} \sum_{k=1}^{RK} (n_k + n_{0k} - n_{uk}) - (k - 1)N_h
\]

where \( \hat{n}_s \) is the estimate of the time of the start of the preamble and hence of the whole data frame. The receiver will therefore be synchronized with the received HP to within a few samples. This is arguably sufficient for the localization system.

Fig. 4.12 shows the correlation of an example waveform \( s_{01}[n] \) at a frequency of 40 kHz with itself and other orthogonal waveforms. Each waveform is realized by 512 samples. The autocorrelation looks exactly like that of a narrowband signal; the magnitude grows towards the zero lag point forming both negative and positive peaks. In the ideal case, the positive peak occurs exactly at zero lag. The cross-correlation of \( s_{01}[n] \) and \( s_{11}[n] \) (\( s_{11}[n] \) represents binary 1 for the same frequency) shows four (two negative and two positive) peaks and a valley (actually two valleys, one corresponding to the two negative peaks and one to the two positive peaks. However, they will be treated as a single valley). The valley also coincides with the zero lag in the ideal case. The cross-correlation of \( s_{0k}[n] \) with, respectively, \( s_{02}[n] \) and \( s_{12}[n] \), which are for a frequency of 41 kHz are in general of lower magnitude as compared to those involving the two waveforms of the 40 kHz frequency. This indicates that the autocorrelation (upper left) can be sufficient for the decision concerning the detection of the waveform \( s_{01}[n] \). Alternatively, combining the autocorrelation with the cross-correlation of \( s_{01}[n] \) and \( s_{11}[n] \) in a similar way to Eq. (4.7) may provide a more robust detection. It is noticed that the cross-correlation of two binary digits encoded with two different frequencies is not perfectly zero at almost all lags. This is due to the imperfectness of the orthogonality of the signals under consideration.

Fig. 4.13 shows similar plots to those in Fig. 4.12, but for \( s_{11}[n] \) correlated with
itself and other different waveforms. The autocorrelation in this case is seen to have two smaller peaks beside the main peak and two valleys on both sides of the main peak. These two valleys do not interfere with the valley coinciding with zero lag. This property makes the detection of $s_{11}[n]$ possible in the same manner explained for $s_{01}[n]$.

In Fig. 4.14, the process of preamble decoding exemplified by a single hop decoding is demonstrated. Binary 1 is transmitted over a carrier of 38 kHz preceded by binary 0 transmitted over 35.5 kHz and followed by binary 1 over 39 kHz. The process is started 256 samples from the start of the hop of interest and ended 256 samples after the end of the latter (a hop is 512 samples in this example). In Fig. 4.14, the location of the peak of $|\Delta E_k|$ represents the best alignment between the test signals and the hop of interest (the 40 kHz hop), while the negative sign of the peak of $\Delta E_k$ indicates that the received binary is 1. Note that when the signals are perfectly aligned (i.e., at zero lag), the cross-correlations pertaining to one test signal ($E_{0k}$ or $E_{1k}$) or their absolute values may be sufficient for detecting the received binary. However, since the start and the end of the hop of interest are not known (actually they are to be determined), using both of the cross-correlations helps prevent ambiguities, for example, due to the smaller local peaks. Fig. 4.15 shows the effect of noise by plotting $|\Delta E_k|$ for different SNR values. The figure indicates that noise can have the effect of moving the peak of $|\Delta E_k|$ far from its true location. For SNR equal to 10 and 0 dB, the peak location is found to coincide exactly with that of the noise-free case, which is the true location of the peak. For SNRs of $-10$ and $-20$ dB, the peak location is found to be earlier by 4 and 99 lags, respectively. Fig. 4.16 shows the root mean squared error (RMSE) for the location of the peak of $|\Delta E_k|$ versus $E_b/\eta_0$ (the AWGN case). It can be seen that for $E_b/\eta_0$ lower than approximately 15 dB the RMSE remarkably grows, while $E_b/\eta_0$ above 15 dBs results in reliable performance that is sufficient for localization.

4.4.3 Multiple Transmitters

The previous discussion has considered a single transmitter scenario. Localization, however, requires at least three transmitters. FHSS naturally allows for such multiuser sharing. So, as far as the payload is concerned, the orthogonality of the HPs guarantees that collisions will not have a serious impact on the performance of the system. On the other hand, when the preamble part of the transmission is considered, it is clear that with the same preamble
Figure 4.12: Correlations of $s_{01}[n]$. 
Figure 4.13: Correlations of $s_{11}[n]$. 
Figure 4.14: Decoding of a preamble hop.
Figure 4.15: Noise effect on $|\Delta E_k|$. 

Figure 4.16: RMSE for $|\Delta E_k|$ peak location versus $E_b/\eta_0$. 
frequencies shared by all the transmitters, collision will be an issue. One can think of using different preamble HPs for different transmitters, but that makes preamble acquisition difficult, and still complete collision avoidance is not guaranteed. It is, therefore, reasonable and more efficient to use the same HP for all preambles.

Before going into the details of the proposed solution to the collision problem, it should be noted that the proposed solution is dictated by the lack of synchronization and communication between all the units involved, i.e., the transmitters and the receiver. The proposed algorithm for collision avoidance does not affect the previous discussion on single transmitter acquisition and synchronization. However, it is required to allow the system to scale without serious performance loss.

4.4.3.1 The Proposed Collision Avoidance Approach

- Each transmitter continuously transmits fixed-length frames each with a fixed-length payload and a fixed-length preamble-slot.
- In the preamble-slots, a transmitter either transmits its ID information or goes silent for the whole duration of the preamble-slot.
- The decision as to whether a preamble be transmitted during a preamble-slot or not is determined by a binary random variable whose values are determined independently for each transmitter.

From the first point, the same data frame format in Fig. 4.10 can be adopted. However, according to the second point, a preamble need not be sent with each data frame. In other words, a preamble-slot may occasionally be left empty. The goal of leaving a preamble-slot empty is to avoid permanent collision between the preambles of different transmitters. More specifically, when preamble-slots collide continuously, a certain preamble can still be detected since other preambles are occasionally empty. Continuous collision can alternatively be avoided by varying the frame size from transmission to transmission, but this makes the system more complicated. The third point proposes a random and independent manner for preamble transmission/non-transmission, to assure that collision avoidance works properly.

Now, let us assume that there are $N_b$ unsynchronized transmitters. The preambles of these transmitters are received by the receiver asynchronously
at different random times. During the acquisition process, the receiver is assumed to be stationary. The joint pattern represented by the reception times of all the preambles is periodic due to the periodicity of the transmission. In what follows, we focus on calculation of the probability of acquisition failure for a given preamble. Without loss of generality, let us focus on transmitter-1 and assume that the arrival of the preamble-slot of that transmitter occurs at time $t_1$. The arrival of the preamble-slot of the $i$'th transmitter occurs at time $t_i$. By the arrival time, we mean the instant of time at which the first symbol in a preamble-slot, if any, starts to be received. The time difference $d_i = t_i - t_1$ is the time when the preamble-slot of the $i$'th transmitter arrives relative to that of transmitter-1. The time differences $d_i, i = 2, ... , N_b$ can be looked at as random variables. It will be assumed that $d_i$ are independent and identically and uniformly distributed. The distribution of these random variables is confined to the interval $[-T_f/2, T_f/2]$, where $T_f$ is the duration of the whole data frame. In order to calculate the probability of collision, the following assumptions are made:

1. Assumption 1: for preamble-1 to be detectable, the arrival time of any other preamble must not coincide with the duration of preamble-1.

2. Assumption 2: for preamble-1 to be detectable, any other preamble must not occur at least a duration of $m$ hops before the arrival of preamble-1.

The purpose of assumption 2 is to allow enough time for multipath due to other (preceding) preambles to be remarkably attenuated before the reception of preamble-1, so that collisions due to multipath are negligible. Therefore, the value of the parameter $m$ should be chosen based on knowledge of the transmission channel characteristic and should be a zero for multipath-free channels. According to Chapter 3, the value of $m$ should coincide with the reverberation time of the space under consideration. However, it was empirically found that for the proposed modulation scheme, preambles were detectable even when other preambles occur at times that are considerably less than the reverberation time.

When time overlap between more than one preamble occurs, or a preamble occur less than $m$ hops duration before the arrival of preamble-1, preamble collision is said to have occurred.

The occurrence of preamble collision coincides with another event, that is the coincidence of preamble-slots, which is referred to as slot collision. In other
4.4 THE PROPOSED ACQUISITION AND SYNCHRONIZATION SYSTEM

words, slot collision is a prerequisite for preamble collision. The aforemen-
tioned rules for preamble collision can equally be valid for slot collision.

Based on the above discussion, the probability that slot-i (preamble-slot of the
\(i\)'th transmitter) collides with slot-1 (preamble-slot of transmitter-1) is given by

\[
P_{sc} = \Pr \left[ -(T_p + mT_h) < d_i < T_p \right] = \frac{(2T_p + mT_h)}{T_f} \tag{4.10}
\]

where \(T_p\) is the whole duration of a preamble. The result in (4.10) stems di-
rectly from the assumed uniform distributions of the data. Due to the assumed
identical distribution of \(d_i, \forall i\), \(P_{sc}\) is the same for any value of \(i\). From the as-
sumption that hopping rate is identical for both the preamble and the payload,
Eq. (4.10) can take the form

\[
P_{sc} = \frac{2RK + m}{RK + H} \tag{4.11}
\]

It is clear from (4.11) that \(P_{sc}\) completely depends on three quantities: the
lengths (in hops) of the preamble (given by the product \(RK\)) and the length
of the payload; and the parameter \(m\). Eq. (4.11) can, in fact, be used to deter-
mine any of these values at the design stage. The design should respect the
fact that \(P_{sc} \leq 1\). This gives rise to the inequality

\[
H \geq RK + m. \tag{4.12}
\]

Considering the inequality (4.12), coupled with the fact that the length of the
preamble is dictated by the minimum amount of data required to be commu-
nicated for beacon identification, and the fact that the value of \(m\) is required to
be above a certain lower limit to avoid multipath, it is reasonable to think of
using \(H\) as the adjustable design parameter for performance optimization. As
can be seen from Eq. (4.11), \(P_{sc}\) decreases as \(H\) increases, which is, in general, a
desirable effect. However, other effects have to be considered before adopting
the latter conclusion, as will be demonstrated subsequently. Fig. 4.17 plots \(P_{sc}\)
versus \(H\) for \(RK = 8\) and two different values of \(m; m = 4\) (a case of practi-
cal interest) and \(m = 0\) (the ideal case). The figure shows how \(P_{sc}\) decreases
as \(H\) increases. The figure also shows how putting a practical factor such as
multipath in consideration can result in a generally higher probability of slot
collision, as reflected in the \(m = 4\) graph.

Now, without loss of generality, let \(b_1(nT_f)\) and \(b_i(nT_f), i = 2, \ldots, N_b, n = \)
0, 1, 2, ... be independent Bernoulli random variables with identical distributions and a parameter \( p \). The values of these random variables control whether preamble-slots of, respectively, transmitter-1 and transmitter-i are occupied or not. The parameter \( p \) represents the probability that a preamble-slot being occupied, i.e., the probability that the corresponding random variable takes the value ‘1’. The probability that transmitter-i will cause a collision (to transmitter-1)—in terms of actual preambles not slots—is the probability that the two preamble-slots will collide and that slot-i will be occupied, which is equivalent to \( pP_{sc} \). The probability that \textit{at least} a single transmitter (of the \( Nb - 1 \) transmitters) will cause a collision will be given by

\[
P_c = 1 - (1 - pP_{sc})^{Nb-1}.
\]

(4.13)

Now, the probability of failure of the detection of preamble-1 in a cycle can be stated as the probability that either slot-1 is empty, or is occupied while a preamble collision is occurring. This probability also represents the noise-free probability of acquisition failure and is given by

\[
P_{f,0} = (1 - p) + pP_c.
\]

(4.14)

To obtain the aggregate probability of acquisition failure, bit error due to noise must be put into consideration. In the absence of error correction capability, a single bit error in the preamble will result in a wrong transmitter’s ID. It can be assumed that when no preamble collision occurs, bit error is solely due to AWGN noise. This is a reasonable assumption considering the definition of collision that has been discussed previously. The probability of erroneous reception of any bit is given by (4.4). The probability that at least one of the preamble bits is in error can directly be stated as

\[
P_{e,1} = 1 - (1 - P_e)^{RK}.
\]

(4.15)

Fig. 4.18 plots \( P_{e,1} \) versus \( E_b/\eta_0 \) for \( RK = 8 \). Generally, the probability of error is low for reasonably high values of \( E_b/\eta_0 \).

Now, the final aggregate probability of acquisition failure in a cycle considering the effect of noise can be stated as

\[
P_f = (1 - p) + p [P_c + (1 - P_c)P_{e,1}]
\]

(4.16)

Note that Eq. (4.14) can be obtained from Eq. (4.16) by setting the noise contri-
4.4 THE PROPOSED ACQUISITION AND SYNCHRONIZATION SYSTEM

Figure 4.17: The probability of slot collision ($P_{sc}$) versus the number of payload hops ($H$). The number of preamble hops is $RK = 8$.

Figure 4.18: The probability of at least one bit error ($P_{e,1}$) for DBPSK modulation versus the transmitted energy per bit to noise power ratio ($E_b/\eta_0$). The number of digits $RK = 8$
bution \((P_{e,1})\) equal to zero. Now, rearranging Eq. (4.16) and substituting for \(P_e\) from Eq. (4.13), \(P_{e,1}\) from Eq. (4.15) and \(P_{sc}\) from Eq. (4.11), we obtain

\[
P_f = 1 - (1 - P_{e,1})p (1 - pP_{sc})^{N_b - 1}
\]

\[
= 1 - (1 - P_e)^{RK} p \left[ 1 - p \left( \frac{2RK + m}{RK + H} \right) \right]^{N_b - 1}.
\] (4.17)

It can be seen from Eq. (4.17) that the probability of acquisition failure depends on the effect of noise and different parameters, namely, \(K, R, H, m, N_b\) and \(p\). The latter parameter, in particular, needs to be adjusted such that performance is optimized. Once the design value of \(p\) is found, the process that controls the transmission of preambles can be realized. Note that \(p\) is the only parameter that is pertaining to a stochastic process, and is hence the parameter that gives a stochastic nature to the process that represents the success/failure of the preamble detection.

To obtain the optimal value of the parameter \(p\), the well-known first derivative approach can be utilized, that is, the partial derivative of \(p_f\) in (4.17) is set equal to zero, which gives

\[
\frac{\partial P_f}{\partial p} = -(1 - P_{e,1})(1 - pN_bP_{psc})(1 - pP_{sc})^{N_b - 2} = 0.
\] (4.18)

Eq. (4.18) has \(N_b - 2\) roots of \(p = 1/P_{sc}\) and one root of \(p = 1/(N_bP_{psc})\). Since we are dealing with probabilities, search for the optimal value of \(p\) must be confined to the interval \([0, 1]\). Consequently, the \(p = 1/P_{sc}\) roots will be out of the question, while the \(p = 1/N_bP_{psc}\) root may give a valid optimal value for \(p\) in the correct interval in some cases. Therefore, the only possible optimal value of \(p\) is given by

\[
p_{opt} = \frac{1}{N_bP_{psc}} = \frac{RK + H}{N_b(2RK + m)}.
\] (4.19)

It can be seen from (4.19) that the validity of \(p_{opt}\) depends on different parameters, and that noise is irrelevant to the optimization process. For a set of system parameters that produce a valid \(p_{opt}\) value, the latter value will represent the optimal average preamble transmission rate (per hopping cycle). On the other hand, when \(p_{opt}\) is out of the valid interval, it can be concluded that the function \(P_f\) given by (4.17) does not have an extremum in the interval \([0, 1]\). Contemplating Eq. (4.18) in the open interval \((0, p_{opt})\), it can be shown that the
first derivative \( \frac{\partial P_f}{\partial p} \) is negative in the whole interval, meaning that \( P_f \) is a decreasing function in \( p \). Thus, for \( p_{opt} > 1 \), the optimal value of \( p \) will be the largest value in the valid probability interval, namely, \( p = 1 \). This result is mathematically convincing, however, practically, \( p = 1 \) means that preambles are transmitted in all slots, which may result in continuous collision. To avoid such a situation, \( p \) should be set to a value that is slightly less than unity. It should be noted here that for the case where an extremum is detected in the interval \([0, 1]\), it can be verified to be a minimum using a well-known second derivative test whereby the second derivative must satisfy the condition \( \frac{\partial^2 P_f}{\partial p^2} = 0 \) in order for the extremum to be a minimum.

By substituting (4.19) back in (4.17), the minimum probability of acquisition failure will be given by

\[
p_{f,\text{min}} = 1 - \frac{1}{N_b p_{sc}} (1 - P_e) \left( \frac{N_b - 1}{N_b} \right)^{N_b-1} 
= 1 - \frac{RK + H}{Nb(2RK + m)} (1 - P_e)^{RK} \left( \frac{N_b - 1}{N_b} \right)^{N_b-1}.
\]  

(4.20)

Fig. 4.19 plots \( P_f \) versus \( p \) for different values of \( N_b \), different values of \( E_b/\eta_0 \), \( RK = 8 \), \( m = 4 \) and \( H = 14 \). Fig. 4.20 is similar to Fig. 4.19, but the number of transmitters is fixed at \( N_b = 3 \), while \( H \) is varied. The two figures show a general trend of increase in \( P_f \) when \( E_b/\eta_0 \) decreases. In Fig. 4.19, it can be seen that \( P_f \) notably increases as the number of transmitters increases. For all the cases shown in Fig. 4.19, \( P_f \) is seen to have minima for \( p \) values that fall in the valid probability range. The value of \( p \) where the minimum occurs (\( p_{opt} \)) decreases with increase in \( N_b \) and does not vary with \( E_b/\eta_0 \). This means that increasing the number of transmitters requires decreasing the mean preamble transmission rate in order to reduce collision. On the other hand, in Fig. 4.20, for a fixed \( N_b \), the probability of collision is seen to decrease as \( H \) increases, while \( p_{opt} \) increases consistently with increase in \( H \). The latter observation agrees with the fact that for large \( H \), preamble duration becomes relatively small compared to the whole frame length, allowing for a higher preamble transmission rate, with relatively low collision rate. For the two graphs pertaining to the largest two values of \( H \) (\( H = 56 \) and \( H = 70 \)), the values of \( p_{opt} \) are greater than 1, and are therefore invisible in the plot.

From the assumption that \( b_i(nT_f), i = 1, \ldots, N_b, n = 0, 1, 2, \ldots \) are independent and identically distributed (i.i.d.) random variables, and that bit error is bitwise independent, the probability of acquisition failure will be the same for
all cycles. The probability that acquisition takes place in the \( x \)th cycle is equivalent to the probability that acquisition fails in all the previous cycles and succeeds in the \( x \)th cycle. The acquisition time \( (T_{acq}) \) is equivalent to \( xT_f \). Thus, the probability of acquisition time being equal to \( x \) frame durations can be stated as

\[
P(T_{acq} = x) = P_f^{x-1}(1 - P_f)
\]

which represents the probability density function (PDF) for acquisition time. The cumulative distribution function (CDF) will be given by

\[
P(x) = P(T_{acq} \leq x) = \sum_{u=1}^{x} P(u) = \sum_{u=1}^{x} P_f^{u-1}(1 - P_f) = 1 - P_f^x.
\]

An important parameter is the mean acquisition time, which can be obtained from (4.21) as

\[
\mu_{acq} = (RK + H) \sum_{x=1}^{\infty} xP(x) = (RK + H) \sum_{x=1}^{\infty} xP_f^{x-1}(1 - P_f)
\]

\[
= \frac{RK + H}{1 - P_f} = \frac{RK + H}{(1 - P_e)RK p [1 - p \left( \frac{2RK + m}{RK + H} \right)]^{N_b-1}}
\]

Similarly, the variance can be expressed as

\[
\sigma_{acq}^2 = \frac{(RK + H)^2}{(1 - P_e)RK p [1 - p \left( \frac{2RK + m}{RK + H} \right)]^{N_b-1}}
\]

The multiplication by \( RK + H \) and \( (RK + H)^2 \) in (4.23) and (4.24), respectively, is required to obtain the result in units of hop durations (or simply hops) rather than frame durations. From (4.23) and (4.24) the mean squared value (MSV) of
acquisition time is given by

\[
\zeta_{acq}^2 = \mu_{acq}^2 + \sigma_{acq}^2 = (1 + P_f) \left( \frac{RK + H}{1 - P_f} \right)^2
\]

\[
= \left\{ 2 - (1 - P_e)^{RK} p \left[ 1 - p \left( \frac{2RK + m}{RK + H} \right)^{N_b - 1} \right] \right\}^2
\]

\[
\times \left\{ \frac{RK + H}{(1 - P_e)^{RK} p \left[ 1 - p \left( \frac{2RK + m}{RK + H} \right)^{N_b - 1} \right]} \right\}^2
\]

(4.25)

where \( \zeta_{acq} \) represents the root mean squared value (RMSV) of acquisition time.

It can be indicated from Eqs. (4.23) and (4.24) that the mean, the variance and the MSV of acquisition time are all dependent on the same parameters as \( P_f \), but in somewhat different ways. To obtain the optimal value of the average preamble transmission rate that minimizes \( \mu_{acq} \), Eq. (4.23) is differentiated and equated to zero. This results in exactly the same solutions as for \( P_f \). Applying the same procedure to Eq. (4.24) and Eq. (4.25) gives the same result as for Eq. (4.23) plus additional roots that fall out of the interval \([0, 1]\). It can therefore be concluded that the optimal average preamble transmission rate is given by (4.19), or if (4.19) gives an invalid rate, an average preamble transmission rate that is close—but not equal—to unity should be adopted.

Fig. 4.21 plots \( \mu_{acq} \) versus \( p \) for different number of transmitters \( (N_b) \), different values of \( E_b/\eta_0, RK = 8, m = 4 \) and \( H = 14 \). Fig. 4.22 is similar to Fig. 4.21, but the number of transmitters is fixed at \( N_b = 3 \), while \( H \) is varied. The two figures show a general trend of increase in \( \mu_{acq} \) as \( E_b/\eta_0 \) decreases in a similar way to that in Fig. 4.19 and Fig. 4.20. Similar to \( P_f \) in Fig. 4.19, in Fig. 4.21, \( \mu_{acq} \) increases as the number of transmitters increases, however, the effect is more pronounced for the higher values of \( p \). Also, in Fig. 4.21, the minima of \( \mu_{acq} \) coincide with those of \( P_f \) in Fig. 4.19, again emphasizing the importance of decreasing the average preamble transmission rate in optimizing performance when the number of transmitters is increased.

Contrary to Fig. 4.20, in Fig. 4.22, \( \mu_{acq} \) does not really decrease as \( H \) increases—the general trend is that increasing \( H \) has the effect of increasing \( \mu_{acq} \) for a lower portion of the values of \( p \), while no systematic performance can be noticed for larger values of \( p \). This can be attributed to the \( RK + H \) term in the numerator of Eq. 4.23 that makes the numerator grows linearly with \( H \), while at the same time, the denominator grows nonlinearly with \( H \). This way of inclusion of the variable \( H \) in the expression for \( \mu_{acq} \) makes the relationship between \( \mu_{acq} \)
and $H$ rather complex. This complication can be interpreted physically by that increasing $H$ increases the preamble transmission cycle, and thus contributes to increasing $\mu_{\text{acq}}$. On the other hand, increasing $H$ helps in decreasing collision rate, thus contributing to the reduction of $\mu_{\text{acq}}$. Still, for a fixed value of $H$, the variation of $\mu_{\text{acq}}$ with $p$ agrees with that of $P_f$ in Fig. 4.20 in terms of the location of the minima.

Fig. 4.23 and Fig. 4.24 are the counterparts of Fig. 4.21 and Fig. 4.22 plotting the standard deviation ($\sigma_{\text{acq}}$). The figures are not much different from Fig. 4.21 and Fig. 4.22. This can be realized from Eqs.(4.23) and (4.24), where $\sigma_{\text{acq}}$ can be written as $\sigma_{\text{acq}} = \sqrt{P_f} \mu_{\text{acq}}$. Since $\sqrt{P_f} << \mu_{\text{acq}}$ and the range for $P_f$ is very limited, the contribution of $\sqrt{P_f}$ is hardly visual in Figs. 4.23 and 4.24. It is also noticed in Figs. 4.21, 4.22, 4.23 and 4.24 that both $\mu_{\text{acq}}$ and $\sigma_{\text{acq}}$, for some values of $N_b$ and $H$, take very large values (extremely large values were excluded from the plots to improve visualization) close to the two extreme values of $p$, 0 and 1.

By showing the plots for $\mu_{\text{acq}}$ and $\sigma_{\text{acq}}$, the properties of both the components of the MSV of acquisition time are identified. In fact, the variation of the MSV (or equivalently, the RMSV) of acquisition time with different system parameters, is not much different from that of the two components, $\mu_{\text{acq}}$ and $\sigma_{\text{acq}}$. The MSV of acquisition time is the ultimate performance metric that we adopt for the selection of the system parameters. In this regard, two parameters are of interest, namely, $p$ and $H$ (assuming that the latter is not subject to other practical restrictions).

So far, it has been shown that an optimal value of $p$, $p_{\text{opt}}$, can be calculated for fixed system parameters (including $H$), and is the same for both $\mu_{\text{acq}}$ and $\sigma_{\text{acq}}$, and consequently for $\zeta_{\text{acq}}$. However, in our case, it is required to optimize over the two-dimensional space spanned by both $p$ and $H$. In other words, we are looking for the point $(p, H)_{\text{opt}}$ where the function $\zeta_{\text{acq}}^2$ is minimized. This can be obtained by solving the pair of simultaneous equations resulting from equating the two partial derivatives of $\zeta_{\text{acq}}^2$ with respect to $p$ and $H$ to zero. This procedure is found to be involved and requires nonlinear regression in order to obtain $(p, H)_{\text{opt}}$. Even though, it is not guaranteed that a valid pair $(p, H)_{\text{opt}}$ will be obtained. For instance, if $(p, H)_{\text{opt}}$ includes a $p > 1$ value, truncating $p$ to $p \approx 1$ as for the one-dimensional case, does not guarantee that the pair with the truncated value will coincide with the minimum of $\zeta_{\text{acq}}^2$.

Alternatively, $(p, H)_{\text{opt}}$ is selected based on simulation, whereby $\zeta_{\text{acq}}$ is evaluated for the whole range of values of $p$ with $H$ varied in a similar manner.
to that in Fig. 4.22 and Fig. 4.24, but here \( H \) is varied more smoothly. For each value of \( H = H_i \), the minimum is located and a pair \((p_{\text{opt},i}, H_i)\) is identified together with the corresponding \( \zeta_{\text{min},i} \), where \( p_{\text{opt},i} \) is the optimal value of \( p \) in the interval \([0, 1]\) and \( \zeta_{\text{min},i} \) is the corresponding (minimum) value of \( \zeta_{\text{acq}} \). The process is repeated for all values of \( H \) in the range of interest. At the end of the procedure \((p, H)_{\text{opt}}\) is selected as the pair \((p_{\text{opt},i}, H_i)\) that coincides with the minimum \( \zeta_{\text{min},i} \).

Fig. 4.25 plots \( \zeta_{\text{min}} \) (the set of all values of \( \zeta_{\text{min},i} \)) versus \( H \) for different values of \( N_b \) and different values of \( E_b/\eta_0 \). The variation with both \( N_b \) and \( E_b/\eta_0 \) is obvious. The minimum value of \( \zeta_{\text{min}} \) and the corresponding value of \( H \) increase as \( N_b \) increases (see the ‘x’ marks in on the graphs). Unlike all the previous cases, the locations of the minima of \( \zeta_{\text{min}} \) (i.e., values of \( H \) corresponding to the minima of \( \zeta_{\text{min}} \)) are affected by noise. It can be seen that as \( E_b/\eta_0 \) decreases, the locations of the minima of \( \zeta_{\text{min}} \) move slightly towards lower values of \( H \). Plots similar to Fig. 4.25 can be used to select the optimal value of \( H \) for a specific scenario. The corresponding (optimal) values of \( p \) are not shown, since once the optimal value of \( H \) is identified, the corresponding value of \( p \) calculated directly from (4.19) and truncated in the way mentioned previously if required.

### 4.4.4 Payload Synchronization

By payload synchronization we mean both coarse and fine synchronization applied to the payload part of the data frame after the HP has been acquired. The same procedure that has been proposed for preambles can be used for payload synchronization. In the case of the payload, the HP and the transmitted data is known in advance, which makes fine synchronization (or payload decoding) somewhat a redundant operation. However, this redundancy can be exploited to confirm the reliability of the process. In other words, the binary digits resulting from the decoding can be compared to the known digits and according to that the receiver forms an idea on how accurate the process was.

Due to the absence of error detection capability in the preamble decoding operation, the receiver also needs to compare the HPs acquired from the preamble and from the payload as a form of error detection. If the HP detected in the payload does not match the one associated with the acquired transmitter’s ID, this indicates that an error has occurred either in the preamble decoding or in the payload HP detection. To prevent the receiver from being overwhelmed
Figure 4.19: Probability of acquisition failure ($P_f$) versus average preamble transmission rate ($p$) for different number of transmitters ($N_b$), different values of $E_b/\eta_0$, $RK = 8$, $m = 4$ and $H = 14$. 
Figure 4.20: Probability of acquisition failure ($P_f$) versus average preamble transmission rate ($p$) for different payload lengths ($H$), different values of $E_b/\eta_0$, $RK = 8$, $m = 4$ and $N_b = 3$. 
Figure 4.21: Mean acquisition time ($\mu_{acq}$) versus average preamble transmission rate ($p$) for different number of transmitters ($N_b$), different values of $E_b/\eta_0$, $RK = 8$, $m = 4$ and $H = 14$. 

78
4.4 THE PROPOSED ACQUISITION AND SYNCHRONIZATION SYSTEM

Figure 4.22: Mean acquisition time ($\mu_{acq}$) versus average preamble transmission rate ($p$) for different payload lengths ($H$), different values of $E_b/\eta_0$, $RK = 8$, $m = 4$ and $N_b = 3$. 

(a) $E_b/\eta_0 = \inf$

(b) $E_b/\eta_0 = 6$ dB

(c) $E_b/\eta_0 = 3$ dB

(d) $E_b/\eta_0 = 0$ dB
Figure 4.23: Standard deviation of acquisition time ($\sigma_{acq}$) versus average preamble transmission rate ($p$) for different number of transmitters ($N_b$), different values of $E_b/\eta_0$, $RK = 8$, $m = 4$ and $H = 14$. 
Figure 4.24: Standard deviation of acquisition time ($\sigma_{acq}$) versus average preamble transmission rate ($p$) for different payload lengths ($H$), different values of $E_b/\eta_0$, $RK = 8$, $m = 4$ and $N_b = 3$. 
Figure 4.25: Minimum root mean squared value of acquisition time ($\zeta_{min}$) versus payload length ($H$) for different number of transmitters ($N_b$), different values of $E_b/\eta_0$, $RK = 8$ and $m = 4$. The locations of the minima of $\zeta_{min}$ are marked with ‘x’.
by errors in the payload such that acquisition failure is repetitively reported, the receiver may allow for a reasonable number of errors in the detection of the payload HP. In other words, the receiver only needs to confirm that the detected HP sufficiently matches that associated with the acquired transmitter’s ID.

4.5 Practical System and Experiments

The system consisted of three ultrasonic transmitters ($N_b = 3$ is the minimum required for 3-D positioning) transmitting at a hopping rate of approximately 326.8 hops/second, which corresponds to a hop and DBPSK symbol durations of, respectively, 3.06 and 1.53 ms. The hopping rate was chosen based on the discussion in Chapter 3. Other system design parameters were $R = 2$, $K = 4$ and $m = 4$. The value of $m$ was meant to allow for a duration of 12.24 ms (which corresponds to a distance of approximately 4.2 meters) for multipath pertaining to all previous preambles to be attenuated.

Noise levels in the operational environment are quite unpredictable. The range 3–30 dB for $E_b/\eta_0$ was found convenient for the design to be carried out. For this range, the optimal values of $H$ are found to range from 55 to 58 hops, with the corresponding value of $p$ equal to 1 in all of the cases. For system implementation, the value $H = 56$ was first considered that corresponds to 4 times of the number of available payload channels. To avoid permanent collision, an average preamble transmission rate of 0.90 was used. From Figs. 4.22 and 4.24, it can be seen that for $(p, H) = (0.9, 56)$, the performance is close to that of other $(p, H)$ pairs, for instance, $(0.37, 14)$ and $(0.90, 70)$. For some reasons, one may prefer one of these pairs or the other on the $(0.9, 56)$ pair. For instance, if $H = 14$ is closer to the number of channels available for payload transmission, the $(0.37, 14)$ pair will be favoured. However, the $(0.90, 70)$ may be preferred due to the low preamble processing required (preamble-slot rate is relatively low due to the relatively large value of $H$). Note that if the latter choice is made, any deficiency in the number of available channels has to be compensated for.

As an example, the $(0.90, 70)$ pair. In this case, the payload will consist of 78 hops, which gives a preamble-slot transmission rate of approximately 4.19 preamble-slots/s (one preamble-slot every 239 ms), and an average preamble transmission rate of 3.77 preamble/s. Fig 4.26 shows the mean, the stan-
standard deviation and the RMSV (in milliseconds) that correspond to \((p, H) = (0.90, 70)\) against \(E_b/\eta_0\). It can be seen that for the range of \(E_b/\eta_0\) of interest, the RMSV of acquisition time asymptotically saturates to a value of approximately 543 ms.

### 4.5.1 Experimental Setup

In this subsection, we report on experiments that have been carried out to test the proposed methods for acquisition and synchronization. Since the performance in terms of the MSV of acquisition time is well characterized analytically, in this subsection we will focus on preamble detection and decoding for acquiring the transmitter’s ID. The latter process entails both the required coarse and fine synchronization.

Throughout the experiments, the signals were sampled at a rate of 167.857 samples/s, which results in approximately 512 samples per hop. The frequency range was 35 to 50 kHz, with a frequency separation of 750 Hz and 1 kHz for payload and preamble, respectively. The frequency channels 38, 39, 40 and 41 kHz were dedicated to preamble transmission. The test space is depicted in Fig. 4.27, with all the locations of the transmitters and receivers marked and the transmitted preamble codes are given in the attached table.

Three different scenarios were considered. In the first scenario, transmitter-1 (at the location marked Tx\(_1\)) was set to transmit FHSS signals with preamble-1. The receiver was located at the location marked Rx\(_1\). The purpose of this setup was to keep the receiver far from rigid reflectors to avoid serious effects due to multipath. Each test consisted of the receiver capturing a portion of the signal that is sufficiently long to contain a preamble. The receiver then detected and decoded the preamble therein. The test was repeated 10 times for preamble-1. Thereafter, preamble-1 was replaced by preamble-2 and later by preamble-3, and the test was repeated 10 times for each preamble. In all of the 30 tests, the receiver managed to detect and decode each of the three preambles correctly without any bit error being detected.

In the second scenario, the receiver was moved to the location marked Rx\(_2\) to be close to a rigid wall—the receiver was approximately 49 cm from the wall. Again FHSS signals containing preambles were transmitted from Tx\(_1\) and preambles were detected and decoded. The tests were repeated 15 times for each preamble. The results are summarized in Table 4.1, which reveals that 1-bit errors are dominant, while larger bit errors are quite rare.
In the third scenario, transmitter-1, transmitter-2 and transmitter-3 located at, respectively, the locations marked Tx\textsubscript{1}, Tx\textsubscript{2} and Tx\textsubscript{3}, were allowed to transmit simultaneously. The timing was adjusted so that on reception, preamble-3 was preceded by preamble-1 and preamble-2 directly, with each consecutive pair of the three preambles separated by at most 4 hops (but no preamble overlapping was allowed). The receiver was setup to detect and decode preamble-1. According to the design procedure presented above, this represents the worst acceptable scenario for preamble-1. In 30 tests that were carried out, the result was zero bit error.

Table 4.1: The number of occurrences of different number of bit errors in 15 tests per preamble when the receiver was close to the wall.

<table>
<thead>
<tr>
<th>Preamble</th>
<th>1-bit error</th>
<th>2-bit error</th>
<th>3-bit error</th>
<th>&gt;3-bit errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preamble-1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Preamble-2</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Preamble-3</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

4.6 Summary

This chapter presented an approach for FHSS code acquisition and synchronization. The proposed approach utilized preambles to transmit the transmit-
TERS’ IDs using DBPSK modulation. The chapters discussed the motivations for selecting the latter modulation scheme. A digitally-implemented demodulator was presented that made it possible for the whole system to be realized digitally.

In this chapter, FHSS code acquisition and synchronization were achieved simultaneously—preamble detection and decoding were demonstrated to coincide with coarse and fine synchronization. Preamble detection was based on detecting different frequency channel powers in the DFT domain, while decoding was achieved using cross-correlation of the received signal with the known preambles.

An approach that allows for the detection and decoding of multiple preambles pertaining to multiple transmitters was proposed. The analytical performance of the whole system in terms of the mean squared value of acquisition time was presented and used to evaluate the performance and select the final design parameters.

Finally, experimental results confirming that the proposed method for acquisition and synchronization is practically feasible were presented. A shortcoming was found to be that multipath might have serious effects on performance, in particular, when the receiver is close to a wall. It is therefore recommended that for the receiver device to start up, it should be placed in a good location.
initially such that errors due to multipath are avoided.

A thorough study of the effect of multipath is an important element of the future work for the proposed acquisition and synchronization approach. First, the effect of multipath on preamble detection needs to be understood. Furthermore, the effect of multipath on DBPSK demodulation needs to be investigated. A solution of the multipath problem could be a combination of signal design and signal processing whereby the effect of the impairment is mitigated.
CHAPTER 5

SPATIAL-DIVERSITY-BASED PHASE-DIFFERENCE AMBIGUITY RESOLUTION

This chapter focuses on the phase-difference ambiguity problem due to the wrapping of the phase of the cross power spectrum of the signals received by a pair of widely-spaced sensors. The spatial diversity provided by a third collinear sensor is exploited to resolve the ambiguity in the case of a single-frequency signal. The identifiability condition and the disambiguation algorithm are presented and analyzed under various conditions. Spatial diversity is found useful in resolving ambiguities, however, performance is found to have significant spatial variation, which represents the main drawback of the proposed method.

5.1 Introduction

In this chapter, the phase-difference ambiguity problem is discussed and a solution is proposed. As has been alluded to in Chapter 2, the focus is on phase-difference estimates obtained from the cross power spectrum (CPS) of signals received by two spatially separated sensors. The solutions being sought are for the purpose of estimating the AOAs of FHSS signals. The signals of interest are bandpass in nature and phase-difference estimates obtained from all the frequency bins of a CPS are subject to phase wrapping. Therefore, either phase unwrapping should be performed for the phase-differences obtained from each frequency bin independently, or an approach that is capable of exploiting the frequency diversity of a bandpass signal should be devised.

This chapter deals with the problem of phase-difference ambiguity for the
single-frequency case. Since frequency diversity in the bandpass case of interest is seemingly inadequate (due to the possibility of the occurrence of phase wrapping even in the lowest frequency), the solution sought aims to exploit another form of diversity, namely, spatial diversity. By adding an extra sensor to augment the two-sensor structure conventionally used for time-delay estimation, the problem of phase-difference ambiguity, in the single-frequency case, is found solvable without additional requirements. Such a solution is capable of handling the bandpass signal case of interest, by applying this solution to the phase-difference estimates obtained from each frequency bin independently.

The approach discussed in this chapter, is similar in concept to that reported in [25]. However, the identifiability condition of the proposed approach is flexible such that the proposed solution can tolerate detectable frequency shifts and the sensor configuration can be designed to accommodate the multi-frequency case in a straightforward way.

This chapter is organized as follows. Section 5.2 states the phase-difference ambiguity problem. Section 5.3 derives the condition for the occurrence of the ambiguity and establishes the relationship with the spatial parameters of the signal. Section 5.4 presents a general formulation for solving the ambiguity problem. Section 5.5 details the proposed disambiguation method for the ideal far-field case. Section 5.6 considers the disambiguation problem outside the ideal far-field region. Section 5.7 is devoted to error analysis considering the effect of different sources of error, giving analytic formulae for performance and presenting simulation and analytical results. In Section 5.8, sample experimental results are presented. Finally, Section 5.9 concludes the chapter by giving a summary of the contents.
5.2 THE PHASE-DIFFERENCE AMBIGUITY PROBLEM

Consider two spatially separated sensors. The signals received by the two sensors are

\[ x_1(nT_s) = s(nT_s) + w_1(nT_s) \]  
\[ x_2(nT_s) = s(nT_s - \tau) + w_2(nT_s) \]  

where \( s(nT_s) \) (\( n = 0, 1, 2, \ldots \) and \( T_s \) is the sampling interval) is a sampled and scaled version of a transmitted (which is in the general) multi-frequency signal \( s(t) \); \( w_1(nT_s) \) and \( w_2(nT_s) \) are samples of two added white Gaussian noise (AWGN) processes; and \( \tau \) is the delay of the received signal between the two sensors. For simplicity, in Eq. (5.1) and Eq. (5.2), it is assumed that the signals received by the two sensors are of equal energy, an assumption that can be justified by the fact that the sensor separation is small compared to the distance between the sensors and the transmitter.

The CPS of \( x_1(nT_s) \) and \( x_2(nT_s) \) can be estimated by dividing each of \( x_1(nT_s) \) and \( x_2(nT_s) \) into a number \( (N_f) \) of (possibly overlapping) frames and estimating the complex CPS from these frames according to [95]

\[ \hat{G}_{x_1x_2}(\omega) = \frac{1}{N_f} \sum_{i=0}^{N_f-1} X_{1i}(\omega) X_{2i}^*(\omega) \]  

where \( \omega \) is the radian frequency which is assumed to be discrete; \( X_{1i}(\omega) \) and \( X_{2i}(\omega) \) are the discrete Fourier transforms (DFTs) of the \( i \)'th frames of \( x_1(nT_s) \) and \( x_2(nT_s) \), respectively, each frame is multiplied by an appropriate window function; and \( "*" \) denotes the complex conjugate operation.
The CPS $\hat{G}_{x_1x_2}(\omega)$ can be related to that of the transmitted signal by [20]

$$\hat{G}_{x_1x_2}(\omega) \approx \hat{G}_{ss}(\omega)e^{j\omega\tau + \nu}$$  \hspace{1cm} (5.4)

where $\hat{G}_{ss}(\omega)$ is an estimate of the real power spectrum of $s(nT_s)$, and $\nu$ is a phase error due to the effect of noise, finite data record, etc. For the sake of simplicity, we will only consider the contribution of (AWGN) noise.

The phase-difference pertaining to the two received signals at each frequency can be estimated from the phase of the CPS according to

$$\hat{\phi}(\omega) = \arg[\hat{G}_{x_1x_2}(\omega)] = \omega\tau + \nu$$  \hspace{1cm} (5.5)

where $\arg(.)$ denotes the angle of a complex quantity. The phase in Eq. (5.5) can be expressed in the form

$$\hat{\phi} = (\varphi + \epsilon) + 2\pi k$$  \hspace{1cm} (5.6)

where $k \in \mathbb{Z}$ is a parameter that represents the number of phase cycles in $\hat{\phi}$ taken to the closest integer number of cycles; $(\varphi + \epsilon) \in [-\pi, \pi]$ is a noisy principal phase component with $\varphi$ being the true principal phase and $\epsilon$ representing the contribution of noise. The parameter $k$ indicates the possibility of the occurrence of phase wrapping; when $k$ takes any value other than zero, phase wrapping will occur.

Naturally, phase-difference estimation using (5.5) yields only the noisy principal phase component $(\varphi + \epsilon)$, while the value of the parameter $k$ remains unknown. In other words, the phase-difference estimation process does not reveal any information about the number of complete cycles in $\hat{\phi}$; it reveals only a subcomponent that is confined to the interval $[-\pi, \pi]$. Therefore, an ambiguity may occur such that mapping from the principal phase domain to the true phase domain is a one-to-many operation. To find the correct mapping for a particular case, the value of $k$ needs to be identified.

It should be noted here that, in some cases, the effect of noise could result in erroneous values for $k$ due to cycle slips. The effect of noise in introducing cycle slips depends on the value of $\varphi$ and that of $\nu$ in (5.5). In most cases, the effect is found to be insignificant for the signal-to-noise ratio (SNR) range of interest. However, for some particular scenarios, it is found that the effect can be serious. These scenarios will be excluded from the discussion in this chapter, however, the problem will be revisited in Chapter 7. When no cycle
slipping occurs, \( \nu \) and \( \epsilon \) in (5.6) will be the same process, as is assumed in the following discussion.

### 5.3 Condition for the Occurrence of Ambiguity

The occurrence of phase wrapping is directly linked to the spatial parameters of the received signal, nominally, its AOA. To demonstrate this, let us assume a source in the far-field of two sensors (such as sensor 1 and sensor 2 in Fig. 5.1) and noise-free scenario. The phase-difference in (5.5) can be expressed as

\[
\phi = 2\pi f \tau = \pi D \sin(\theta) \quad (5.7)
\]

where \( 2\pi f = \omega \), with \( f \) being the discrete frequency in Hertz; \( \theta \) is the AOA of the received signal; and \( D \) is the distance between the two sensors in half-wavelengths. From (5.7) and noting that phase is naturally restricted to the interval \([-\pi, \pi]\), the necessary and sufficient condition to avoid phase wrapping can be stated as

\[
D \leq \frac{1}{\left| \sin(\theta) \right|}. \quad (5.8)
\]

In the general case, the signal can be received from any direction, and the general condition to avoid phase wrapping can be obtained from (5.8) by setting \( \sin(\theta) \) equal to unity. Hence, the latter condition becomes

\[
D \leq 1. \quad (5.9)
\]

According to (5.7), a system that does not satisfy (5.9) can still undergo no phase wrapping if the AOA is restricted to the interval

\[
\Theta = \left[ -\arcsin\left(\frac{1}{D}\right), \arcsin\left(\frac{1}{D}\right) \right] \quad (5.10)
\]

where \( \Theta \) contains all the directions that do not undergo phase wrapping, and is referred to as the wrapping-free region. On the other hand, \( \Theta' \), the remaining space outside \( \Theta \), is referred to as the wrapping region.

The results summarized in Eqs. (5.7–5.10) is approximately applicable in the case of a source in the near-field, or the near far-field of two sensors. Since the relative distance between a transmitter and the receiver, in the case of a
Figure 5.3: Deviations (in degrees) of AOA calculated based on Eq. (5.7) from the true emitter angle viewed from a) sensor 1; b) sensor 2; c) the center of the two sensor baseline. Sensor separation is 20 mm with the sensor baseline aligned along the x-axis and centered at the origin.
location system, varies significantly, all possible scenarios have to be consid- 
red. Namely, the interest herein is in the far-field and the near far-field cases. 
Therefore, to cover these cases, the general model depicted in Fig. 5.2 should 
be utilized. This model can be used in all cases—the far-field can be viewed as 
a special case for which $\theta_1 = \theta_2 = \theta$, where $\theta_1$ and $\theta_2$ are the angles viewed 
from the centers of the two sensor baselines.

By considering the angle from the center of the two sensors baseline as the 
AOA of interest, it is found that the relationship in (5.7) is fairly accurate even 
in the near-field case. Fig. 5.3 (a, b and c) shows the deviation of the AOA, cal-
culated based on (5.7) ($\theta = \arcsin(2f \tau/D)$), from the true AOA viewed from 
sensor 1, sensor 2 and the mid-point between sensor 1 and sensor 2, respecti-
vely for each subfigure. It can be seen that the error for the case that coincides 
with the center of the sensor baseline as a point of view is less than 0.1° for 
the whole 12 m × 12 m space of consideration, except for a tiny region that 
is close to the point of view and that is not of interest for most localization 
applications. It will therefore be assumed that Eqs. (5.7 – 5.10) are sufficiently 
accurate for all the cases of interest.

It can be concluded that for any two sensors with full angle scanning range, 
the sufficient condition for ambiguities to arise at a particular frequency is that 
the sensor separation exceeds a half-wavelength of that frequency. Note that 
the occurrence of ambiguities does not imply the occurrence of phase wrap-
ning, but the possibility of the occurrence of phase wrapping. In other words, 
the possibility of the occurrence of phase wrapping with different number of 
cycles results in an unidentifiability or ambiguity problem.

5.4 General Formulation of the Disambiguation

Method

In this section, we describe a general formulation for using a sensor triplet to 
achieve phase-difference disambiguation. We will first adopt the general mo-
del depicted in Fig. 5.2. Based on this model, the generic formulae required for 
the problem solution will be established. In the subsequent section, disambi-
guation in what we call the ideal far-field case will be approached as a special 
case of the general model. In a later section, disambiguation in scenarios that 
do not satisfy the ideal far-field requirement will be discussed.
First, the signal model for three sensors is
\[
x_1(nT_s) = s(nT_s) + w_1(nT_s) \\
x_2(nT_s) = s(nT_s - \tau_{12}) + w_2(nT_s) \\
x_3(nT_s) = s(nT_s - \tau_{12} - \tau_{23}) + w_3(nT_s)
\] (5.11)

where \(\tau_{12}\) is the signal delay between sensor 1 and sensor 2; \(\tau_{23}\) is the signal delay between sensor 2 and sensor 3; and \(w_1(nT_s), w_2(nT_s)\) and \(w_3(nT_s)\) are samples of three AWGN processes.

Using the above signal model, the CPS of \(x_1(nT_s)\) and \(x_2(nT_s)\), \(\hat{G}_{x_1x_2}\), can be obtained from Eq. (5.3), and \(\hat{G}_{x_2x_3}\) can be obtained in the same way. For simplicity of exposition, and due to the linear relationship between phase-difference and delay, we will formulate the problem in terms of delays rather than phase-differences, and the focus will be to find the true delay observed between a sensor pair.

Based on Eq. (5.6), we obtain the following pair of equations:
\[
\hat{\tau}_{12} = (\delta_{12} + \varepsilon_{12}) + \frac{k_{12}}{f}, k_{12} \in \mathbb{Z} \\
\hat{\tau}_{23} = (\delta_{23} + \varepsilon_{23}) + \frac{k_{23}}{f}, k_{23} \in \mathbb{Z}
\] (5.12) (5.13)

where \(\hat{\tau}_{12} \in (-D_{12}/2f, D_{12}/2f)\) and \(\hat{\tau}_{23} \in (-D_{23}/2f, D_{23}/2f)\) are estimates of the true unknown delays, with \(D_{12} = 2d_{12}/\lambda\) and \(D_{23} = 2d_{23}/\lambda\) represent the inter-sensor spacings in half-wavelengths; \{\((\delta_{12} + \varepsilon_{12}), (\delta_{23} + \varepsilon_{23})\) \in [-1/2f, 1/2f]\) are the corresponding noisy principal delays whose values are known; \(\varepsilon_{12}\) and \(\varepsilon_{23}\) represent delay estimation errors due to noise. Where cycle slipping due to noise is ignored, the statistical properties of these errors will be similar to those of the random variable \(\nu/2\pi f\), where \(\nu\) is the phase error in Eq. (5.5). Note that the intervals to which \(\hat{\tau}_{12}\) and \(\hat{\tau}_{23}\) have been confined are open since we ignore the trivial cases of \(\theta_1 = \pm 90^\circ\) and \(\theta_2 = \pm 90^\circ\). The goal of disambiguating is to find the value of the integer \(k_{12}\), or, equivalently, the delay estimate \(\hat{\tau}_{12}\), exploiting the additional information obtained from the third sensor. This is equivalent to the phase unwrapping problem that has initially been discussed. To establish the method, first a noise-free model will be assumed. The effect of noise will be considered in Section 5.7.
5.4 GENERAL FORMULATION OF THE DISAMBIGUATION METHOD

By setting the noise contributions in Eq. (5.12) and Eq. (5.13) to zero, we obtain

\[ \tau_{12} = \delta_{12} + \frac{k_{12}}{f}, k_{12} \in \mathbb{Z} \]  

(5.14)

\[ \tau_{23} = \delta_{23} + \frac{k_{23}}{f}, k_{23} \in \mathbb{Z} \]  

(5.15)

where \( \tau_{12} \in (-D_{12}/2f, D_{12}/2f) \) and \( \tau_{23} \in (-D_{23}/2f, D_{23}/2f) \) are the true unknown delays; \( \{\delta_{12}, \delta_{23}\} \subset [-1/2f, 1/2f] \) are the corresponding principal delays. The problem is now to find \( k_{12} \), or equivalently, \( \tau_{12} \) given \( \delta_{12} \) and \( \delta_{23} \).

As has been mentioned previously, the AOA varies depending on the point of observation on the sensor baseline. However, it was shown that Eq. (5.7) provides an accurate value for the AOA observed from the mid-point of the baseline between two sensors. Based on this and using (5.7), we can obtain

\[ \sin(\theta_1) \approx 2f \frac{\tau_{12}}{D_{12}} \]

\[ \sin(\theta_2) \approx 2f \frac{\tau_{23}}{D_{23}} \]  

(5.16)

where \( \theta_1 \) and \( \theta_2 \) are the AOA observed from the centers of the two baselines as in Fig. 5.2. The two angles, \( \theta_1 \) and \( \theta_2 \), are generally different and the difference varies depending on the emitter location. Generally speaking, the values of \( \theta_1 \) and \( \theta_2 \) are very close to each other when the emitter is sufficiently far from the sensors. From (5.16), the following relationship can be established:

\[ \frac{\tau_{23}}{\tau_{12}} = \frac{\sin(\theta_2)}{\sin(\theta_1)} \frac{D_{23}}{D_{12}} = \rho \frac{D_{23}}{D_{12}} \]  

(5.17)

where \( \rho \triangleq \sin(\theta_2) / \sin(\theta_1) \) is introduced to parameterize the effect of the geometry due to the emitter location.

Now, using (5.17) to eliminate \( \tau_{23} \) from (5.15), subtracting (5.14) from the result and manipulating yields

\[ \tau_{12} = \mu_\rho \left( \delta_{23} - \delta_{12} + \frac{k_{23} - k_{12}}{f} \right) \]  

(5.18)

where

\[ \mu_\rho = \left( \frac{D_{12}}{\rho D_{23} - D_{12}} \right). \]  

(5.19)

Since the difference \( k_{23} - k_{12} \) is unknown, many candidate values for \( \tau_{12} \) can be calculated assuming different values for the difference \( k_{23} - k_{12} \). Consequently,
(5.18) can be written as

\[ \tau_{12}(k) = \mu_p \left( \delta_{23} - \delta_{12} + \frac{k}{f} \right), \quad k \in \mathbb{Z} \quad (5.20) \]

where \( k \) is an unknown integer that corresponds to the difference \( k_{23} - k_{12} \), and \( \tau_{12}(k) \) are candidate values for \( \tau_{12} \). In (5.20), the disambiguation problem has been converted into a problem of finding the true value of the integer \( k \) (denoted as \( k_o \)). To find that value, we propose restricting the sensor configuration in such a way that \( k_o \) produces a unique, and hence identifiable, delay (\( \tau_{12}(k_o) \)). For \( \tau_{12}(k_o) \) to be unique, a criterion could be that it is the only candidate delay that falls in the interval \((-D_{12}/2f, D_{12}/2f)\) which corresponds to the visible AOA range, i.e.,

\[ \exists! \tau_{12}(k = k_o) \in \left( \frac{-D_{12}}{2f}, \frac{D_{12}}{2f} \right), \quad k \in \mathbb{Z}. \quad (5.21) \]

Eq. (5.21) will be used as the criterion for identifying the true delay exploiting a sensor configuration that will be revealed in the subsequent discussion.

Now, consider the true candidate delay \( \tau_{12}(k_o) \). Any other false candidate value of the delay \( \tau_{12}(k_o \pm q), q \in \mathbb{N} \) can be expressed, based on (5.20), as

\[ \tau_{12}(k_o \pm q) = \tau_{12}(k_o) \pm q \frac{\mu_p}{f} \quad (5.22) \]

From (5.21), it follows that the false delays in (5.22) should satisfy

\[ \left| \tau_{12}(k_o) \pm q \frac{\mu_p}{f} \right| \geq \frac{D_{12}}{2f}, \quad \forall q \quad (5.23) \]

where \(|.|\) denotes the absolute value. Considering all possible sign (+/-) combinations of \( \tau_{12}(k_o) \) and \( \mu_p \), a sufficient and necessary condition for the true phase to be uniquely identifiable can be stated as

\[ |\mu_p| \geq \frac{D_{12}}{2} + |f \tau_{12}(k_o)|. \quad (5.24) \]

Substituting for \( \tau_{12}(k_o) = D_{12} \sin(\theta_1)/2f \) and for \( \mu_p \) from (5.19), dividing both sides of (5.24) by \( D_{12} \) and manipulating, the necessary and sufficient condition can now be stated as

\[ |(\rho - 1)D_{12} + \rho \Delta_d| \leq \frac{2}{1 + |\sin(\theta_1)|} \quad (5.25) \]
where $\Delta_d \triangleq D_{23} - D_{12}$. A sufficient but not necessary condition can be derived from (5.25) by setting $\sin(\theta_1) = 1$, i.e.,

$$|(\rho - 1)D_{12} + \rho \Delta_d| \leq 1 \quad (5.26)$$

It is clear from (5.25) and (5.26) that the identifiability depends on the emitter location as reflected by the parameters $\rho$ and $\theta_1$. This is impractical since the emitter location is unknown.

5.5 Disambiguation in the Ideal Far-Field

We define the ideal far-field as any situation for which $\rho$ approaches the value ‘1’.

**Theorem 5.1** For a system of three collinear sensors in the ideal far-field and under noise-free conditions, a sufficient condition for the identifiability of the true delay observed between any pair of sensors is that the difference of the two smaller inter-sensor spacings be no greater than a half-wavelength of the impinging signal.

The theorem does not assume any upper limit on the inter-sensor distances. Arbitrary spacing is allowed so long as the ideal far-field assumption is valid.

**Proof.** The proof of Theorem 5.1 follows directly from (5.26) by substituting $\rho = 1$, which results in the condition

$$|\Delta_d| \leq 1. \quad (5.27)$$

Based on the condition in (5.27), the true value of $k$ (denoted as $k_o = k_{23} - k_{12}$) is expected to have a maximum absolute value equivalent to 1. This is due to the fact that the maximum possible absolute phase change over a distance of $\lambda/2$ is equal to $\pi$. Consequently, the search for $k_o$ is confined to the set

$$K = \{-1, 0, 1\}. \quad (5.30)$$
Figure 5.4: The values of $\rho$ for $f = 40$ kHz at a propagation speed of 343 m/s and $d_{12}$ set equal to: a) $1.01\lambda/2$; b) $2\lambda/2$; c) $6\lambda/2$; and d) $20\lambda/2$; $\Delta_\rho \approx 0.7\lambda/2$. The sensor baseline is aligned along the x-axis with $d_{12}$ centered at the origin. Resolution is 50 mm.

Hence, for a collinear sensor triplet that satisfies (5.27), the true delay can be identified by evaluating (5.20) for $k \in \mathbf{K} = \{-1, 0, 1\}$ and selecting the value that satisfies (5.21) as the true delay value.

5.6 Disambiguation under Non-Ideal Far-Field Conditions

In this section, we consider scenarios, where the unknown parameter $\rho$ (defined in (5.17)) is not equal to unity. Fig. 5.4 plots the values of $\rho$ for different locations in a $12 \times 12$ m space for different sensor separations. Fig. 5.5 shows a zoom-in around $0^\circ$ with the y-coordinate of the emitter fixed at 5 m. From the two figures, it can be seen that for large areas of the space considered,
even close to the sensors, $\rho$ may remain close to unity (i.e., the ideal far-field assumption is sufficiently satisfied). Nevertheless, $\rho$ can take values that are extremely far from unity close to the $0^\circ$ region. The values that $\rho$ takes are also seen to be dependent both on the emitter location and inter-sensor spacing.

The direct result of these deviations in the value of $\rho$ is that the satisfaction of the necessary condition for disambiguation given by (5.25) cannot be guaranteed. Based on the success or failure of the disambiguation condition, we introduce the following definition:

**Definition 1.** For a triplet of collinear sensors and a source that does not fall in the ideal far-field region, the geometrically adequate region (GAR) is defined as the portion of space for which (5.25) holds.

**Definition 2.** For a triplet of collinear sensors and a source that does not fall in the ideal far-field, the geometrically inadequate Region (GIR) is defined as the portion of space for which (5.25) does not hold.

The notion of these definitions is that, in the context of disambiguation away from the ideal far-field, certain locations can approximately be treated as if they are ideal far-field locations. Other locations are different and do not satisfy the identifiability condition and should be looked at in a different way in terms of disambiguation. Fig. 5.6 plots the GAR and the GIR for different sensor configurations. It can be seen that the width of the GIR is proportional to the inter-sensor spacing. Comparing Fig. 5.4 and Fig. 5.6, the relationship between the success/failure of the uniqueness condition and the value of $\rho$ is obvious.
Figure 5.6: The GAR (red) and the GIR (blue) for $f = 40$ kHz at a propagation speed of 343 m/s and $d_{12}$ set equal to: a) $1.01\lambda/2$; b) $2\lambda/2$; c) $6\lambda/2$; and d) $20\lambda/2$; $\Delta_d \approx 0.7\lambda/2$. The sensor baseline is aligned along the x-axis with $d_{12}$ centered at the origin.

Due to the failure of the disambiguation condition in the GIR and the unavailability of the value of $\rho$ to identify whether the emission is from the GAR or the GIR, uniqueness cannot directly be used as a criterion for disambiguation in this case.

5.6.1 Disambiguation in the GAR

Since $\rho$ and hence $\mu_\rho$ are not precisely known, we consider setting $\rho = 1$ and using (5.28) to produce rough estimates for the candidate delays

$$\hat{\tau}_{12}(k) = \mu_1 \left( \delta_{23} - \delta_{12} + \frac{k}{f} \right), \ k \in \mathbb{Z}. \quad (5.31)$$
Figure 5.7: The wrapping (blue) and wrapping-free (red) regions for $f = 40$ kHz at a propagation speed of 343 m/s and $d_{12}$ set equal to: a) $1.01\lambda/2$; b) $2\lambda/2$; c) $6\lambda/2$; and d) $20\lambda/2$; $\Delta_d \approx 0.7\lambda/2$. The sensor baseline is aligned along the x-axis and centered at the origin.
Figure 5.8: The values of $k$ for $f = 40$ kHz at a propagation speed of 343 m/s and $d_{12}$ set equal to: a) $1.01\lambda/2$; b) $2\lambda/2$; c) $6\lambda/2$; and d) $20\lambda/2$; $\Delta_d \approx 0.7\lambda/2$. The sensor baseline is aligned along the x-axis with $d_{12}$ centered at the origin.
The ratio of the approximate candidate values produced by (5.31) to the true ones from (5.20) recalling (5.19) is given by

\[
G_\rho = \frac{\hat{\tau}_{12}(k)}{\tau_{12}(k)} = \frac{\rho D_{23} - D_{12}}{D_{23} - D_{12}}
\]  

(5.32)

where \(G_\rho\) represents a scale factor that can be applied to Eq. (5.20) to produce (5.31). Noting that the uniqueness of a delay in (5.20) coincides with the fact that the delay has the smallest absolute value amongst the candidate delays, coupled with the fact that the scale factor \(G_\rho\) is constant for all the candidate values, the order of the absolute values of these candidate delays is expected to be maintained in (5.31). Consequently, the scaled version of the true delay will have the minimum absolute value among all the scaled candidate delays resulting from (5.31). This means that an approximate version of the true delay can be estimated as

\[
\hat{\tau}_{12,m} = \hat{\tau}_{12}(\hat{k}_o), \text{ where } \hat{k}_o = \arg \min_k |\hat{\tau}_{12}(k)|, \ k \in K.
\]  

(5.33)

Note that for the GAR, the estimate \(\hat{k}_o\) in (5.33) is error free (i.e, \(\hat{k}_o = k_o\)), as can be deduced from the above discussion.

When \(\rho\) is sufficiently close to unity, (5.33) can produce a fairly accurate estimate of the delay \(\tau_{12}\). On the other hand, (5.33) is expected to diverge when \(\rho\) is far from unity. In general, (5.33) introduces some bias in the estimated delay value. One way to remove this bias is to exploit the fact that the ambiguity is pertaining to an integer value, as reflected in (5.14). The following steps can be carried out to remove the bias. First, the delay value obtained from (5.33) is substituted in (5.14) to obtain an estimate of the integer value \(k_{12}\),

\[
\hat{k}_{12} = \Psi[f(\hat{\tau}_{12,m} - \delta_{12})]
\]  

(5.34)

where \(\Psi(.)\) is the round-off of the value. Next, \(\hat{k}_{12}\) is inserted back into (5.14) to obtain the final estimate of the true delay as

\[
\hat{\tau}_{12} = \delta_{12} + \frac{\hat{k}_{12}}{f}.
\]  

(5.35)
5.6.2 Disambiguation in the GIR

The GIR coincides with the failure of the uniqueness condition on which disambiguation relies. Contemplating Fig. 5.6 in the light of Fig. 5.7, which depicts the wrapping and the wrapping-free regions based on Eq. (5.10) for different inter-sensor separations, it can be seen that for a wide range of inter-sensor spacings, the GIR regions generally coincide with wrapping-free regions. Note that for the largest inter-sensor spacing depicted, this is not precisely true. In such cases, the true delay $\tau_{12}$ is the principal delay $\delta_{12}$. However, this fact needs to be established.

It is noticed that in the wrapping-free case, due to the smallness of the delays, the integers $k_{12}$ and $k_{23}$ are both expected to be equal to zero. Hence, the true value of $k$ is also equal to zero. Fig. 5.8 depicts the values of $k$ for the various configurations. By comparing Fig. 5.6 and Fig. 5.8, it can be seen that generally in the GIR, $k$ is zero, except in Fig. 5.8 (d), where isolated pockets of other values of $k$ also appear in the GIR. It can be concluded that for moderate sensor separation, $k$ is zero in the GIR, and this is what will be assumed in the following discussions.

Now, consider Eq. (5.31) evaluated for $k \in K$. In the GIR case and for moderate sensor separation, a rough estimate of the true delay $\hat{\tau}_{12}(0)$ can be expressed as

$$\hat{\tau}_{12}(0) = \mu_1 (\delta_{23} - \delta_{12}) = \mu_1 \delta$$ (5.36)

where $\delta$ is a small value due to $\delta_{12}$ and $\delta_{23}$ being small. The other two false images can be expressed as

$$\hat{\tau}_{12}(-1) = \hat{\tau}_{12}(0) - \frac{\mu_1}{f} = \hat{\tau}_{12}(0) - \frac{D_{12}}{f \Delta_d}$$

$$\hat{\tau}_{12}(1) = \hat{\tau}_{12}(0) + \frac{\mu_1}{f} = \hat{\tau}_{12}(0) + \frac{D_{12}}{f \Delta_d}.$$ (5.37)

Note that the level of approximation in (5.36) and (5.37) can be excessively large since $\rho$ may take values that are extremely far from unity in the GIR (see Fig. 5.4 and Fig. 5.5). Since $\hat{\tau}_{12}(0)$ is very small, it is anticipated to satisfy $\hat{\tau}_{12}(0) \ll \frac{D_{12}}{2f}$ to a sufficient degree. Therefore, for $\Delta_d \leq 1$, both $\hat{\tau}_{12}(-1)$ and $\hat{\tau}_{12}(1)$ will be expected to fall outside the interval $(-D_{12}/2f, D_{12}/2f)$. Consequently, the approximate estimate of the true delay is unique in the sense revealed in (5.21). Note that the uniqueness in this case is true for the approximate estimates of the candidate delays not the true candidate ones (in (5.20))
for which the uniqueness condition is not satisfied in the GIR. This uniqueness can hence be used to identify the approximate true delay. Since uniqueness and having the smallest absolute value are equivalent, Eq. (5.33) can again be used to obtain an approximate version of the true delay. Further, Eq. (5.34) and Eq. (5.35) can be used to remove the bias.

It is noticed that, in the GIR, $|\mu_1 \delta| < |\mu_\mu \delta|$, where $\mu_\mu \delta$ represents the true delay from (5.20). Therefore, it is anticipated that Eq. (5.34) will be able to restore the correct value of $k_{12}$ (presumably zero in this case). However, failure is also expected in some regions in space due to the inaccuracy of the underlying assumptions.

5.6.3 Summary of the Disambiguation Algorithm

The proposed disambiguation algorithm can be summarized as follows. For three collinear sensors that satisfy the condition $\Delta_d \leq 1$, given the ambiguous delays $\delta_{12}$ and $\delta_{23}$, the true delay $\tau_{12}$ can be identified using the procedure:

I. Calculate the rough candidate estimates of the true delay $\hat{\tau}_{12}(k)$ using (5.31) for $k \in \{-1, 0, 1\}$.

II. Obtain the first estimate ($\hat{\tau}_{12,m}$) for the true delay using (5.33).

III. Use $\hat{\tau}_{12,m}$ to obtain an estimate for $k_{12}$ using (5.34).

IV. Obtain the final estimate of the true delay from (5.35).

Step II of the disambiguation algorithm will be referred to as the minimum absolute value (MAV) step, and step III will be denoted as the round-off (RO) step. The algorithm above can equally be used in both cases, when the ideal far-field assumption is satisfied and when it is not. However, for some situations when the latter assumption is not satisfied, the approximations involved may cause the algorithm to fail in identifying the true delay. Such effects, together with the noise effect, will be considered in the subsequent section.

5.7 Error Analysis

In this section, sources of error in the disambiguation process are considered. Errors are expected to arise primarily due to the approximate nature of the
CHAPTER 5: SPATIAL-DIVERSITY-BASED PHASE-DIFFERENCE AMBIGUITY ...

disambiguation approach when the ideal far-field assumption is not satisfied. The deviation of the value of $\rho$ from unity is expected to result in some inaccuracies at least in some regions in space. The other source of error for the proposed disambiguation method is noise, which is present in all practical systems. The question that arises is: how well does the proposed disambiguation algorithm—that has primarily been derived based on a noise-free data model—perform in the presence of noise?

Using Eq. (5.12) and Eq. (5.13), the noisy counterpart of Eq. (5.31), can be obtained as

$$
\hat{\tau}_{12,\varepsilon}(k) = \mu_1 \left( \delta_{23} - \delta_{12} + \frac{k}{f} \right) + \mu_1 (\varepsilon_{23} - \varepsilon_{12}) \\
= \tau_{12}(k) + \nu_k + \varepsilon, \ k \in \mathbb{Z}
$$

(5.38)

where $\varepsilon \triangleq \mu_1 (\varepsilon_{23} - \varepsilon_{12})$, $\nu_k$ is a $k$-value dependent bias due to model mismatch and is defined as

$$
\nu_k \triangleq (\mu_1 - \mu_p) \left( \delta_{23} - \delta_{12} + \frac{k}{f} \right).
$$

(5.39)

In the presence of noise, $\hat{\tau}_{12,\varepsilon}(k)$ will represent the candidate delay estimates that are calculated in step I of the proposed disambiguation algorithm. In the following subsections, error analysis will consider the next three steps of the disambiguation algorithm with $\hat{\tau}_{12,\varepsilon}(k)$ being the input to step II. The error terms, $\nu_k$ and $\varepsilon$, in (5.38) determine the probability of success/failure of the disambiguation process and the error in the final delay estimate.

Now, without loss of generality, assume that $\varepsilon_{12}$ and $\varepsilon_{23}$ are two i.i.d. (independent and identically distributed) random variables with normal distributions, a zero mean, and a variance $\sigma_p^2$. The assumption will be justified later for the CPS case. From basic probability theory, $\varepsilon$ will also be a zero-mean process with variance $\sigma^2 = 2\mu_1^2 \sigma_p^2$ and cumulative distribution function (CDF)

$$
\Phi_\varepsilon(x) = \frac{1}{2} \left[ 1 + \text{erf} \left( \frac{x}{\sigma \sqrt{2}} \right) \right], \ x \in \mathbb{R}
$$

(5.40)

where $\text{erf}(\cdot)$ is the so-called error function. Subsequently, the CDF for any Gaussian random variable with a zero mean will be obtained from (5.40) by substituting $\sigma$ equal to the standard deviation of the new random variable, and changing the subscript $\varepsilon$ as appropriate.
5.7 ERROR ANALYSIS

5.7.1 Error Analysis for MAV Step

The probability of success of the MAV step can be defined as the probability of picking the estimate that corresponds to the right value of \( k \), i.e., \( k_o \) by applying (5.33) to the data given in (5.12) and (5.13). This probability can be expressed as

\[
P_{s,m} = P\left( |\hat{\tau}_{12,e}(k_o)| < |\hat{\tau}_{12,e}(k_1)| \right) \land \left( |\hat{\tau}_{12,e}(k_o)| < |\hat{\tau}_{12,e}(k_2)| \right)
\]

\[
= P\left( |\tau_{12}(k_o) + \nu_{k_o} + \epsilon| < |\tau_{12}(k_1) + \nu_{k_1} + \epsilon| \right) \land \left( |\tau_{12}(k_o) + \nu_{k_o} + \epsilon| < |\tau_{12}(k_2) + \nu_{k_2} + \epsilon| \right)
\]

(5.41)

where \( k_1 \) and \( k_2 \) are the values of \( k \) that correspond to the false delays; and “\&” denotes the logical AND operation. The probability \( P_{s,m} \) is dependent on different variables all of which, excluding \( \epsilon \), depend on the emitter location. A procedure for evaluating this probability for different cases is presented in Appendix A.1.

The probability of failure of the MAV step \( P_{f,m} = 1 - P_{s,m} = P_{k_1} + P_{k_2} \) is the probability that any of the two false delays is obtained as an outcome of the MAV step. The definitions of \( P_{k_i}, i = 1, 2 \) can be obtained from (5.41) by swapping the role of \( \tau_{12}(k_o) + \nu_{k_o} \) and \( \tau_{12}(k_i) + \nu_{k_i} \). The procedure for evaluating the probabilities \( P_{k_1} \) and \( P_{k_2} \) is similar to that of \( P_{s,m} \) and can also be found in Appendix A.1.

Based on the probabilities \( P_{s,m}, P_{k_1} \) and \( P_{k_2} \), the mean squared error (MSE) for the delay obtained from the MAV step can be expressed in the form

\[
\varepsilon^2_m = \Gamma_{s,m}P_{s,m} + \Gamma_{k_1,m}P_{k_1} + \Gamma_{k_2,m}P_{k_2}
\]

(5.42)

where \( \Gamma_{s,m}, \Gamma_{k_1,m} \) and \( \Gamma_{k_2,m} \) are, respectively, the MSEs associated with the three cases representing success and failure, and can be calculated

\[
\Gamma_{s,m} = 2\mu_1^2\sigma_p^2 + \nu_{k_0}^2
\]

\[
\Gamma_{k_1,m} = 2\mu_1^2\sigma_p^2 + \left[ \tau_{12}(k_1) + \nu_{k_1} - \tau_{12}(k_o) \right]^2
\]

\[
\Gamma_{k_2,m} = 2\mu_1^2\sigma_p^2 + \left[ \tau_{12}(k_2) + \nu_{k_2} - \tau_{12}(k_o) \right]^2.
\]

(5.43)

Note that each of the terms \( \Gamma_{s,m}, \Gamma_{k_1,m} \) and \( \Gamma_{k_2,m} \) contains an error term that corresponds to a random variable, and an error term that corresponds to a bias.
5.7.2 Error Analysis for the RO Step

The input to the RO step is the output of the MAV step, which has the form given in Eq. (5.38), with \( k \) value depending on the success/failure status. The probability of success of the RO step and hence the whole disambiguation method is the probability that the correct integer \( k_{12} \) is restored, i.e., \( P_s = P(\hat{k}_{12} = k_{12}) \). This can be expanded based on Eq. (5.34) to

\[
P_s = P \{ \Psi \{ f \left[ \delta_{12} + \varepsilon_{12} \right] = k_{12} \} \}.
\]

(5.44)

Realizing the fact that \( \hat{\tau}_{12,m} \) may take one of three different values, each associated with a generally different probability, \( P_s \) can be expressed as

\[
P_s = P \{ \Psi \{ f \left[ \tau_{12}(k_o) + \nu_{k_o} + \varepsilon - (\delta_{12} + \varepsilon_{12}) \right] = k_{12} \} \} \ P_{s,m}
+ P \{ \Psi \{ f \left[ \tau_{12}(k_1) + \nu_{k_1} + \varepsilon - (\delta_{12} + \varepsilon_{12}) \right] = k_{12} \} \} \ P_{k_1}
+ P \{ \Psi \{ f \left[ \tau_{12}(k_2) + \nu_{k_2} + \varepsilon - (\delta_{12} + \varepsilon_{12}) \right] = k_{12} \} \} \ P_{k_2}.
\]

(5.45)

By substituting for \( \tau_{12}(k_i) = \tau_{12}(k_o) + q_i \mu_p / f, i = 1, 2, \) as similar to Eq. (5.22), but here \( q_i \) is a positive or negative integer value equal to \( k_i - k_o \); and for \( f[\tau_{12}(k_o) - \delta_{12}] = k_{12} \) and manipulating; we obtain

\[
P_s = P \{ \Psi \left[ k_{12} + f(\varepsilon - \varepsilon_{12}) + f \nu_{k_o} \right] = k_{12} \} \ P_{s,m}
+ P \{ \Psi \left[ k_{12} + f(\varepsilon - \varepsilon_{12}) + f \nu_{k_1} + q_1 \mu_p \right] = k_{12} \} \ P_{k_1}
+ P \{ \Psi \left[ k_{12} + f(\varepsilon - \varepsilon_{12}) + f \nu_{k_2} + q_2 \mu_p \right] = k_{12} \} \ P_{k_2}
\]

(5.46)

or equivalently

\[
P_s = P \left[ -0.5 \leq f(\varepsilon - \varepsilon_{12}) + f \nu_{k_o} < 0.5 \right] \ P_{s,m}
+ P \left[ -0.5 \leq f(\varepsilon - \varepsilon_{12}) + f \nu_{k_1} + q_1 \mu_p < 0.5 \right] \ P_{k_1}
+ P \left[ -0.5 \leq f(\varepsilon - \varepsilon_{12}) + f \nu_{k_2} + q_2 \mu_p < 0.5 \right] \ P_{k_2}.
\]

(5.47)
Eq. (5.47) can finally be converted to the form

\[
P_s = \left[ \Phi_{\varepsilon_\text{to} - \varepsilon_\text{to}} \left( \frac{0.5}{f} + \nu_k \right) - \Phi_{\varepsilon_\text{to} - \varepsilon_\text{to}} \left( \frac{-0.5}{f} + \nu_k \right) \right] P_{s,m} \\
+ \left[ \Phi_{\varepsilon_\text{to} - \varepsilon_\text{to}} \left( \frac{0.5 + q_1 \mu_{\rho}}{f} + \nu_k \right) - \Phi_{\varepsilon_\text{to} - \varepsilon_\text{to}} \left( \frac{-0.5 + q_1 \mu_{\rho}}{f} + \nu_k \right) \right] P_{k_1} \\
+ \left[ \Phi_{\varepsilon_\text{to} - \varepsilon_\text{to}} \left( \frac{0.5 + q_2 \mu_{\rho}}{f} + \nu_k \right) - \Phi_{\varepsilon_\text{to} - \varepsilon_\text{to}} \left( \frac{-0.5 + q_2 \mu_{\rho}}{f} + \nu_k \right) \right] P_{k_2} 
\]

(5.48)

where \( \Phi_{\varepsilon_\text{to} - \varepsilon_\text{to}} \) is the CDF of the random variable \( \varepsilon - \varepsilon_\text{to} \), which has a Gaussian distribution with a zero mean and a variance equal to \( \sigma^2 + \sigma_{\rho}\). The MSE of the final delay obtained from step IV of the disambiguation algorithm takes the form

\[
\zeta^2 = \Gamma_3 P_s + \Gamma_f (1 - P_s) 
\]

(5.49)

where \( \Gamma_3 = \sigma_{\rho}^2 \) is the MSE contribution from the success cases, while an approximation for the contribution from the failure cases, \( \Gamma_f \), is derived in Appendix A.2.

### 5.7.3 Noise-Free Success/Failure

In the noise-free case, the only source of error is due to the emitter location. In this section we evaluate the performance of the MAV and the RO steps in the noise-free case. In the absence of noise, the expressions in Eqs. (5.41) and (5.44) loose their probabilistic nature and turn into binary expressions that give the success/failure (1/0) of the corresponding algorithm. The same result can be obtained form Appendix A.1 and Eq. (5.48) by setting the variance \( \sigma_{\rho} \) equal to zero. Also, the MSEs in (5.42) and (5.49), in the absence of noise, will reflect only the effect of the bias due to deviation from the ideal far-field assumption. The noise-free performance, and in particular, the noise-free success/failure is important since it implies where (in space) the proposed methods can or cannot be applied.

Fig. 5.9 and Fig. 5.10 depict the noise-free success/failure for the MAV and the RO steps, respectively. The figures show a general trend of an increase in the size of the failure area as the inter-sensor spacing increases. The failure regions appear to center around the excessive \( \rho \) values regions (see Fig. 5.4), yet they do not exactly coincide. It can be seen that the MAV step exhibits
smaller failure regions compared to the RO step. It can also be seen that the failure region of the MAV step is always contained inside the failure region of the RO step. The failure region of the RO step may dominate the space in some cases, e.g., Fig. 5.10 (c). The plots underscore the effect of the deviation from the ideal far-field assumption on the success/failure of the proposed disambiguation method. They clearly show that this deviation sets an upper limit on the inter-sensor spacing that can be used in a particular scenario. The figures also give guidelines on the choice of a particular sensor configuration for a particular scenario such that failure is avoided. In this regard, the RO step is more limiting since it generally has a larger failure area. However, as it will be shown subsequently, the inclusion of the RO step in the disambiguation algorithm can significantly improve performance in terms of the MSE, when the emitter is not located in the failure region.

### 5.7.4 Success/Failure under Noise

Fig. 5.11 and Fig. 5.12 plot examples of the probability of success of the MAV and the RO steps using Eq. (5.41) and Eq. (5.44), respectively. The figures show the probability of success for \( \frac{d_{12}}{\lambda} = 20 \) and \( \frac{d_{23}}{\lambda} = 2 \) for different values of \( \sigma_p^2 \). A general trend of a decrease in the probability of success as noise increases can be seen. However, in some cases, the noise may have the effect of improving the probability of success relative to the noise-free case (e.g., in some parts of the regions that were failure regions under noise-free conditions). This can be explained by that the noise effect may counter the effect of the error due to ideal far-field assumption in these particular cases.

### 5.7.5 Performance Versus SNR

To obtain the performance directly against SNR, the variance \( \sigma_p^2 \) needs to be obtained as a function of SNR. For a deterministic signal (which is the signal type of interest for this work), this will be derived in this subsection.

In general, for a large number of frames (\( N_f \)), the error associated with the phase at the frequency \( \omega, \nu \) (see Eq. (5.5)), is approximately Gaussian [20, 22]. Assuming no cycle slipping will take place, both \( \varepsilon_{12} \) and \( \varepsilon_{23} \) will also be approximately Gaussian and the corresponding variance \( \sigma_p^2 \) can be approximated as [20, 22]
Figure 5.9: The success (red) and failure (blue) of the MAV algorithm under no noise for $f = 40$ kHz at a propagation speed of 343 m/s, and $d_{12}$ set equal to: a) $1.01\lambda/2$; b) $2\lambda/2$; c) $6\lambda/2$; and d) $20\lambda/2$; $\Delta_d \approx 0.7\lambda/2$. The sensor baseline is aligned along the x-axis with $d_{12}$ centered at the origin.
Figure 5.10: The success (red) and failure (blue) of the RO step under no noise for $f = 40$ kHz at a propagation speed of 343 m/s and $d_{12}$ set equal to: a) $\lambda/2$; b) $2\lambda/2$; c) $6\lambda/2$; and d) $20\lambda/2$; $\Delta_{d} \approx 0.7\lambda/2$. The sensor baseline is aligned along the x-axis with $d_{12}$ centered at the origin.
Figure 5.11: The probability of success for the MAV algorithm for $f = 40$ kHz at a propagation speed of 343 m/s and $d_{12}$ set equal to $20\lambda/2$; $\Delta_d \approx 0.7\lambda/2$: a) $\sigma_P^2 = -150$ dB; b) $\sigma_P^2 = -130$ dB; c) $\sigma_P^2 = -110$ dB. The sensor baseline is aligned along the x-axis with $d_{12}$ centered at the origin.
Figure 5.12: The probability of success for the MAR algorithm for $f = 40$ kHz at a propagation speed of 343 m/s and $d_{12}$ set equal to $20\lambda/2$; $\Delta_d \approx 0.7\lambda/2$: a) $\sigma_p^2 = -150$ dB; b) $\sigma_p^2 = -130$ dB; c) $\sigma_p^2 = -110$ dB. The sensor baseline is aligned along the x-axis with $d_{12}$ centered at the origin.
5.7 ERROR ANALYSIS

\[ \sigma_p^2 \approx \frac{1}{\omega^2} \frac{1 - |\gamma(\omega)|^2}{2N_f|\gamma(\omega)|^2} \]  

(5.50)

where \( |\gamma(\omega)|^2 \) is the magnitude squared coherence (MSC) at frequency \( \omega \), which can be defined as [97]

\[ |\gamma(\omega)|^2 = \frac{G_{ss}^2(\omega)}{[G_{ss}(\omega) + G_{nn}(\omega)]^2} \]  

(5.51)

where \( G_{nn} \) is the noise power spectrum. It should be noted that Eqs. (5.50) and (5.51) rely on the assumption that the signal power is equal at the two receivers and so are the noise powers. The former assumption can be satisfied when the sensor baseline is small compared to the distance to the transmitter, and the sensors have identical gains. For a deterministic signal, \( G_{ss}(\omega) \) is available. This fact is exploited in Appendix A.3 to express \( \sigma_p^2 \) in terms of SNR as follows:

\[ \sigma_p^2 \approx \frac{1}{2N_f\omega^2} \left\{ \frac{2\alpha(\omega)L\Lambda + 1}{|\alpha(\omega)L\Lambda|^2} \right\} \]  

(5.52)

where \( \alpha(\omega) \) is the ratio of the power of the received signal at the frequency \( \omega \) to the total signal power, \( L \) is the length of the CPS of the signals and \( \Lambda \) is the linear SNR. It is noted that for a deterministic signal, the required parameter \( \alpha(\omega) \) is fixed for a fixed DFT length and does not depend on the signal amplitude, hence it can be calculated directly from the CPS of a known version of the signal. This means that the channel should not alter the signal frequencies, which agrees with the assumed linear channel model in (5.1) and (5.2). It should be noted here that we assume a single-sided CPS. In the case of a double-sided CPS, Eq. (5.52) is still applicable, however, the values of \( L \) and \( \alpha(\omega) \) need to be set to the ones that correspond to the double-sided CPS case.

By using Eq. (5.52) to calculate the required variance, \( P_{s,m}, P_s, \zeta_m^2 \) and \( \zeta^2 \) can be plotted directly against the SNR. It can be indicated from (5.52) that the performance depends, among other factors, on the parameters that determine \( \sigma_p^2 \); nominally; \( N_f, \alpha(\omega), L \) and the SNR. Performance is expected to improve as each of these four parameters increases. This can be attributed to the inverse proportionality of \( \sigma_p^2 \) with the former parameter, and the dominance of the squared value in the denominator for the latter three parameters. However, herein, only the effect of the SNR will be of interest.
5.7.6 Simulation and Analytical Results

In this section, the proposed disambiguation method is applied to delay disambiguation when the signal is a sinusoid. For all of the results presented in this section, the emitter and the sensors are assumed to be in the same plane, with the origin of the assumed coordinate system coinciding with the midpoint between sensor 1 and sensor 2. All distances are given in units of half-wavelengths. The test signal is a sinusoid with an acoustic frequency of 40 kHz (sampled at 160 KHz) and speed of 343 m/s. The number of samples used to estimate the CPS are 640 in 20 non-overlapping rectangular windows. A 32-point Fast Fourier Transform (FFT) is used to realize the DFTs. All the results presented in this section are based on phase-differences obtained from a single frequency bin that corresponds to the peak of the CPS.

First, simulation results are compared to the analytical predictions. Fig. 5.13 (a and b) plots the MSE for the MAV step and for the whole disambiguation method. The figure is plotted for $D_{12} = 5$, $\Delta = 1$ and emitter location $[1166, 1166, 0]^T$ (1166 is approximately 5 m and the location corresponds to $\theta = 45^0$). Simulation results were obtained from $10^4$ Monte Carlo trials. It can be seen that the analytical performance prediction closely approximates the performance obtained from simulation. It can be seen that the MSE increases as SNR decreases. For low SNR, the effect of outliers is more noticeable in the form of large MSE values. The MAV step is seen to have a constant bias thus its performance remains constant as SNR decreases at high SNRs. This bias makes the MAV step less interesting, and hence will not be considered further. On the other hand, the MSE for the whole algorithm (Fig. 5.13 (b)) increases consistently as the SNR decreases, up to a certain point where a threshold effect can be seen such that the MSE increases significantly beyond the threshold point. This threshold is similar to that of the maximum likelihood (ML) estimator (e.g., see [98, 99]), which for sparse arrays (like the ones under consideration) is mainly due to the effect of grating lobes (see [46] and Chapter 2). However, establishing the relationship between grating lobes and the threshold effect pertaining to the proposed disambiguation method is not straightforward. This is due to the fact that the approach under consideration is two-stage—ambiguos estimation followed by a disambiguation—whereas array processing techniques (like the ML) use the spatial and temporal samples of the signals directly to estimate the AOA.

Due to the closeness of the analytical performance to simulation, subsequently
only the analytical formulae—for the whole algorithm—will be used to compare different configurations.

Fig. 5.14 plots the probability of success corresponding to Fig. 5.13. Again, a threshold effect that agrees with that in Fig. 5.13 is manifested—the probability of success falls below the nominal SNR threshold value. Above the threshold, the probability of success approaches the value ‘1’. It can therefore be concluded that the success of the proposed disambiguation method is almost guaranteed when it is applied in low to moderate noise conditions.

Fig. 5.15 plots the MSE for different values of $\Delta_d$, fixed $D_{12} = 5$ and emitter at $[1166,1166,0]^T$. The figure demonstrates the effect of the value of $\Delta_d$. It can be seen that for a fixed $D_{12}$ and fixed emitter location, $\Delta_d$ determines the threshold SNR for the triplet. The threshold SNR is found to decrease as $\Delta_d$ increases, with the lowest threshold SNR occurring at the maximum permissible value of $\Delta_d$, i.e., $\Delta = 1$. This can be explained by that, for a fixed $D_{12}$ and fixed location, $\Delta_d$ determines the value of the parameter $\mu_1$. Since $\mu_1$ increases as $\Delta_d$ decreases, an increase in $\mu_1$ results in an increased variance of $\varepsilon$ (see Eq. (5.38)) leading to an increased outlier contribution in Eq. (5.42) and Eq. (5.49) at relatively higher SNRs. Therefore, it is recommended that $\Delta_d$ be set to its maximum value.

Fig. 5.16 plots the MSE for different values of $D_{12}$, fixed $\Delta_d = 1$ and emitter at $[1166,1166,0]^T$. The threshold SNR is seen to increase as $D_{12}$ increases. Again this can be explained by an increase in $\mu_1$. It can also be seen that for $D_{12} = 20$ the solution is biased and unreliable. This can be explained by a complete failure of the disambiguation algorithm even in the noise-free case. In other words, the emitter is located in the failure region (see Fig. 5.10) and disambiguation does not work. This emphasizes the fact that for a fixed emitter location, there exists an upper limit on the inter-sensor spacing. It should be noted here that while the asymptotic performance for time-delay estimation is identical for all values of $D_{12}$ other than $D_{12} = 20$, it is straightforward that larger values of $D_{12}$ will result in better performance in terms of AOA estimation. This is due to the fact that angular MSE is inversely proportional to the square of the sensor separation [28].

In Fig. 5.17, it is shown that complete failure can be avoided if the emitter is moved sufficiently far from the sensors, regardless of the inter-sensor spacings. With fixed $D_{12} = 20$ and $\Delta_d = 1$, the emitter location is varied. The locations are selected such that the AOA, as viewed from the mid-point between sensor 1 and sensor 2, is fixed ($\theta = 45^\circ$). The first location is $[1166,1166,0]^T$ (Note
that in the figure, \([x, y, z]\) is used instead of \([x, y, z]^T\). The other locations are obtained by multiplying the coordinates of the first location by a constant. It is seen that by increasing the emitter distance, complete failure is avoided and performance is improved consistently as the distance increases. Performance tends to saturate when the emitter is sufficiently far from the receiver. This can be attributed to the convergence of the situation to the ideal far-field, and hence reduction in the effects caused by deviation from the assumed model. More specifically, the parameter \(\rho\) approaches unity, and hence the bias terms in (5.38) are reduced. It should be noted here that different geometries analyzed here will have different spherical spreading loss that may result in different SNR values for each geometry. However, this effect has not been taken into account in the graphs. In other words, it is assumed that the signal power levels received by the array are equal in all cases.

5.8 Experimental Results

The proposed disambiguation method was tested experimentally in a normal office (see Fig. 5.18). The test signal was a sinusoid with an acoustic frequency equal to 40 kHz, sampled at approximately 168 kHz. The frame size was 512 with 384 overlap between successive frames, and a 512-point DFT was used to calculate the CPS of the received signals. Throughout the tests, the array parameter \(\Delta_d\) was fixed at approximately 2.5 mm, while four different values of \(d_{12}\) were tested. The proposed method was tested for different emitter angles.
5.8 EXPERIMENTAL RESULTS

Figure 5.14: The (analytical) probability of success of the MAV step and that of the whole proposed disambiguation algorithm corresponding to Fig. 5.13.

Figure 5.15: The MSE for time-delay estimation for different values of $\Delta_d$, fixed $D_{12} = 5$ and emitter at $[1166, 1166, 0]^T$.

Figure 5.16: The MSE for time-delay estimation for different values of $D_{12}$, fixed $\Delta_d = 1$ and emitter at $[1166, 1166, 0]^T$. 

121
Figure 5.17: The MSE for time-delay estimation for different emitter distances/locations, fixed $D_{12} = 20$ and fixed $\Delta_d = 1$.

Figure 5.18: The test room with the locations of the transmitter (Tx) and the receiver (Rx) marked. Heights (hgt) are quoted between brackets.

In all of the tests, the phase-differences were estimated from the peaks of the two CPS computed from the three received signals.

Table 5.1 presents sample results for an emitter angle of $60^\circ$. Each entry in the table was estimated from 10 independent tests. The column titled “Mean angle” gives the mean values of the angle estimates that correspond to the MSE estimates. Unlike the simulation and analytical results, the experimental results do not show any particular trend. This can be attributed to the presence of some practical factors that have not been considered in the simulation and the theoretical analysis (e.g., reverberation). However, the table confirms the effectiveness and the feasibility of the proposed disambiguation approach when it is applied in the real environment.
### 5.9 Summary

In this chapter, a method for phase-difference ambiguity resolution was presented. A sensor configuration consisting of a triplet of collinear sensors, with the difference of the two smaller inter-sensor spacings being not larger than $\lambda/2$ of the received signal, coupled with a proposed algorithm, was demonstrated to resolve the ambiguity problem in most cases. The chapter focused on the situation when the emitter is not sufficiently far from the sensor array such that the far-field assumption is not perfectly satisfied. In such case, not only the distance, but also the angle of the emitter relative to the sensors determines performance. Formulae were proposed that can be used to analytically determine the performance for the case of a sinusoid signal under Gaussian white noise, given the emitter location and the sensor configuration. Analytical performance was found to be close to that obtained from simulation. The MSE was seen to exhibit a threshold phenomenon against the SNR. This threshold was the main performance feature that has been used for comparing different scenarios.

Various sensor configurations were tested. The effects of different parameters were studied separately. It was found that, in the under non-ideal far-field, the inter-sensor spacing can be increased up to a certain point, above which complete failure results. On the other hand, for a fixed sensor configuration and a fixed AOA, performance tends to improve as the emitter-array distance increases. Finally, sample experimental results obtained from tests carried out in the real environment were presented to affirm the feasibility of the approach.

<table>
<thead>
<tr>
<th>$d_{12}$ (mm)</th>
<th>MSE ($[\mu s]^2$)</th>
<th>Mean angle (degrees)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.0</td>
<td>0.027</td>
<td>60.2</td>
</tr>
<tr>
<td>12.0</td>
<td>6.268</td>
<td>55.5</td>
</tr>
<tr>
<td>15.0</td>
<td>0.716</td>
<td>61.3</td>
</tr>
<tr>
<td>17.5</td>
<td>1.619</td>
<td>61.6</td>
</tr>
</tbody>
</table>
This chapter presents a continuation of the work described in the previous chapter proposing an alternative solution to the phase-difference ambiguity problem. In the previous chapter, spatial diversity was exploited to resolve ambiguities pertaining to a single-frequency signal. In this chapter, the proposed solution is based on frequency diversity, where the signal is composed of more than one frequency. More specifically, the method proposed herein focuses on disambiguation using two frequency components. The method presented in this chapter does not require extra sensors since frequency diversity is found sufficient for ambiguity resolution.

6.1 Introduction

The focus of this chapter is on phase-difference ambiguity resolution exploiting frequency diversity. The problem is closely related to the problem of phase unwrapping (exploiting frequency diversity) that is frequently discussed in the literature and has importance in different fields of application. The relationship between the general phase unwrapping and phase-difference unwrapping is that in the former case, phase can be a nonlinear function of frequency, while in the latter case, phase-difference (or the phase of the cross-power spectrum (CPS)) has a strictly linear relationship with frequency. This asserts that general phase unwrapping methods can be directly applied to phase-difference unwrapping.

A few phase unwrapping methods that are based on frequency diversity have
been proposed in the literature (e.g., [16, 33, 34, 35, 36]). In general, these methods rely on the fact that the signal spectrum is continuous, and that it starts at zero frequency (i.e., the wideband signal case) or some wrapping-free lower frequency that can be used as the starting point of the unwrapping process. The phase is unwrapped progressively starting from the lowest frequency, leaving the success of the whole process dependent on the success of the unwrapping at low frequencies.

As has been mentioned in the previous chapter, in the case of frequency hopping spread spectrum (FHSS) signals, the signals are bandpass in nature. Therefore, even the lowest frequency is subject to phase wrapping when the signals are received by sensors whose separation exceeds $\lambda_\text{max}/2$, half of the maximum wavelength. Conventional phase unwrapping methods cannot be used in this case.

In this chapter, we propose a new linear phase unwrapping method that unwraps the phase of the CPS (or the phase-difference) at each frequency independently (not progressively). The proposed method exploits the knowledge of the upper limit of the phase-frequency line slope of the CPS phases (which is equivalent to the maximum observable time-delay) and does not require continuous phase or phase that starts from zero. Therefore, the proposed method is capable of handling the case where the phase of the lowest available frequency is uncertain, for instance, the degenerate case of the bandpass signal. In fact, the proposed method needs as few as two frequency components for the unwrapping to be carried out. This chapter will focus on phase-difference unwrapping using only two frequency components. In the following chapter, a frequency pairing approach will be proposed to extend the proposed method to the case where more frequency components are available.

This chapter is organized as follows. In Section 6.2, the proposed phase unwrapping method for two frequencies is developed. Section 6.3 studies the effect of noise on the proposed phase unwrapping method. In Section 6.4, the proposed unwrapping method is tested experimentally and results are presented. Section 6.5 gives a summary of the whole chapter.
6.2 Phase Unwrapping for Two Frequency Components

The proposed frequency-diversity-based disambiguation method requires only two sensors. Therefore, the three-sensor model utilized in the previous chapter will reduce to the two-sensor one given by the following pair of equations:

\[ x_1(nT_s) = s(nT_s) + w_1(nT_s) \]  \hspace{1cm} (6.1)
\[ x_2(nT_s) = s(nT_s - \tau) + w_2(nT_s) \]  \hspace{1cm} (6.2)

where \( s(nT_s) \) (\( n = 0, 1, 2, \ldots \) and \( T_s \) is the sampling interval) is a sampled version of a transmitted multi-frequency signal \( s(t) \); \( w_1(nT_s) \) and \( w_2(nT_s) \) are samples of two added white Gaussian noise (AWGN) processes; and \( \tau \) is the delay of the received signal between the two sensors. As preceded in Chapter 5, \( \tau = \sin(\theta)d/c \); where \( d \) is the receiver separation, \( c \) is the speed of propagation of the wave and \( \theta \) is the angle of arrival of the wave as viewed from the receiver baseline midpoint.

Similar to the discussion in the previous chapter, the estimated phase at each frequency of the complex CPS can be expressed as

\[ \hat{\phi} = (\varphi + \epsilon) + 2\pi k \]  \hspace{1cm} (6.3)

where \( k \in \mathbb{Z} \) are phase wrapping parameters; \( (\varphi + \epsilon) \in [-\pi, \pi] \) are noisy principal phase components with \( \varphi \) being the true principal phase and \( \epsilon \) representing the contribution of noise.

Now, consider two frequencies \( \omega_u \) and \( \omega_v \). The phase estimates at these two frequencies can be expressed following (6.3) as

\[ \hat{\phi}_u = (\varphi_u + \epsilon_u) + 2\pi k_u \]  \hspace{1cm} (6.4)
\[ \hat{\phi}_v = (\varphi_v + \epsilon_v) + 2\pi k_v. \]  \hspace{1cm} (6.5)

6.2.1 The Noise-Free Case

Theorem 6.1 In a noise-free situation, a sufficient condition for the true phases in Eqs. (6.4) and (6.5) to be identifiable from the two principal components, is that the inter-frequency separation (in radians) be not greater than the reciprocal of the
maximum possible slope of the phase-frequency line multiplied by \( \pi \), i.e. \(|\Delta \omega| \triangleq |\omega_v - \omega_u| \leq \pi c/d\).

**Proof.**

The proof of Theorem 6.1 is similar in spirit to that of Theorem 5.1 presented in the previous chapter. Nevertheless, for the sake of more compact notations, herein, the presentation will be given in terms of phases rather than delays.

From the linearity of phase, we have

\[
\phi_v = \frac{\omega_v}{\omega_u} \phi_u
\]

(6.6)

where \( \phi_u \) and \( \phi_v \) are the error-free versions of \( \hat{\phi}_u \) and \( \hat{\phi}_v \), respectively. It should be noted here that Eq.(6.6) is true only when the effect of dispersion [21] is ignored. For the frequency range of interest for this work, dispersion in air at these frequencies is so small that it can be neglected [96], and hence, Eq.(6.6) will approximately hold. By setting the error terms in (6.4) and (6.5) equal to zero, substituting Eq. (6.6) in (6.5), and subtracting (6.4) from the result, yields

\[
\phi_u = \mu_{uv} (\varphi_u - \varphi_v + 2\pi k_{uv})
\]

(6.7)

where \( k_{uv} \triangleq k_v - k_u \) and

\[
\mu_{uv} \triangleq \frac{\omega_u}{\omega_v - \omega_u}.
\]

(6.8)

Now phase unwrapping is transformed into a problem of determining the correct value of the integer \( k_{uv} \). Hence, it is convenient to write (6.7) in the form

\[
\phi_u(k) = \mu_{uv} (\varphi_u - \varphi_v + 2\pi k)
\]

(6.9)

where \( k \in \mathbb{Z} \) is a general integer variable whose true value \( k_{uv} \) is being sought. Now, define the identifiability criterion for the true phase \( \phi_u(k_{uv}) \) as being the only valid phase amongst the phases \( \phi_u(k) \), with the validity of phase defined as falling in the interval \((-\omega_u d/c, \omega_u d/c)\), i.e.,

\[
\exists! \phi_u(k = k_{uv}) \in \left(-\omega_u \frac{d}{c}, \omega_u \frac{d}{c}\right), k \in \mathbb{Z}.
\]

(6.10)

Now, consider the true phase value \( \phi_u(k_{uv}) \). Any other false candidate value of the phase \( \phi_u(k_{uv} \pm q), q \in \mathbb{N} \) can be expressed, based on (6.9), as

\[
\phi_u(k_{uv} \pm q) = \phi_u(k_{uv}) \pm 2\pi q \mu_{uv}.
\]

(6.11)
From the identifiability criterion of the true phase in Eq. 6.10, for \( \phi_u(k_{uv} \pm q) \), \( \forall q \) to be invalid phases, the following inequality must be satisfied:

\[
|\phi_u(k_{uv})| \geq \frac{\omega_u d}{c}, \forall q. \tag{6.12}
\]

Considering all the sign (+/-) combinations of \( \phi_u(k_{uv}) \) and \( \mu_{uv} \), a necessary and sufficient condition for the true phase to be uniquely identifiable can be stated as

\[
2\pi |\mu_{uv}| \geq \omega_u \frac{d}{c} + |\phi_u(k_{uv})|. \tag{6.13}
\]

Since the true phase \( \phi_u(k_{uv}) \) is generally unknown, it is more convenient to obtain a more strict—but accessible—version of the condition in (6.13) by setting \( |\phi_u(k_{uv})| = |\phi_u(k_{uv})|_{\text{max}} = \omega_u \frac{d}{c} \), which after substituting for \( \mu_{uv} \) from (6.8) and manipulating results in the sufficient—but not generally necessary—condition

\[
|\omega_v - \omega_u| \leq \frac{\pi c}{d}. \tag{6.14}
\]

The condition given in (6.14) is the end of the proof of Theorem 6.1.

In the following, we show how the condition in (6.14) can be exploited to identify the true value of \( \phi_u \). The same logic can be applied to \( \phi_v \). Otherwise, the true value of \( \phi_v \) can directly be determined from the true value of \( \phi_u \) using (6.6).

By inspecting Eq. (6.9) after substituting Eq. (6.8) and assuming that the condition in Eq. (6.14) is satisfied, the following relationship is obtained

\[
|\phi_u(k)| \geq \omega_u \frac{d}{c} \left| \frac{\phi_u - \phi_v + 2\pi k}{\pi} \right|, \forall k. \tag{6.15}
\]

The relationship given by (6.15) asserts that \( |\phi_u(k)| \), for some \( k \), is either equal to or more than the product of \( \omega_u d/c \) and \( |(\phi_u - \phi_v + 2\pi k)/\pi| \). For \( \phi_u(k) \) to be a valid phase, the quantity \( |(\phi_u - \phi_v + 2\pi k)/\pi| \) must be less than unity. Contemplating the expression \( |(\phi_u - \phi_v + 2\pi k)/\pi| \), considering the valid range of \( \phi_u \) and \( \phi_v \), it can be deduced that any value of \( k \) such that \( |k| > 1 \) will result in a value of \( |(\phi_u - \phi_v + 2\pi k)/\pi| \) that surely exceeds unity. It can therefore be concluded that there are only three possible values for \( k \) and that at most one of them is anticipated to yield a valid phase (i.e., \( |\phi_u(k)| \leq \omega_u d/c \)). Literally, the search should be confined to the subset \( \{-1, 0, 1\} \).

To summarize, in the noise-free case, and for \( \omega_u \) and \( \omega_v \) that satisfy (6.14),
the true phase can be restored by evaluating Eq. (6.9) for \( k \in \{-1, 0, 1\} \) and selecting the value that falls in the interval \((-\omega_u d/c, \omega_u d/c\)). Only one such value will exist due to the restriction in (6.14).

### 6.2.2 The Noisy Case

By considering the noise terms in Eqs. (6.4) and (6.5), one can use the same procedure used to obtain (6.9), to obtain the noisy version of the same equation as

\[
\hat{\phi}_u(k) = \mu_{uv} (\phi_u - \phi_v + 2\pi k) + \epsilon_{uv}
\]

where \( \epsilon_{uv} \triangleq \mu_{uv} (\epsilon_u - \epsilon_v) \) and represents the total effect of noise. It is clear that due to the noise term in (6.16), one can no longer talk about uniqueness in the sense reflected in (6.10). However, the uniqueness in (6.10) implies that the true phase \( \phi_u(k = k_{uv}) \) is also identifiable in a different way, that is, the true phase is also the one that has the minimum absolute value. This new criterion is useful since the minimum absolute value property is maintained even in the noisy case, as can be deduced directly from Eq. (6.16). Thus, from the discussion of the noise-free case, the true phase estimate in the noisy case can be recovered using

\[
\hat{\phi}_{u,1} = \hat{\phi}_u(k_{uv}), \text{ where } k_{uv} = \arg \min_k |\hat{\phi}_u(k)|, \quad k \in \{-1, 0, 1\}.
\] (6.17)

In Eq. 6.17, \( \hat{k}_{uv} \) in represents an estimate for the true integer \( k_{uv} \). Therefore, in the case of success (i.e., when \( \hat{k}_{uv} = k_{uv} \)), \( \hat{\phi}_{u,1} \) according to (6.16) will be given by

\[
\hat{\phi}_{u,1} = \phi_u(k_{uv}) + \epsilon_{uv} = \phi_u + \epsilon_{uv}
\] (6.18)

It can be seen from Eq. (6.18) together with Eq. (6.4) and the definition of the error term \( \epsilon_{uv} \) that the unwrapping results in an increased error due to the accumulation of error (from the noisy principal phase components at each of the two frequencies). To reduce this effect, the fact that \( k_u \) is an integer value can be exploited. Literally, the estimate \( \hat{\phi}_{u,1} \) can be substituted into Eq. (6.4)
(for $\phi_u$), and the integer $k_u$ can be estimated as

$$k_u = \Psi \left( \frac{\hat{\phi}_{u,1} - (\varphi_u + \epsilon_u)}{2\pi} \right)$$

(6.19)

where $\Psi(.)$ is the round-off of the value.

Finally, the true phase is re-estimated from Eq. (6.4) as

$$\hat{\phi}_u = (\varphi_u + \epsilon_u) + 2\pi \hat{k}_u.$$  

(6.20)

In the case where (6.19) is capable of restoring the correct integer value (i.e., when $\hat{k}_u = k_u$), the phase-difference resulting from (6.20) will be given by

$$\hat{\phi}_u = (\varphi_u + 2\pi k_u) + \epsilon_u$$

$$= \varphi_u + \epsilon_u$$

(6.21)

which corresponds to the true phase-difference plus the noise term $\epsilon_u$. It is clear when comparing (6.18) and (6.21) that (6.20) can potentially give a better estimate for the true phase-difference compared to (6.17). However, that depends on the success of Eq. (6.19) in restoring the correct integer. The decision as to whether (6.17) is better or (6.20) requires a noise analysis, which will be discussed in Section 6.3. The analysis therein reveals that the latter is asymptotically better than the former. Therefore, Eqs. (6.19) and (6.20) will be considered as part of the proposed disambiguation method whose summary will be given in the following subsection.

### 6.2.3 Summary of the Disambiguation Algorithm

To recover the estimates of the true phase-difference, $\hat{\phi}_u$, from the noisy versions of the principal phase-differences, $(\varphi_u + \epsilon_u)$ and $(\varphi_v + \epsilon_v)$, when (6.14) is satisfied, the following procedure can be used:

**I.** Calculate the candidate estimates of the true phase-difference $\hat{\phi}_u(k)$ (as in 6.16) for $k \in \{-1, 0, 1\}$ by setting $\{\varphi_u, \varphi_v\} = \{(\varphi_u + \epsilon_u), (\varphi_v + \epsilon_v)\}$ in Eq. (6.9).

**II.** Obtain a first estimate ($\hat{\phi}_{u,1}$) of the true phase-difference using (6.17).

**III.** Use $\hat{\phi}_{u,1}$ to obtain an estimate for $k_u$ from (6.19).

**IV.** Obtain the final estimate of the true phase-difference using (6.20).
Subsequently, step II will be referred to as the *minimum absolute value* (MAV) step, and step III as the *round-off* (RO) step.

### 6.3 Noise Effect

In this section, the effect of noise on the performance of the proposed phase unwrapping method is considered. For the purpose of analyzing performance under noise, it will be assumed, without loss of generality, that $\epsilon_u$ and $\epsilon_v$ (in Eqs. (6.4) and (6.5), respectively) are two zero-mean Gaussian random variables with variances $\sigma_u^2$ and $\sigma_v^2$, respectively. Hence, $\epsilon_{uv}$, from its definition, will be a zero-mean Gaussian random variable with a variance given by

$$\sigma_{uv}^2 = \mu_{uv}^2 \left( \sigma_u^2 + \sigma_v^2 \right). \quad (6.22)$$

#### 6.3.1 Noise Effect on the MAV Step

First consider the MAV stage of the proposed algorithm (see Section 6.2.3), where a choice between three phase-difference estimates is made. Each of these three estimates corresponds to an integer; the integer corresponding to the correct selection is $k_{uv}$; the other two integers corresponding to the false phase-differences will be denoted as $k_1$ and $k_2$. Due to the effect of noise represented in the error term $\epsilon_{uv}$ (see Eq. (6.16)), a wrong choice might be made in Eq. (6.17) corresponding to the MAV step. The probability that the MAV step produces the correct answer, i.e, the probability of success for the MAV step, can be written as

$$P_{s,m} = P \left[ |\hat{\phi}_u(k_{uv})| < |\hat{\phi}_u(k_1)| \quad \& \quad |\hat{\phi}_u(k_{uv})| < |\hat{\phi}_u(k_2)| \right]$$

$$= P \left[ |\phi_u(k_{uv}) + \epsilon_{uv}| < |\phi_u(k_1) + \epsilon_{uv}| \right.$$  

$$\& \left. |\phi_u(k_{uv}) + \epsilon_{uv}| < |\phi_u(k_2) + \epsilon_{uv}| \right] \quad (6.23)$$

where “&” is the logical AND operation. The probability $P_{s,m}$ depends on the actual values of $\phi_u(k_{uv})$, $\phi_u(k_1)$ and $\phi_u(k_2)$. A procedure for evaluating $P_{s,m}$ from the CDF of the random variable $\epsilon_{uv}$ is described in Appendix B.1.

The probability of failure of the MAV step is given by $1 - P_{s,m} = P_{k_1} + P_{k_2}$, where each of $P_{k_1}$ and $P_{k_2}$ represents the probability of obtaining one of the two false phase-differences from the MAV step. The probabilities $P_{k_i}, i = 1, 2$ can be obtained from (6.23) by swapping the role of $\phi_u(k_{uv})$ and $\phi_u(k_i)$. The
procedure for evaluating the probabilities $P_{k_1}$ and $P_{k_2}$ is also listed in Appendix B.1.

Based on the probabilities $P_{s,m}$, $P_{k_1}$ and $P_{k_2}$, an approximation for the mean squared error (MSE) corresponding to the phase-difference estimate obtained from the MAV step can be obtained. For the purpose of maintaining consistency and facilitating comparison with the MSEs from the previous chapter, the MSE of the delay corresponding to the estimated phase-difference will be given. The latter MSE can be written in the form

$$\tilde{\epsilon}_{m}^{2} = \Gamma_{s,m} P_{s,m} + \Gamma_{k_1,m} P_{k_1} + \Gamma_{k_2,m} P_{k_2}$$  \hfill (6.24)

where $\Gamma_{s,m}$, $\Gamma_{k_1,m}$ and $\Gamma_{k_2,m}$ are, respectively, the MSEs associated with the three cases representing success and failure, and are defined as

$$\Gamma_{s,m} = \frac{\sigma_{u}^{2} + \sigma_{v}^{2}}{(\omega_{v} - \omega_{u})^{2}}$$

$$\Gamma_{k_1,m} = \frac{\sigma_{u}^{2} + \sigma_{v}^{2}}{(\omega_{v} - \omega_{u})^{2}} + \left[ \frac{\phi_{u}(k_1) - \phi_{u}(k_{uv})}{\omega_{u}} \right]^{2}$$

$$\Gamma_{k_2,m} = \frac{\sigma_{u}^{2} + \sigma_{v}^{2}}{(\omega_{v} - \omega_{u})^{2}} + \left[ \frac{\phi_{u}(k_2) - \phi_{u}(k_{uv})}{\omega_{u}} \right]^{2}. \hfill (6.25)$$

Note that the effect of $\Gamma_{k_1,m}$ and $\Gamma_{k_2,m}$ will be manifested in the form of a bias in the MSE of the MAV step.

### 6.3.2 Noise Effect on the RO Step

The probability of success for the RO step, and hence for the whole disambiguation method, is the probability that the correct integer $k_{u}$ is restored. That can be expressed as

$$P_{s} = P \left\{ \psi \left[ \phi_{u,1} - (\phi_{u} + \epsilon_{u}) \right] = k_{u} \right\} \hfill (6.26)$$

where $\phi_{u,1}$ can take one of three different values depending on the suc-
cess/failure of the MAV algorithm. Accordingly, $P_s$ can take the form

$$P_s = P \left\{ \psi \left( \frac{\phi_u(k_{uv}) + \epsilon_{uv} - (\varphi_u + \epsilon_u)}{2\pi} \right) = k_u \right\} P_{s,m} + P \left\{ \psi \left( \frac{\phi_u(k_1) + \epsilon_{uv} - (\varphi_u + \epsilon_u)}{2\pi} \right) = k_u \right\} P_{k_1} + P \left\{ \psi \left( \frac{\phi_u(k_2) + \epsilon_{uv} - (\varphi_u + \epsilon_u)}{2\pi} \right) = k_u \right\} P_{k_2}, \quad (6.27)$$

According to (6.11), $\phi_u(k_i)$ can be expressed as $\phi_u(k_i) = \phi_u(k_{uv}) + 2\pi q_i \mu_{uv}$, where here $q_i$ is a positive or negative integer value equal to $k_i - k_{uv}$. Substituting for this together with $[\phi_u(k_{uv}) - \varphi_u]/2\pi = k_u$ and manipulating, yields

$$P_s = P \left\{ \psi \left( k_u + \frac{\epsilon_{uv} - \epsilon_u}{2\pi} \right) = k_u \right\} P_{s,m} + P \left\{ \psi \left( k_u + \frac{\epsilon_{uv} - \epsilon_u + 2\pi q_1 \mu_{uv}}{2\pi} \right) = k_u \right\} P_{k_1} + P \left\{ \psi \left( k_u + \frac{\epsilon_{uv} - \epsilon_u + 2\pi q_2 \mu_{uv}}{2\pi} \right) = k_u \right\} P_{k_2}, \quad (6.28)$$

which can be expressed as

$$P_s = P \left( -0.5 \leq \frac{\epsilon_{uv} - \epsilon_u}{2\pi} < 0.5 \right) P_{s,m} + P \left( -0.5 \leq \frac{\epsilon_{uv} - \epsilon_u + 2\pi q_1 \mu_{uv}}{2\pi} < 0.5 \right) P_{k_1} + P \left( -0.5 \leq \frac{\epsilon_{uv} - \epsilon_u + 2\pi q_2 \mu_{uv}}{2\pi} < 0.5 \right) P_{k_2}, \quad (6.29)$$

Eq. (6.29) can finally be converted to the form

$$P_s = [\Phi_{\epsilon_{uv} - \epsilon_u} (\pi) - \Phi_{\epsilon_{uv} - \epsilon_u} (-\pi)] P_{s,m} + [\Phi_{\epsilon_{uv} - \epsilon_u} (\pi + 2\pi q_1 \mu_{uv}) - \Phi_{\epsilon_{uv} - \epsilon_u} (-\pi + 2\pi q_1 \mu_{uv})] P_{k_1} + [\Phi_{\epsilon_{uv} - \epsilon_u} (\pi + 2\pi q_2 \mu_{uv}) - \Phi_{\epsilon_{uv} - \epsilon_u} (-\pi + 2\pi q_2 \mu_{uv})] P_{k_2}, \quad (6.30)$$

where $\Phi_{\epsilon_{uv} - \epsilon_u}$ is the CDF of the random variable $\epsilon_{uv} - \epsilon_u$, which has a Gaussian distribution with a zero mean and a variance equal to $\sigma_{uv}^2 + \sigma_u^2$.

The MSE for the delay associated with the phase-difference obtained from the
6.3 NOISE EFFECT

The whole disambiguation algorithm is given by
\[ \zeta^2 = \Gamma_s P_s + \Gamma_f (1 - P_s) \]  
(6.31)
where \( \Gamma_s = \sigma_u^2 \), while an approximation for \( \Gamma_f \) is derived in Appendix B.2.

6.3.3 Performance Versus SNR

The MSEs, \( \zeta_m^2 \) and \( \zeta^2 \), are functions of the variances \( \sigma_u^2 \) and \( \sigma_v^2 \). To obtain the performance directly against SNR, these variances need to be expressed in terms of SNR. This will be discussed in the following for the case of the CPS-based phase-difference estimation of a deterministic signal, which is the case of interest for this work.

As in Chapter 5, for a large number of frames \( (N_f) \), \( \epsilon_u \) and \( \epsilon_v \) are approximately Gaussian and \( \sigma_i^2, i = u, v \) can be approximated as [20, 22]
\[ \sigma_i^2 \approx \frac{1 - |\gamma(\omega_i)|^2}{2N_f|\gamma(\omega_i)|^2} \]  
(6.32)
where \( |\gamma(\omega_i)|^2 \) is the magnitude squared coherence (MSC) at frequency \( \omega_i \), which can be defined as [97]
\[ |\gamma(\omega_i)|^2 = \frac{G_{ss}^2(\omega_i)}{[G_{ss}(\omega_i) + G_{nn}(\omega_i)]^2} \]  
(6.33)
where \( G_{nn} \) is the noise power spectrum. It should be noted that Eqs. (6.32) and (6.33) rely on the assumption that the signal power is equal at the two receivers and so are the noise powers. The former assumption can be satisfied when the sensor baseline is small compared to distance to the transmitter, and the sensors have identical gains.

It is noticed that for a deterministic signal \( s(nT_s) \), \( G_{ss} \) can be known. Similar to Chapter 5, \( \sigma_i^2 \) can be expressed in terms of the SNR as follows:
\[ \sigma_i^2 \approx \frac{1}{2N_f} \left\{ \frac{2\alpha(\omega_i)\Lambda + 1}{[\alpha(\omega_i)\Lambda]^2} \right\} \]  
(6.34)
where \( \alpha(\omega_i) \) is the ratio of the power of the received signal at the frequency \( \omega_i \) to the total signal power, \( L \) is the length of the CPS of the signals and \( \Lambda \) is the linear SNR. It is noticed that for a deterministic signal, the required parameter \( \alpha(\omega_i) \) is fixed for a fixed DFT length and does not depend on the signal am-
plitude. Hence, it can be calculated directly from the CPS of a known version of the signal. This requires that the channel should not introduce any effect on the signal frequencies, which agrees with the linear channel model in (6.1) and (6.2). It should be noted here that we assume a single-sided CPS. In the case of a double-sided CPS, Eq. (6.34) is still applicable, however, the values of \( L \) and \( \alpha(\omega_i) \) need to be set to the ones that correspond to the double-sided CPS case.

By using Eq. (6.34) to calculate the required variances, \( P_{s,m}, P_s, \zeta_m^2 \) and \( \zeta_s^2 \) can be plotted directly against the SNR. It can be seen from (6.34) that the performance depends, among other factors, on the parameters that determine \( \sigma_i^2 \); nominally; \( N_f, \alpha(\omega_i), L \) and the SNR. Performance is expected to improve as each of these four parameters increases. This can be attributed to the inverse proportionality of \( \sigma_i^2 \) to the former parameter, and the dominance of the squared value in the denominator for the latter three parameters. However, herein, only the effect of the SNR will be of interest.

### 6.3.4 Simulation and Analytical Results

In this section, the proposed disambiguation method is applied to unwrap the phase-differences obtained from the CPS of two signals received by two widely-spaced sensors. For all of the scenarios considered herein, the emitter and the receivers were in the same plane, with the origin of the assumed coordinate system coinciding with the mid-point between the two receivers. The test signal is constructed as the sum of two sinusoids at two different ultrasonic frequencies and speed of 343 m/s. The frequency of one of the two sinusoids is fixed at \( \omega_u = 2\pi \times 40 \text{ k rad} \) (which is equivalent to 40 kHz), while the frequency of the other one is selected according to a specific value of \( \Delta_\omega \). In each test, 20 non-overlapping rectangular frames of each signal are used to estimate the CPS (i.e, \( N_f = 20 \)). Each frame consists of 32 signal samples. In all cases, the sampling frequency is fixed at 160 KHz and the frame length is the same as the DFT length. All the results presented in this section are based on phase-differences obtained from only two frequency bins that are the two peaks of the CPS corresponding to the two sinusoidal frequencies.

First, simulation results are compared to the analytical predictions. Fig. 6.1 (a) and (b) plots the MSE for the MAV step and that for the whole disambiguation algorithm. The figure is plotted for \( d = 5 \lambda_u/2 \) (which is equivalent to approximately 2.14 cm), \( \Delta_\omega = 5 \times 2\pi \text{ k rad} \), 128-point DFT, and emitter location \([1166,1166,0]^T\) (1166 is approximately 5 m and the location corresponds to
6.3 NOISE EFFECT

\[ \theta = 45^\circ \). Simulation results were obtained from 10⁴ Monte Carlo trials. It can be seen that the analytical performance prediction closely approximates the performance obtained from simulation for most of the cases of interest. The deviations at low SNR can be explained by the several approximations that have been introduced in the derivation of the probability of success/failure and the MSE. For both Fig. 6.1 (a) and Fig. 6.1 (b), the MSE increases consistently as the SNR decreases until a certain threshold point on the SNR axis, where the effect of outliers starts to increase significantly. In particular, in Fig. 6.1 (b), the threshold is similar to that of the maximum likelihood estimator (e.g., see [98, 99]), which indicates a good property of the whole disambiguation algorithm. It can also be concluded that the MAV step is not of interest as it does not perform as well as the whole algorithm in terms of the MSE.

Fig. 6.2 plots the probabilities of success corresponding to Fig. 6.1. Again, a threshold effect that agrees with that in Fig. 6.1 is manifested—the probability of success falls below the nominal SNR threshold value. Above the threshold, the probability of success is almost unity. It can therefore be concluded that the success of the proposed disambiguation method is almost guaranteed when it is applied in low to moderate noise conditions.

Due to the closeness of the analytical performance to the performance from simulation, in what follows, only the analytical formulae—of the whole algorithm—will be used for the purpose of comparison.

Fig 6.3 plots the MSE for different values of \( \Delta_\omega \), \( d = 5\lambda_u/2 \), 128-point DFT, and emitter location \([1166,1166,0]^T\). All the values of \( \Delta_\omega \) that are considered are positive; negative values are found to be equivalent so long as the absolute values are equal. For this scenario, it can be seen that \( \Delta_\omega \) determines the value of the threshold SNR. The threshold SNR is found to decrease as \( \Delta_\omega \) increases. This can be explained by noting that, \( \Delta_\omega \) determines the value of the parameter \( \mu_{uv} \), which increases as \( \Delta_\omega \) decreases, resulting in an increase in the variance \( \sigma_{uv}^2 \). Apart from the threshold effect, the asymptotic performance does not depend on the value of \( \Delta_\omega \).

Fig 6.4 plots the MSE for different values of \( d \), fixed \( \Delta_\omega = 5 \times 2\pi \text{ k rad} \), 32-point DFT and emitter at \([1166,1166,0]\). The inter-sensor spacing does not seem to have effect on the value of the threshold SNR or the asymptotic performance. However, below the threshold SNR, increasing \( d \) results in an increased MSE. This can be explained by that increasing \( d \) results in moving the system more towards the an unidentifiability region (see Eq. 6.13 and Eq. 6.14). Increasing \( d \) further can even result in unidentifiability, as reflected by the graphs.
Figure 6.1: The MSE in the estimation of the delay corresponding to the unambiguous phase-difference obtained from a) the MAV step and b) the whole disambiguation algorithm.

Figure 6.2: The (analytical) probability of success of the MAV step and that of the whole proposed disambiguation algorithm corresponding to Fig. 6.1.

Finally, it should be noted here that, contrary to the spatial-diversity-based disambiguation method presented in the previous chapter, the frequency-diversity-based method does not show any interesting performance variations with space. In other words, the location of the emitter is not found to have significant effect on the attainable performance. Note that the latter conclusion does not take the effect of the receiver directivity into consideration.
6.3 NOISE EFFECT

Figure 6.3: The MSE in the estimation of the delay corresponding to the unambiguous phase-difference obtained from the proposed disambiguation algorithm, for different values of frequency separation and $\Delta \omega$ ($2\pi$ k rad).

Figure 6.4: The MSE in the estimation of the delay corresponding to the unambiguous phase-difference obtained from the proposed disambiguation algorithm, for different values of sensor spacing and $d$ ($\lambda_u/2$).
6.4 Experimental Results

The proposed method for phase-difference disambiguation was tested experimentally. The experiments were carried out in a normal office (see Fig. 6.5). Ultrasonic FHSS signals consisting of 20 equally-spaced frequencies in the range 35 - 49.5 kHz were used. A sampling frequency of approximately 168 kHz, a frame size of 512 with 384 overlap between successive frames, together with a 512-point DFT were used. Several tests with different frequency pairs and four different sensor separations were performed. Also, different angles of emitter were tested. Table 6.1 presents sample results for emitter angle of 60°, $\omega_u = 40 \times 2\pi$ k rad and $\omega_v = 45 \times 2\pi$ k rad. Each entry in the table was estimated from 10 independent tests, each consisted of disambiguating the phase-difference estimated from the frequency bin corresponding to $\omega_u$ utilizing the phase-difference estimated from the frequency bin corresponding to $\omega_v$. The column titled “Mean angle” gives the mean values of the angle estimates corresponding to the MSE estimates. Unlike the simulation and analytical results presented in Section 6.3.3, the experimental results do not show any particular trend due to the presence of practical factors that were not considered in the simulation and the analytical performance evaluation (e.g., reverberation). However, the table does emphasize the effectiveness and the feasibility of the proposed frequency-diversity-based disambiguation method when it is applied in the real environment.
Table 6.1: Sample experimental results for emitter at 60°.

<table>
<thead>
<tr>
<th>d (mm)</th>
<th>MSE ([μs]^2)</th>
<th>Mean angle (degrees)</th>
</tr>
</thead>
<tbody>
<tr>
<td>12.5</td>
<td>2.002</td>
<td>55.7</td>
</tr>
<tr>
<td>15.0</td>
<td>0.224</td>
<td>61.2</td>
</tr>
<tr>
<td>17.5</td>
<td>0.653</td>
<td>61.7</td>
</tr>
<tr>
<td>20.0</td>
<td>0.076</td>
<td>59.6</td>
</tr>
</tbody>
</table>

6.5 Summary

This chapter presented a method for phase-difference unwrapping of an ambiguous phase-difference estimate obtained for a certain frequency, exploiting another ambiguous estimate obtained from a different frequency. The identifiability condition derived was that the inter-frequency separation (in radians) be not greater than the reciprocal of the maximum possible slope of the phase-frequency line multiplied by \( \pi \).

Analytical performance was derived and compared to simulation results. It was found that above a certain SNR threshold, the true phase-difference is almost surely identifiable. Asymptotic performance, in terms of the MSE, is found to be independent of the inter-frequency separation and the inter-sensor spacing. On the other hand, asymptotic performance is determined by the parameters that determine the variance of the observed phase-difference estimates, including the SNR. Since the operation of the proposed disambiguation method does not depend on spatial parameters, its performance does not vary notably with the emitter location. Finally, experimental results were presented that confirm the feasibility of the proposed method in practical environments.
This chapter focuses on the problem of estimating the angle-of-arrival (AOA) for FHSS signals. The phase-difference disambiguation methods proposed in the two preceding chapters are employed for AOA estimation using the phase data obtained from the CPS of the received signals. The chapter considers different factors that affect AOA estimation performance. First, the single-source case is considered. Next, it is shown that by exploiting some signal properties, the methods can be extended for the multi-source case. AOA estimation performance using each of the two disambiguation methods is compared. Methods with superior performance to be incorporated in the final system are identified. Comparisons are based on experimental results obtained from tests conducted in the real environment.

7.1 Introduction

This chapter is concerned with the problem of angle-of-arrival (AOA) estimation using phase data. It is well known that the phase of the cross-power spectra (CPS) of a signal received at spatially separated sensors can be utilized to estimate the time-delay of the signal, and hence the AOA (see [13, 20, 21, 22, 16, 23]).

As has been highlighted in Chapter 2, the main problem in using the phase of the CPS to estimate the AOA of a signal is the occurrence of phase wrapping [20] when the receivers are widely-spaced. To avoid phase wrapping in all cases, the receiver separation must not exceed the minimum half-wavelength of the frequency components that are present in the received signal. The possibility of the occurrence of phase wrapping results in ambiguities, that is the estimated AOA is not consistent.

In Chapter 5 and Chapter 6, two different methods have been proposed as so-
solutions to the ambiguity problem. The method in Chapter 5 exploits spatial
diversity while the one in Chapter 6 is based on frequency diversity. In Chap-
ner 5, the discussion was limited to the single-frequency case while in Chapter 6
the method requires a minimum of two frequencies in order for the ambiguity
to be resolved. In this chapter, the two methods will be applied in more ge-
neral scenarios; when more frequency components are present in the received
signal, and also when more than one source are present. The single-source case
(when the signal is also multi-frequency) is handled using a general approach.
On the other hand, the multi-source case is handled using an approach that
exploits some specific properties of the signal type of interest, namely, the or-
thogonality of FHSS signals. However, the latter approach can be applied to
other signals that possesses a similar orthogonality property.

The remainder of this chapter is organized as follows. Section 7.2 discusses
the AOA estimation problem for the single-source case. Section 7.3 extends
the methods in Section 7.2 to the multi-source case exploiting signal ortho-
gonality. Section 7.4, proposes grid-search methods that use the ambiguous
phase-difference estimates to find the AOA without explicitly disambiguating
the phase-differences. Section 7.5 discusses the effect of cycle slipping and
how it can be mitigated. In Section 7.6, experimental results for the various
methods are presented. In Section 7.7, a summary of the chapter is given.

## 7.2 AOA Estimation: the Single-Source Case

Consider the signal model for the three-sensor case presented in Chapter 5:

\[
\begin{align*}
x_1(nT_s) &= s(nT_s) + w_1(nT_s) \\
x_2(nT_s) &= s(nT_s - \tau_{12}) + w_2(nT_s) \\
x_3(nT_s) &= s(nT_s - \tau_{12} - \tau_{23}) + w_3(nT_s)
\end{align*}
\]  

(7.1)

where \(\tau_{12}\) is the signal delay between sensor-1 and sensor-2; \(\tau_{23}\) is the signal
delay between sensor-2 and sensor-3; and \(w_1(nT_s), w_2(nT_s)\) and \(w_3(nT_s)\) are
samples of three noise processes. Herein, this signal model will be used in
both cases when the spatial-diversity-based disambiguation method is applied
and when the frequency-diversity-based disambiguation method is applied.
However, for the frequency-based method, only sensor-1 and sensor-2 will be
utilized. In both cases, it will be assumed that the goal is to obtain an estimate
for the AOA (denoted as \(\theta\)) of the signal observed from the mid-point between
sensor-1 and sensor-2.

As in Chapter 5, the CPS can be estimated as

\[
\hat{G}_{x_1,x_2}(\omega) = \frac{1}{N_f} \sum_{i=0}^{N_f-1} X_1(i)X_2^*(i) \\
\hat{G}_{x_2,x_3}(\omega) = \frac{1}{N_f} \sum_{i=0}^{N_f-1} X_2(i)X_3^*(i)
\]  

(7.2)

where \( N_f \) is the number of frames of each signal used in calculating the CPS. Generally speaking, the averaging entailed in the estimation of the CPS, reduces the effect of uncorrelated noise.

It will be assumed that the CPS of the received signals consist of \( M \) frequencies \( \omega_m, m = 0, \ldots, M - 1 \) in the passband. The unwrapped phase-difference estimates pertaining to each sensor pair at each frequency are given by

\[
\hat{\phi}_{12,m} = \omega_m \tau_{12} + \nu_{12,m} \\
\hat{\phi}_{23,m} = \omega_m \tau_{23} + \nu_{23,m}
\]  

(7.3)

where \( \nu_{12,m} \) and \( \nu_{23,m} \) are random variables that carry the effect of noise and other error sources.

Now, assuming that the estimates \( \hat{\phi}_{12,m} \) are available, the time-delay, \( \tau_{12} \), can be estimated by incorporating the contributions from all of the available frequency bins according to

\[
\hat{\tau}_{12} = \frac{\sum_{m=0}^{M-1} \psi_m \hat{\phi}_{12,m} \omega_m}{\sum_{m=0}^{M-1} \psi_m \omega_m^2}
\]  

(7.4)

where \( \psi_m \) are generic weights (see [16, 20, 21, 22]), for which a unity value coincides with the linear least squares estimator of \( \tau_{12} \) [39]. Subsequently, the linear least squares estimator will be referred to as the least squares estimator, or simply, the LS estimator. When \( \nu_{12,m} \) are Gaussian (this holds when the system noise is Gaussian and sources of error other than noise are negligible) and \( \psi_m \) are set equal to the reciprocal of the variance of the noise terms \( \nu_{12,m} \), i.e., \( \psi_m = [\text{var}(\nu_{12,m})]^{-1} \), Eq. (7.4) gives the best linear unbiased estimator of \( \tau_{12} \) [39].

The AOA, as viewed from the center of the baseline of sensor-1 and sensor-2, is calculated as

\[
\hat{\theta} = \arcsin \left( \frac{c \hat{\tau}_{12}}{\hat{d}_{12}} \right)
\]  

(7.5)
where $c$ is the speed of propagation of the signals, and $d_{12}$ is the distance between sensor-1 and sensor-2.

### 7.2.1 AOA Estimation with Phase Unwrapping

When the system is susceptible to phase wrapping due to wide sensor separation, the ambiguous phase-difference estimates, $\hat{\phi}_{12}$ and $\hat{\phi}_{23}$, relate to the true estimates in (7.3) by

$$
\hat{\phi}_{12, m} = \phi_{12, m} - 2\pi k_{12, m} \\
\hat{\phi}_{23, m} = \phi_{23, m} - 2\pi k_{23, m}
$$

where $k_{12, m}$ and $k_{23, m}$ are the integers corresponding to the ambiguity at each frequency for each pair of sensors. In practice, only the ambiguous estimates in Eq. (7.6) are available. The disambiguation methods from Chapter 5 and Chapter 6 can then be applied to obtain the estimates for the true phase differences as follows.

#### 7.2.1.1 Using Spatial-Diversity-Based Disambiguation

For each frequency bin, $m$, the two ambiguous estimates, $\hat{\phi}_{12, m}$ and $\hat{\phi}_{23, m}$, obtained from the two sensor pairs are paired and used as the input to the three-sensor based disambiguation algorithm. This results in the unambiguous estimate, $\hat{\phi}_{12, m}$. Note that the disambiguation method in Chapter 5 is presented using delays. Using it for the disambiguation of phases is straightforward.

#### 7.2.1.2 Using Frequency-Diversity-Based Disambiguation

Here only $\hat{\phi}_{12, m}$ are used. These are paired such that each pair will satisfy the identifiability condition for the frequency-based disambiguation. An ambiguous phase-difference may have multiple pairs from different frequency bins. Therefore, the results can be combined in different ways depending on the size of the available data. For example, the mean of the estimates can be taken as the final estimate. However, due to the probability of the occurrence of outliers, the statistical mode may be more useful in suppressing the contribution of the latter effect in the final result. It may be more practical to take the mode of the estimates of the integer $\hat{k}_{12}$ from all the pairs, and then calculate the final phase-difference estimate. The latter approach will be adopted and hence
the process of disambiguation of the phase-difference $\hat{\phi}_{12,u}$ (at frequency $\omega_u$) can be summarized as

$$\hat{\phi}_{12,u} = \hat{\phi}_{12,u} + 2\pi \text{mode}(\hat{k}_{12,v}); v = 0, ..., M - 1; |\omega_u - \omega_v| \leq \frac{c}{d_{12}}. \quad (7.7)$$

Having obtained $\hat{\phi}_{12,m}, \forall m$, using any of the two disambiguation methods, Eq. (7.4) can be used to estimate the time-delay, then Eq. (7.5) is used to determine the AOA. In this chapter, the LS estimator of the time-delay will be adopted and all the weights in Eq. (7.4) will be set equal to one.

Motivated by the fact that the LS approach for time-delay estimation works poorly when outliers occur (the estimator simply weighs all the data equally), another approach is sought. The new approach again relies on taking the mode of the data to protect against outliers. Namely, the estimation process of the time-delay is carried out according to

$$\hat{\tau}_{12} = \text{mode}(\hat{\tau}_{12,m}), m = 1, ..., M - 1 \quad (7.8)$$

where $\hat{\tau}_{12,m}$ are obtained simply by dividing $\hat{\phi}_{12,m}$ by the corresponding radian frequencies. It should be noted here that, since the delays $\hat{\tau}_{12,m}$ have a continuous distributions, the data needs to be grouped into bins, where each bin coincides with an interval of the occurrence of the variable of interest. These groups of bins construct the histogram of the data. The mode of the data can be obtained from the histogram—the mode will be the mid-value of the bin that coincides with the peak of the histogram. The bin width should be sufficiently small to allow for reasonable precision. On the other hand, for the discrete data in Eq. (7.7), the mode can be obtained in a more straightforward way [102].

### 7.2.2 Parameter Consideration

Based on the discussions in the two previous chapters, in the case of a deterministic signal, and when sources of error other than noise are ignored, the variances pertaining to phase errors, $\nu_{12,m}$ and $\nu_{23,m}$ are equal and are approximately given by

$$\sigma_m^2 = \frac{1}{2N_f} \left\{ \frac{2\alpha(\omega_m) L\Lambda + 1}{\alpha(\omega_m)L\Lambda} \right\}. \quad (7.9)$$

Note that Eq. 7.9 is obtained for the added white Gaussian noise (AWGN) case, which will be assumed in the following discussion. All of the parameters that
determine \( \sigma^2_m \), excluding the SNR \((\Lambda)\), are controllable (in some way) and can be tuned to enhance performance. The effect of these parameters is as follows:

- The number of frames, \( N_f \): \( \sigma^2_m \) is inversely proportional to \( N_f \). Therefore, increasing the number of frames is desirable. The number of frames might be limited by a target system update rate, or by some signal characteristics, e.g., stationarity [16]. For deterministic signals, the latter consideration is irrelevant.

- Length of the CPS, \( L \): \( L \) is directly determined by the DFT length \((L \) is equal to half or all of the DFT length depending on whether the spectra are single-sided or double-sided\). The variance \( \sigma^2_m \) decreases as \( L \) increases due to the dominance of the squared term in the denominator versus the linear term in the numerator. Therefore, \( L \) should be kept as large as possible. In general, it is also desirable to increase \( L \) to obtain reasonable frequency resolution. The latter is more important for the frequency-diversity-based methods. Poor resolution may mean that no frequency pairs that satisfy the identifiability condition exist. Limiting factors for increasing \( L \) could be the update rate and the computational complexity of the DFTs. Moreover, it can be mutually exclusive to optimize over the values of both the parameters \( L \) and \( N_f \).

- The fractional power, \( \alpha(\omega_m) \): The effect of \( \alpha(\omega_m) \) is similar to that of \( L \). Frequency bins with higher \( \alpha(\omega_m) \) are expected to yield estimates with higher quality than those frequency bins with lower power. Therefore, on forming the final estimate (e.g., using Eq. (7.4)), the contributions from those frequency bins with higher power should be more accentuated, for example, by giving them more weight when the results are averaged. Note that the best linear unbiased estimator accounts for that by weighting the results by the reciprocals of the variances. The least squares weighing will be optimal only when all of the frequencies incorporated in the process have equal powers. For FHSS signals, due to the peaky CPS, one may consider employing only those estimates obtained from the frequency bins that coincide with the peaks. This will also help in reducing the computational cost.
7.3 AOA Estimation: the Multi-Source Case

When more than one source is present, the signal model becomes

\[
\begin{align*}
x_1(nT_s) &= \sum_{i=1}^{N_b} s(nT_s) + w_1(nT_s) \\
x_2(nT_s) &= \sum_{i=1}^{N_b} s(nT_s - \tau_{12,i}) + w_2(nT_s) \\
x_3(nT_s) &= \sum_{i=1}^{N_b} s(nT_s - \tau_{12,i} - \tau_{23,i}) + w_3(nT_s)
\end{align*}
\] (7.10)

where \(N_b\) is the number of sources received at each sensor; \(\tau_{12,i}\) and \(\tau_{23,i}\) are the time-delays for source \(i\).

In such a multi-source scenario, Eq. (7.2) cannot always be used to estimate the CPS, simply due to the incoherence of the mixtures. Addition of frames containing multiple signals with different time-delays leads to phase distortions in the general case. Therefore, in the multi-source case, the CPS can only be estimated using a single frame (i.e., \(N_f = 1\)) (e.g., see [17]). This leads to increased variances due to the number of frames being reduced. In fact, poor estimates of the CPS result in this case that do not have any advantage in mitigating the effect of noise (the effect of noise is reduced through averaging). In the case of FHSS signals, AOA estimation can be confined to the spectral peaks to increase the parameter \(\alpha(\omega_m)\). Since a substantial proportion of the FHSS signal power concentrates around these peaks, \(\alpha(\omega_m)\) can be sufficiently high to produce good estimates even when \(N_f = 1\). Further, estimation error can be reduced by fusing multiple data from different frequencies.

Due to the presence of multiple sources, any CPS will represent a mixture of the CPS of the individual sources plus noise (this is due to the linearity of the DFT [42]). To obtain the phase-difference estimates for each source, the orthogonality of FHSS signals in the frequency (or time-frequency) domain is exploited. Namely, in the frequency-domain, the probability that two FHSS sources have the same frequency at the same time is very low, and we can approximately write

\[
S_k(\omega_m,i)S_l(\omega_m,i) \approx 0, \forall m, \forall i, \forall \{k,l\} \subset \{1, ..., N_b\}, k \neq l
\] (7.11)

where \(\hat{S}_k(\omega_m,i)\) and \(\hat{S}_l(\omega_m,i)\) are, respectively, the values of the Fourier transforms of the \(i\)'th frames of source \(k\) and source \(l\) at frequency \(\omega_m\). Based on
Eq. (7.11), it can be deduced that, at most, only one source will be active at each instant \((\omega_m, i)\), and consequently

\[
\hat{G}_{x_1 x_2} (\omega_m, i) \approx \hat{G}_{s_a s_a} (\omega_m, i) e^{j\phi_{12,m}}
\]

\[
\hat{G}_{x_2 x_3} (\omega_m, i) \approx \hat{G}_{s_a s_a} (\omega_m, i) e^{j\phi_{23,m}}
\]

(7.12)

where \(a\) is the index of the active source at the instant \((\omega_m, i)\).

As a result of this orthogonality property, the phase-difference estimates from the instant \((\omega_m, i)\) will belong to a single source, which is the active source at that particular instance. This property has effectively been exploited to blindly estimate the AOAs of multiple sources as in [44] and [17], for example. Herein, it will be assumed that the receiver knows the hopping patterns (HPs) of—and is synchronized to—each individual source. Based on this assumption, the receiver can associate phase-difference estimates with different sources (the receiver knows which source is active at which instant). By separating the estimates belonging to each individual source, disambiguation and AOA estimation can be carried out in a similar way to the single-source case. Depending on the size of the DFT and the hop length, the frequency-based disambiguation, in particular, may need phase-difference estimates from more than one frame.

In practice, the orthogonality of FHSS signals is not perfect and the property degrades as the number of sources increases. This is mainly due to two effects. First, collision may occur such that two sources are using the same frequency simultaneously. The second effect is due to the tails from neighboring frequencies from other interfering sources. Both effects can be reduced by judicious signal design, as is described in Chapter 3. The first effect requires that the HPs be designed with as low probability of collision as possible. When collision occurs, the incorporation of multiple data (from several frames) to form the final estimate can greatly help in rejecting outlying estimates resulting from collisions. The second effect can be mitigated by maximizing the orthogonality of the signals at the design stage. Typically, FHSS signals are designed such that the minimum difference between any two hopping frequencies is at least equal to \(1/T_h\), where \(T_h\) is the hop duration. Such a design criterion results in sufficient orthogonality that the effect of the tails from neighboring hops can be ignored, especially, when the estimation is confined to near-peak portions of the CPS.

Depending on the level of approximation in Eq. (7.11) and Eq. (7.12), the discussion concerning the effect of the different parameters in Section 7.2.2, will
be approximately valid in the multi-source case. In other words, due to the orthogonality, the multi-source estimation problem can be viewed as a number of disjoint single-source problems.

7.4 Search-Based AOA Estimation Methods

7.4.1 Spatial Diversity Based

Based on Theorem 5.1, the following result can directly be reached:

Corollary 7.1 For a system of three collinear sensors in the far-field and under noise-free conditions, if the difference of the two smaller inter-sensor spacings is not greater that a half-wavelength of the impinging signal, a pair of principal phase-differences, \( \varphi_{12} \) and \( \varphi_{23} \), is unique for each unique AOA. In other words, for each pair of principal phase-differences, there exists one and only one mapping to the visible AOA range.

The implication of Corollary 7.1 is that a triplet of widely-spaced sensors that satisfies the constraint \( \Delta_d \leq 1 \), spatially samples a signal in such a way that the principal components of phase-difference uniquely identify the AOA. Furthermore, it is straightforward that a spatial sample, when \( \Delta_d \leq 1 \) is satisfied, is unique for each AOA. The latter statement implies that spectral-based and parametric AOA estimation methods (see [2]) can directly be applied on any spatial sample collected at the sensors. However, these methods are not of interest for this work, as discussed in Chapter 2.

To estimate the true AOA when only principal estimates are available, the whole space can be searched for the angle that produces a pair of principal phase-differences that are closest to the observed pairs of principal estimates. This could be the AOA that produces the minimum mean squared error (MSE)—between the observed principal estimates and calculated ones. However, since the MSE criterion has the disadvantage that it heavily weighs outliers [101], herein, the mean absolute error (MAE) is employed instead. Using the MAE, we define the cost function as

\[
J(\theta) = \frac{1}{N_{\varphi}} \sum_{i,j} |\hat{\varphi}_{12,i} - \hat{\varphi}_{12,i}(\theta)| + |\hat{\varphi}_{23,i} - \hat{\varphi}_{23,i}(\theta)|
\]  
(7.13)

where \( \hat{\varphi}_{12,i} \) and \( \hat{\varphi}_{23,i} \) are the pair of principal estimates with index \( i \); \( N_{\varphi} \) is the total number of estimate pairs; \( \hat{\varphi}_{12,i} \) and \( \hat{\varphi}_{23,i} \) are calculated using hypothesized
values of $\theta$ according to

$$
\begin{align*}
\phi_{12,i}(\theta) &= \phi_{12,i}(\theta) - 2\pi k_{12,i}(\theta) = \frac{\omega_i d_{12} \sin(\theta)}{c} - 2\pi \Psi_0 \left[ \frac{\omega_i d_{12} \sin(\theta)}{2\pi c} \right] \\
\phi_{12,i}(\theta) &= \phi_{23,i}(\theta) - 2\pi k_{23,i}(\theta) = \frac{\omega_i d_{23} \sin(\theta)}{c} - 2\pi \Psi_0 \left[ \frac{\omega_i d_{23} \sin(\theta)}{2\pi c} \right]
\end{align*}
$$

(7.14)

where $\omega_i$ is the frequency corresponding to $\phi_{12,i}$ and $\phi_{23,i}$; $\Psi_0(\cdot)$ is a rounding function that has the tie-breaking rule of rounding to the nearest integer towards zero.

Now, the AOA can be estimated according to

$$
\hat{\theta} = \arg \min_{\theta} J(\theta)
$$

(7.15)

In the noise-free case, the minimum value of $J(\theta)$ is expected to be perfectly zero in the far-field case. Fig. 7.1 shows examples of $J(\theta)$ for an acoustic frequency of 40 kHz received at an AOA of 0°. The figure is plotted from a pair of principal phases under noise-free conditions. The figure shows that the minimum of $J(\theta)$ is perfectly zero for all sensor separations. It can be seen that the larger sensor spacings are associated with an increasing number of local minima, with those local minima located close to the global minima having relatively low amplitudes (the total number of minima is actually proportional to the inter-sensor spacing in half-wavelengths). This is expected to give rise to errors in the presence of noise whereby triplets with larger sensor separations are expected to exhibit less reliability. This is similar to the effect demonstrated in Chapter 5, and can also be related to array beam-patterns and grating lobes as in Chapter 2. The figure also confirms the fact that increasing the inter-sensor spacing has the effect of increasing the resolution as reflected in the form of sharper peaks associated with the larger spacings. A trade-off between reliability (or success) and resolution is clearly demonstrated.

It should be noted here that due to the effect of the deviation from the ideal far-field model that has been discussed in Chapter 5, the emitter range should be restricted in some way to avoid disambiguation failure for a particular configuration.

In the near-field case, since Eq. (7.13) assumes a far-field, deviations may occur that lead to AOA estimates that differ significantly from the actual AOA, in the
same manner that has been referred to as failure in Chapter 5. The method is
however, expected to produce reasonable results when the emitter is not too
close to the receivers, with the definition of the closeness determined by the
sensor configuration.

The minimization of (7.15) can simply be carried out by calculating \( J(\theta) \) for
samples of \( \theta \) covering the whole interval \([-90^\circ, 90^\circ]\). The interval can be sam-
pled using a constant step, the value of which determines the AOA resolution
that can be obtained. An efficient search can be implemented using a relati-
vely large step to obtain a rough estimate of the location of the minimum. This
course search is then followed by a fine search around the expected location
of the minimum as similar to [98]. The step for the coarse search should be
sufficiently small so as to avoid errors due to local minima.

### 7.4.2 Frequency Diversity Based

The following corollary is based on Theorem 6.1:

**Corollary 7.2** In the noise-free case, given a pair of principal values of phase-
difference from two different frequencies such as \( \varphi_{12,u} \) and \( \varphi_{12,v} \), if the absolute diffe-
rence of the two frequencies satisfies \( |\omega_u - \omega_v| < \pi c / d_{12} \), there exists one and only
one real AOA, \( \theta \), that can produce the pair of principal estimates.

Based on this corollary, \( \theta \) can be estimated in the same manner as in Sec-
tion 7.4.1, with the cost function defined here as

\[
J(\theta) = \frac{1}{N_{\phi}} \sum_{<u,v>} |\hat{\varphi}_{12,u} - \varphi_{12,u}(\theta)| + |\hat{\varphi}_{12,v} - \varphi_{12,v}(\theta)|. \tag{7.16}
\]

The summation above may comprise some or all of the principal phase-
difference estimates from the frequency pairs that satisfy the identifiability
condition. The properties of the cost function \( J(\theta) \) are similar to those of the
counterpart cost function for the spatial-diversity-based method depicted in
Fig. 7.1.

### 7.5 The Effect of Cycle Slipping

This section is a note on an important issue that can affect the disambiguation
and AOA estimation processes. In Chapter 5 and Chapter 6, this effect has
Figure 7.1: The error function $J(\theta)$ (normalized) for an AOA equal to $0^\circ$ for selected inter-sensor spacings (given in half-wavelengths). Far-field is assumed.
been ignored. In this section, the effect will be studied and ways to circumvent the problem will be highlighted.

Let us focus on the phase-difference estimates from sensor-1 and sensor-2. The same discussion similarly applies to the estimates from the other sensor pair. Considering cycle slipping, \( \hat{\phi}_{12,m} \) in Eq. (7.3), can be written as

\[
\hat{\phi}_{12,m} = \omega_m \tau_{12} + v_{12,m} = (\varphi_{12,m} + \varepsilon_{12,m}) + 2\pi(k_{12,m} + l_{12,m})
\]  

(7.17)

where \( (\varphi_{12,m} + \varepsilon_{12,m}) \in [-\pi, \pi] \) represents the principal estimate \( \hat{\phi}_{12,m} \). and \( l_{12,m} \in \mathbb{Z} \) is due to cycle slipping. Where no cycle slipping takes place, the integer \( l_{12,m} \) will be equal to zero and \( \varepsilon_{12,m} \) will be equivalent to \( v_{12,m} \). To derive the probability of the occurrence of cycle slipping, consider the noise-free phase-difference

\[
\phi_{12,m} = \omega_m \tau_{12} = \varphi_{12,m} + 2\pi k_{12,m}
\]  

(7.18)

where \( \varphi_{12,m} \in [-\pi, \pi] \). Comparing Eq. (7.17) and Eq. (7.18), the probability of cycle slipping due to noise can be defined as the probability that the sum of the principal phase, \( \varphi_{12,m} \) and the noise term \( v_{12,m} \) falls outside the interval \( \in [-\pi, \pi] \). This can be expressed mathematically as

\[
P_{sc,m} = P(\varphi_{12,m} + v_{12,m} < -\pi) + P(\varphi_{12,m} + v_{12,m} > \pi)
\]  

(7.19)

where \( P_{sc,m} \) is the probability of the occurrence of at least a single cycle slip at frequency \( \omega_m \). To facilitate the analysis, let us focus on the single-source case. When Eq. (7.2) is used to compute the CPS for a sufficiently large value of \( N_f \), the disturbances \( v_{12,m} \) are approximately Gaussian [20, 22]. It will be assumed that the latter assumption concerning the gaussianity of \( v_{12,m} \) holds. Based on this assumption, \( P_{sc,m} \) can be expressed as

\[
P_{sc,m} = 2 - \Phi_{12,m}(\pi - \varphi_{12,m}) - \Phi_{12,m}(\pi + \varphi_{12,m})
\]  

(7.20)

where \( \Phi_{12,m} \) is the CDF of \( v_{12,m} \). It is clear that \( P_{sc,m} \) increases as \( |\varphi_{12,m}| \) approaches \( \pi \). When \( |\varphi_{12,m}| \) is exactly equal to \( \pi \), \( P_{sc,m} \) reaches its maximum value of 0.5, regardless of the variance of the variable \( v_{12,m} \) (being Gaussian with zero mean is sufficient to obtain the latter result). The directions that
coincide with this maximal probability of cycle slipping can be identified via

$$\theta_{cs,m} = \arcsin \left[ \frac{c}{d_{12}} \left( \frac{\pi + 2\pi i}{\omega_m} \right) \right], i = 0, \pm 1, \ldots, \pm N_\pi$$  \hspace{1cm} (7.21)$$

where $N_\pi = \text{floor}(2d_{12}/\lambda)/2 - 1$ when $\text{floor}(2d_{12}/\lambda)$ is even; and $N_\pi = [\text{floor}(2d_{12}/\lambda) - 1]/2$ when $\text{floor}(2d_{12}/\lambda)$ is odd.

Fig. 7.2 shows examples for the probability of cycle slipping in space for different frequencies. The figure is plotted assuming zero-mean gaussian disturbance with variance equal to $3.844 \times 10^{-5}$ (an overly pessimistic value according to physical measurements). To improve visualization, the values of the probability of cycle slipping are thresholded by a value of 0.01; those values above the threshold are set equal to 0.5 (the maximum possible value for the probability of cycle slipping). The probability of cycle slipping is seen to approach zero in most of the space, excluding very small regions coinciding with the angles given by Eq. (7.21). It can be seen that the angles coinciding with the cycle slipping phenomenon differ from frequency to frequency. Albeit, the differences are small close to the boresight of the array. Table 7.1 lists the angles given by Eq. (7.21) for a number of frequencies that are more closely spaced than those in Fig. 7.2. It can be concluded that the angles coinciding with the cycle slipping phenomenon are generally different for different frequencies. Therefore, frequency diversity can be exploited to handle the problem. Note that spatial diversity provided by additional sensor(s) can also be exploited in a similar manner. However, since there is more frequency diversity than spatial diversity in the system of interest (only a limited number of sensor pairs versus many FHSS frequencies), the former property is more adequate to be exploited.

### 7.6 Experimental Results

The various proposed approaches for AOA estimation discussed in this chapter were tested experimentally in a real office environment (see Fig. 7.4). Ultrasonic FHSS signals in the frequency range 35 kHz–49.5 kHz were used (this frequency range coincides with $\lambda_{\text{min}}/2 \approx 3.6$ mm and $\lambda_{\text{max}}/2 \approx 4.9$ mm). The signals were sampled at approximately 168 kHz. In all cases, the frame and DFT lengths were 512, with zero overlap between successive frames. Fig. 7.4 shows the locations of the transmitters and the receiver during the tests, with the heights given in brackets. The receiver was situated in locations such that
7.6 EXPERIMENTAL RESULTS

Figure 7.2: The probability of cycle slipping in space for four different frequencies. The distance between the two sensors is $d_{12} = 2$ cm. The sensor baseline is aligned along the x-axis, with the center of the baseline positioned at the origin. Resolution is 1 cm.

it might catch real effects, e.g., reverberation.

7.6.1 Single-Source

The transmitter and the receiver (containing the sensors) were located at the positions marked Tx3 and Rx1, respectively. The FHSS signal was composed of 20 equally-spaced frequencies with a hop duration of 3.0 ms. At each hop, data was modulated using binary frequency-shift keying (BFSK) applied on randomly generated bits at a data rate equal to the hopping rate. The angles were estimated from the CPS phase after performing disambiguation. The LS and the mode estimators were used in both cases; when the spatial-diversity-based (SDB) disambiguation was applied, and when the frequency-diversity-based
Table 7.1: Cycle slipping angles for different acoustic frequencies and $d_{12} = 2$ cm.

<table>
<thead>
<tr>
<th>f (kHz)</th>
<th>Cycle slipping angles in degrees</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>12.38 40.02 ---</td>
</tr>
<tr>
<td>41</td>
<td>12.07 38.86 ---</td>
</tr>
<tr>
<td>42</td>
<td>11.78 37.77 ---</td>
</tr>
<tr>
<td>43</td>
<td>11.50 36.74 85.63</td>
</tr>
<tr>
<td>44</td>
<td>11.24 35.78 77.02</td>
</tr>
</tbody>
</table>

The (FDB) disambiguation method was applied. The alternative approaches using grid-search (GS) were applied on the ambiguous CPS phases to directly obtain the AOA.

Table 7.2–7.5 present the results that were obtained for four different angles, three different sensor configurations and six different methods. The root mean square error (RMSE) is used as the performance metric. The RMSEs were estimated from 10 independent tests for each angle. Each test involved $10^4$ signal snapshots. All the results that fall outside the visible AOA range were set equal to either $-90^\circ$ or $90^\circ$, depending on the sign of the corresponding time-delay. For the SDB methods, each configuration was augmented with a third sensor placed such that $\Delta_d = 2.5$ mm. The following can be observed from Table 7.2–Table 7.5:

- In most cases the grid-search methods, SDB-GS and FDB-GS, give the best performance. The performance of these methods does not exhibit a notable outlier effect. The only cases where performance can be linked to outliers is for the SDB-GS method for an angle of $20^\circ$, and slightly for $60^\circ$. Except for these two cases, the SDB-GS and FDB-GS methods perform nearly the same in most cases.

- The mode based methods are generally better than the LS methods. The advantage of taking the mode of the time-delays in suppressing outliers is clearly visible. The standard LS estimator seems to be much affected by the occurrence of outliers, and is hence less suitable for situations such as when the sensors are widely-spaced and the occurrence of outliers is more likely to happen (compared to when standard sensor spacing is used). Therefore, the LS estimator will not be of much interest for this work.

- The SDB methods portray the effect of outliers more clearly. This is expected due to the effect of the deviation from the ideal far-field assump-
Table 7.2: Experimental results: The RMSE (in degrees) for $\theta = 0^\circ$.

<table>
<thead>
<tr>
<th>Method</th>
<th>$d_{12} = 10$ mm</th>
<th>$d_{12} = 15$ mm</th>
<th>$d_{12} = 17.5$ mm</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDB-LS</td>
<td>1.2</td>
<td>24.8</td>
<td>21.7</td>
</tr>
<tr>
<td>FDB-LS</td>
<td>1.9</td>
<td>0.3</td>
<td>2.3</td>
</tr>
<tr>
<td>SDB-mode</td>
<td>1.9</td>
<td>13.2</td>
<td>26.4</td>
</tr>
<tr>
<td>FDB-mode</td>
<td>0.9</td>
<td>0.4</td>
<td>2.1</td>
</tr>
<tr>
<td>SDB-GS</td>
<td>0.6</td>
<td>1.3</td>
<td>0.1</td>
</tr>
<tr>
<td>FDB-GS</td>
<td>1.4</td>
<td>0.3</td>
<td>2.2</td>
</tr>
</tbody>
</table>

7.6 EXPERIMENTAL RESULTS

- The best overall method, in terms of performance, is the FDB-GS. The superiority of this method can be attributed to two facts. First, the method is independent of any spatial parameters, and is hence not much affected by the deviation from the ideal far-field model. Second, exhaustive grid-search is naturally more robust than other approaches, though it is computationally more complex.

7.6.1.1 Calibration Error Effect

For the same single-source scenario, a test was performed to study the effect of calibration errors on the performance of the proposed method. The purpose was to see how robust the proposed method is to calibration errors. Random noise was added to the measurements of the distances, $d_{12}$ and $d_{23}$, to give noisy measurements, $d_{12} + e_{12}$ and $d_{23} + e_{23}$. The noise terms, $e_{12}$ and $e_{23}$, were obtained as realizations of two random variables uniformly distributed in the interval $[-B, B]$. The value of $B$ was varied and for each value of $B$, the proposed method was applied. The process was repeated 100 times for each $B$ allowing for new random values of $e_{12}$ and $e_{23}$ to be tested. The RMSE corresponding to each $B$ was estimated from the 100 trials. The tests were carried out for the six AOA methods discussed above.

As an example, Fig. 7.3 shows the variation of the RMSE in degrees with $B$. The figure corresponds to the $\theta = 20^\circ$ scenario whose primary results are reported in Table 7.4. It can be seen that the SDB-LS and SDB-mode methods work only under perfect calibration. The rest of the methods exhibit robustness to calibration errors and their respective RMSEs rise slightly as calibration error increases.
Table 7.3: Experimental results: The RMSE (in degrees) for $\theta = 20^\circ$.

<table>
<thead>
<tr>
<th>Method</th>
<th>$d_{12} = 10$ mm</th>
<th>$d_{12} = 15$ mm</th>
<th>$d_{12} = 17.5$ mm</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDB-LS</td>
<td>0.1</td>
<td>7.8</td>
<td>34.7</td>
</tr>
<tr>
<td>FDB-LS</td>
<td>1.9</td>
<td>5.8</td>
<td>3.2</td>
</tr>
<tr>
<td>SDB-mode</td>
<td>2.0</td>
<td>0.1</td>
<td>62.5</td>
</tr>
<tr>
<td>FDB-mode</td>
<td>2.0</td>
<td>0.5</td>
<td>0.6</td>
</tr>
<tr>
<td>SDB-GS</td>
<td>0.8</td>
<td>0.4</td>
<td>12.4</td>
</tr>
<tr>
<td>FDB-GS</td>
<td>1.1</td>
<td>0.3</td>
<td>0.8</td>
</tr>
</tbody>
</table>

Table 7.4: Experimental results: The RMSE (in degrees) for $\theta = 45^\circ$.

<table>
<thead>
<tr>
<th>Method</th>
<th>$d_{12} = 10$ mm</th>
<th>$d_{12} = 15$ mm</th>
<th>$d_{12} = 17.5$ mm</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDB-LS</td>
<td>2.0</td>
<td>2.9</td>
<td>5.7</td>
</tr>
<tr>
<td>FDB-LS</td>
<td>3.9</td>
<td>7.8</td>
<td>1.4</td>
</tr>
<tr>
<td>SDB-mode</td>
<td>2.2</td>
<td>2.5</td>
<td>2.4</td>
</tr>
<tr>
<td>FDB-mode</td>
<td>2.2</td>
<td>2.5</td>
<td>2.4</td>
</tr>
<tr>
<td>SDB-GS</td>
<td>2.0</td>
<td>2.3</td>
<td>1.9</td>
</tr>
<tr>
<td>FDB-GS</td>
<td>2.7</td>
<td>2.5</td>
<td>2.1</td>
</tr>
</tbody>
</table>

Table 7.5: Experimental results: The RMSE (in degrees) for $\theta = 60^\circ$.

<table>
<thead>
<tr>
<th>Method</th>
<th>$d_{12} = 10$ mm</th>
<th>$d_{12} = 15$ mm</th>
<th>$d_{12} = 17.5$ mm</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDB-LS</td>
<td>2.5</td>
<td>6.2</td>
<td>13.4</td>
</tr>
<tr>
<td>FDB-LS</td>
<td>1.5</td>
<td>3.9</td>
<td>2.9</td>
</tr>
<tr>
<td>SDB-mode</td>
<td>1.9</td>
<td>0.1</td>
<td>28.4</td>
</tr>
<tr>
<td>FDB-mode</td>
<td>1.9</td>
<td>0.6</td>
<td>0.1</td>
</tr>
<tr>
<td>SDB-GS</td>
<td>4.0</td>
<td>1.8</td>
<td>0.9</td>
</tr>
<tr>
<td>FDB-GS</td>
<td>1.5</td>
<td>1.8</td>
<td>0.9</td>
</tr>
</tbody>
</table>

Figure 7.3: Calibration error effect.
7.6 EXPERIMENTAL RESULTS

7.6.2 Multi-Source

In the multi-source case, it was assumed that the receiver was synchronized to the HPs of each source. The AOA for each source was estimated from the peaks of CPS of signals containing 14 hops. This means that only 14 pairs of phase-difference estimates were used to form the final AOA estimate for each source. The CPS to estimate the AOA for each source were calculated separately to ensure that the analysis window was properly aligned with the hops of each respective source. The transmitters’ and receiver’s locations (Tx1, Tx2, Tx3 and Rx2, respectively) are indicated in Fig. 7.4.

Table 7.6 summarizes the results. Each entry was estimated from 10 independent tests. It is clear that only the grid-search approaches have produced meaningful results in this case. The other four approaches are out of the question since the results obtained thereby are far from the true values, as reflected by the RMSE. The deterioration in performance herein, compared to the single-source case, is attributed mainly to the small number of (phase-difference) estimates incorporated in estimating the final AOA in each case. The effect of the signals on one another is an additional source of performance loss. However, as has been demonstrated, search based methods are sufficiently robust in these situations. The SDB-GS results are apparently more accurate than the FDB-GS results, however, this is not conclusive since the variation is not so large. For Example, calibration errors may introduce such performance differences.

Table 7.6: Experimental results: The RMSE (in degrees) for 3 sources at $\theta = -43^\circ$, $\theta = -47^\circ$ and $\theta = 31^\circ$ (corresponding to location Tx1, Tx2 and Tx3 in Fig. 7.4).

<table>
<thead>
<tr>
<th>Method</th>
<th>$d_{12} = 15.0$ mm</th>
<th>$d_{12} = 17.5$ mm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>source 1</td>
<td>source 2</td>
</tr>
<tr>
<td>SDB-LS</td>
<td>30.7</td>
<td>34.8</td>
</tr>
<tr>
<td>FDB-LS</td>
<td>23.8</td>
<td>41.8</td>
</tr>
<tr>
<td>SDB-mode</td>
<td>25.9</td>
<td>29.6</td>
</tr>
<tr>
<td>FDB-mode</td>
<td>25.7</td>
<td>33.3</td>
</tr>
<tr>
<td>SDB-GS</td>
<td>2.8</td>
<td>2.5</td>
</tr>
<tr>
<td>FDB-GS</td>
<td>2.5</td>
<td>3.1</td>
</tr>
</tbody>
</table>
7.7 Summary

In this chapter, a variety of methods for AOA estimation from the phase of the CPS were discussed. The methods employ the two disambiguation methods proposed in the two previous chapters to achieve ambiguity-free AOA estimation. The chapter considered both the single-source and multi-source cases. Parameters that affect the accuracy of the AOA estimation were discussed. The phenomenon of cycle slipping was studied, locations where the effect is significant were pinpointed, and a way to reduce the potential of the effect in causing failure was highlighted.

At the end of the chapter, experimental results were presented for both the single-source and multi-source scenarios. From the results, it was found that some of the proposed methods exhibit substantially better performance than the other ones, with the best of all the methods identified as a grid-search method based on the frequency-diversity-based disambiguation method. Also, the effect of calibration errors on performance was considered. The proposed methods were found to tolerate moderate calibration errors, with variation in the level of this tolerance among the different methods.
This chapter discusses the problem of joint location and orientation estimation of a receiver using only angle-of-arrival (AOA) information. Conventional formulations of the problem consist of a number of nonlinear equations where the number of unknowns exceeds the number of equations. Formulations presented in this chapter simplify the problem in a way that leads to efficient solutions. Two solutions are presented and their performance is compared using simulation.

8.1 Introduction

In this chapter, the problem of receiver’s location and orientation estimation using time-difference of arrival (TDOA) or angle-of-arrival (AOA) information is considered. By TDOA we mean the difference in the delay of a signal received by two sensors that are part of the receiving device. The techniques presented in this chapter, despite focusing more on indoor location-orientation for ubiquitous computing systems, are quite generic and can be applied to various situations in which receiver’s location and orientation are of interest.

In this chapter, we present methods showing that TDOA/AOA information is sufficient for determining both the location and orientation of the receiver in 3D. It should be mentioned that, a similar location-orientation problem for mobile robot was discussed in the literature (e.g., [121]), however it is limited to 2D only. Adding a third dimension increases the complexity of the problem significantly.

In the following section, a general description of the problem is given. Section 8.3 starts by formulating the solution in the most straightforward way.
The discussion reveals that using a straightforward formulation leads to inefficient solutions. In Section 8.3, an alternative formulation is proposed. A general formulation is given that reduces the number of required beacons compared to the straightforward formulation. The section elaborates by proposing two different solutions for the location and orientation estimation problem based on the general formulation. The two solutions are compared in Section 8.4 in simulation. Simulations confirm the feasibility of the proposed approach. Conclusions are given in Section 8.5.

8.2 Problem Description

The scenario considered in this chapter assumes \( N_b \) beacons with fixed and known locations. The receiver contains \( N_s \) sensors that are used to provide TDOA/AOA information with a certain degree of accuracy. Angles are measured relative to the device’s local axes. The goal is to find the location and orientation of the receiver in 3D based on the available TDOA/AOA information only. The location of a 3D object with non-trivial dimensions can be thought of as the location of a point on that object. Orientation can be represented by the global unit vectors corresponding to the directions of each of the receiver’s local Cartesian axes (\( \bar{X}, \bar{Y}, \bar{Z} \) as depicted in Fig. 8.1). The difficulty of the problem stems from the fact that with unknown receiver orientation, AOA information becomes ambiguous since the frame of reference is not fixed.

8.3 Proposed Solutions

The most straightforward formulation of the problem is to use two sensors. In such a case, the location-orientation problem is transformed into a problem of locating the two sensors. The TDOAs of \( N_b \) signals measured between these two sensors result in the following set of equations:

\[
\| \mathbf{b}_k - \mathbf{s}_1 \| - \| \mathbf{b}_k - \mathbf{s}_2 \| = c \tau_{12,k}, \quad k = 1, ..., N_b
\]  

(8.1)

where \( \mathbf{s}_1 \) and \( \mathbf{s}_2 \) are the position vectors of the two sensors; \( \mathbf{b}_k \) is the position vector for beacon \( k \); \( \tau_{12,k} \) is the TDOA of the signal from beacon \( k \) at the two sensors; \( c \) is the speed of signal propagation; and \( \| . \| \) denotes the Euclidean norm. From (8.1), it can be seen that solving for \( \mathbf{s}_1 \) and \( \mathbf{s}_2 \) requires at least
8.3 PROPOSED SOLUTIONS

six independent equations (which requires at least six non-collinear beacons). The equations are nonlinear and their solution is found to be quite involved. Numerical solutions are found to be inefficient with convergence possible only when the initial guess is very close to the true solution. Due to this complication, alternative approaches are sought.

8.3.1 Solutions Using Four Sensors

In this subsection, using four coplanar sensors on the receiver device is proposed as a solution to the location-orientation problem. Precisely, we assume two orthogonal pairs of sensors as depicted in Fig. 8.1. The purpose of this configuration is to provide 3D AOA information. The receiver’s local coordinate system is assumed to coincide with the baselines of these two pairs of sensors (see Fig. 8.1). Based on the TDOA information provided by these sensors, one can formulate the problem in a similar way to Eq. (8.1). Four non-collinear sensors typically provide three distinct TDOAs. Consequently to solve for all the sensor locations requires at least 12 equations from four non-collinear beacons. Solving the latter equations is more tedious than solving the equations in (8.1). Therefore, an alternative approach will be considered.

The approach adopted in this subsection is based on converting TDOA information to AOA information. It is well known that TDOA estimates can be converted to AOA estimates based on the assumption that the emitter is sufficiently far from the sensors (the far-field assumption). In such a case, the hyperboloid defined by TDOA can be approximated by a cone by considering only the asymptotes of the hyperboloid [78]. Based on this assumption, the two TDOA equations from the two orthogonal pairs of sensors define the unit vector of the line of sight (LOS) of each beacon from the centroid of the sensor positions (which coincides with the origin in Fig. 8.1). The unit vector to each beacon can hence be expressed as

\[ \mathbf{u}_k = [\cos \alpha_k \, \cos \beta_k \, \cos \gamma_k]^T, \quad k = 1, \ldots, N_b \] (8.2)

where \( \alpha_k, \beta_k \) and \( \gamma_k \) are the angles with the three local axes \( \tilde{X}, \tilde{Y} \) and \( \tilde{Z} \), respectively (see Fig. 8.1). The values of \( \cos \alpha_k \) and \( \cos \beta_k \) are determined from
TDOAs using

$$\cos \alpha_k = \frac{c \tau_{12,k}}{|| s_2 - s_1 ||}$$

$$\cos \beta_k = \frac{c \tau_{34,k}}{|| s_4 - s_3 ||}$$

(8.3)

where \( \tau_{12,k} \) and \( \tau_{34,k} \) are the TDOAs of the signal from beacon \( k \), between each of the orthogonal pairs of sensors (see Fig. 8.1). The value of \( \cos \gamma_k \) is determined from \( \cos \alpha_k \) and \( \cos \beta_k \) using

$$\cos \gamma_k = \sqrt{1 - \cos^2 \alpha_k - \cos^2 \beta_k}.$$  

(8.4)

In (8.4), only the positive root has been considered. This is motivated by the fact that the maximum value that \( \gamma_k \) can take, without loss of LOS due to the panel where the sensors are mounted, is \( \pi/2 \), as illustrated in Fig. 8.2. It should be noted that \( \bar{u}_k \) in (8.2) are estimates, however, for simplicity of notations, the estimation symbol “\(^\hat{\}\)" is ignored in this case and also in the subsequent discussions.

### 8.3.1.1 Solution 1

Now, taking the unit vectors \( u_k \) pairwise and performing the dot product, the following set of equations is obtained:

$$\begin{align*}
(b_i - m) \cdot (b_j - m) &= ||b_i - m|| ||b_j - m|| \bar{u}^T_i \bar{u}_j, \\
\{i, j\} &\subset \{1, ..., N_b\}, i \neq j
\end{align*}$$

(8.5)
8.3 PROPOSED SOLUTIONS

Figure 8.2: Limitation of the angle $\gamma$. Plain arrows point to the direction of LOS while the arrows with squares point to the direction of NLOS (Non LOS).

where $\mathbf{m} \triangleq [x, y, z]^T$ is the position vector of the centroid of the sensors positions, and is the center of the localization problem; “.” is the dot product operator; and

$$\bar{\mathbf{u}}_i^T \bar{\mathbf{u}}_j \equiv \cos \theta_{ij} \quad (8.6)$$

with $\theta_{ij}$ being the angle between the two LOS lines corresponding to beacon $i$ and beacon $j$ (see $\theta_{12}$ in Fig. 8.1). Eq. (8.5) represents a system of nonlinear equations in $\mathbf{m}$. Since $\mathbf{m}$ is a triad, a minimum of three beacons (i.e. three equations) are required to find the location of the receiver. Eq. (8.5) has the advantage that the unknowns pertaining to the receiver orientation are excluded from the formulation, and the orientation problem can hence be handled subsequently using the location information obtained from the solution of (8.5). This allows for location and orientation determination using only three beacons.

Now, for simplicity, assume that $N_b = 3$, Eq. (8.5) can be written as

$$f_{ij}(\mathbf{m}) = 0, \{i, j\} \subset \{1, ..., 3\}, i \neq j \quad (8.7)$$

where

$$f_{ij}(\mathbf{m}) = (\mathbf{b}_i - \mathbf{m}) \cdot (\mathbf{b}_j - \mathbf{m})$$

$$- \parallel \mathbf{b}_i - \mathbf{m} \parallel \parallel \mathbf{b}_j - \mathbf{m} \parallel \bar{\mathbf{u}}_i^T \bar{\mathbf{u}}_j. \quad (8.8)$$

Eq. (8.7) represents three surfaces each of which is formed by rotating a circle arc around its chord. The chords in this case are the line segments between each pair of beacons. One such surface can be the internal surface of a spindle torus [122] (when $\theta_{ij} < \pi/2$), the external surface of a spindle torus (when $\theta_{ij} > \pi/2$), or a sphere (when $\theta_{ij} = \pi/2$). The intersection of three such surfaces is the location of the receiver.
It should be noted that Eq. (8.7), when expanded, is differentiable in in $x$, $y$ and $z$ and can hence be solved using the well-known Newton-Raphson (NR) method [123]. In fact the problem is found to be well suited for solution using the NR method. The method converges in most cases when a good initial guess is available. A good guess is always found to be a one with the three components values taking sufficiently large absolute values.

Having found the location $m$ of the receiver, orientation can readily be determined. Orientation can be defined as a rotation matrix of the coordinate system. The rotation matrix can be defined as

$$ A \triangleq \begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix} = \begin{bmatrix} \cos \theta_{\tilde{x}\tilde{x}} & \cos \theta_{\tilde{y}\tilde{x}} & \cos \theta_{\tilde{z}\tilde{x}} \\ \cos \theta_{\tilde{x}\tilde{y}} & \cos \theta_{\tilde{y}\tilde{y}} & \cos \theta_{\tilde{z}\tilde{y}} \\ \cos \theta_{\tilde{x}\tilde{z}} & \cos \theta_{\tilde{y}\tilde{z}} & \cos \theta_{\tilde{z}\tilde{z}} \end{bmatrix} \quad (8.9) $$

where $\theta_{\tilde{x}\tilde{v}}$ is the angle that the local $w$-axis of the receiver makes with the global $v$-axis. The rows of $A$ are the directions of orientation of the three receiver axes, respectively. A rotation of any vector from the local receiver coordinate system to the global coordinate system can be achieved using

$$ \tilde{u} = Au. \quad (8.10) $$

The matrix $A$ can be determined if at least three direction vectors are known in both the global and the local coordinate systems. Such vectors can be $\tilde{u}_k$ (as in (8.2)) and $u_k$, $k = 1, ..., 3$. The latter vectors can be determined (when $m$ is known) as

$$ u_k = \frac{b_k - m}{\| b_k - m \|}. \quad (8.11) $$

Now define

$$ U \triangleq [u_1 \ u_2 \ u_3]^T, $$
$$ \tilde{U} \triangleq [\tilde{u}_1 \ \tilde{u}_2 \ \tilde{u}_3]^T. \quad (8.12) $$

From (8.10), it directly follows that

$$ \tilde{U} = AU \quad (8.13) $$

and consequently

$$ A = \tilde{U}U^{-1}. \quad (8.14) $$

From (8.14), it can be observed that $A$ can be determined only if $U$ is non-
singular, a condition that can be satisfied only if the beacons and the receiver do not lie on a plane where one of the three coordinates is a constant. Nevertheless, it is found that, even in the latter case, $A$ can be determined by exploiting the orthogonality of its columns. In the sequel, we determine $A$ when the beacons and the receiver lie on a plane $z = \text{constant}$. In this case, column 3 of the matrix $U$ is expected to contain all zero values. The consequence of this is to eliminate $\cos \theta_{zx}$, $\cos \theta_{zy}$ and $\cos \theta_{zz}$ from Eq. (8.13). Manipulating Eq. (8.13) (when $z = \text{constant}$) results in the following three systems of equations:

$$
Qa_x = q_x \\
Qa_y = q_y \\
Qa_z = q_z
$$  \hspace{1cm} (8.15)

where

$$
\begin{align*}
a_x &\triangleq \begin{bmatrix}
\cos \theta_{xx} \\
\cos \theta_{yx}
\end{bmatrix} \\
a_y &\triangleq \begin{bmatrix}
\cos \theta_{xy} \\
\cos \theta_{yy}
\end{bmatrix} \\
a_z &\triangleq \begin{bmatrix}
\cos \theta_{xz} \\
\cos \theta_{yz}
\end{bmatrix} \\
q_x &\triangleq \begin{bmatrix}
\bar{u}_{1,x} \\
\bar{u}_{2,x} \\
\bar{u}_{3,x}
\end{bmatrix} \\
q_y &\triangleq \begin{bmatrix}
\bar{u}_{1,y} \\
\bar{u}_{2,y} \\
\bar{u}_{3,y}
\end{bmatrix} \\
q_z &\triangleq \begin{bmatrix}
\bar{u}_{1,z} \\
\bar{u}_{2,z} \\
\bar{u}_{3,z}
\end{bmatrix} \\
Q &\triangleq \begin{bmatrix}
u_{1,x} & \bar{u}_{1,x} \\
v_{2,x} & \bar{u}_{2,x} \\
v_{3,x} & \bar{u}_{3,x}
\end{bmatrix} \\
v_{1,y} & \bar{u}_{1,y} \\
v_{2,y} & \bar{u}_{2,y} \\
v_{3,y} & \bar{u}_{3,y}
\end{align*}
$$  \hspace{1cm} (8.16)

where a subscript "k, v" (with k a numerical value and v a character) associated with $u$ and $\bar{u}$ denotes the v component of the vectors $u_k$ and $\bar{u}_k$, respectively. Eq. (8.15) represents three over-determined systems of linear equations that have least squares (LS) solutions

$$
\begin{align*}
a_x &= \left( Q^T Q \right)^{-1} Q^T q_x \\
a_y &= \left( Q^T Q \right)^{-1} Q^T q_y \\
a_z &= \left( Q^T Q \right)^{-1} Q^T q_z. \\
\end{align*}
$$  \hspace{1cm} (8.17)

Eq. (8.17) determines the elements of the two columns $a_1$ and $a_2$. The third column $a_3$ is orthogonal to both $a_1$ and $a_2$, and can hence be determined using the vector cross product as

$$
a_3 = a_1 \times a_2.
$$  \hspace{1cm} (8.18)
8.3.1.2 Solution 2

The following formulation provides an alternative approach to solve the location-orientation problem. From Fig. 8.1, using the cosine rule leads to the following system of equations:

\[ f(r_i, r_j) = 0, i, j \in \{1, ..., 3\}, i \neq j \]  \hspace{1cm} (8.19)

where \( f(r_i, r_j) \) are defined as

\[ f(r_i, r_j) = r_i^2 + r_j^2 - 2r_i r_j \bar{u}_i^T \bar{u}_j - d_{ij}^2 \]  \hspace{1cm} (8.20)

where \( r_k = \| r_k \| \) and \( d_{ij} = \| d_{ij} \| \). Eq. (8.19) represents three elliptic cylinders in the space spanned by the vector \( r \triangleq [r_1 \ r_2 \ r_3]^T \). Each cylinder is aligned along one of the coordinate axes. Again the NR method can be used to solve (8.19) effectively. In fact, the solution of Eq. (8.19) is less cumbersome compared to that of (8.7) in terms of both computational complexity and convergence speed. Better convergence is obtained when the initial guess contains three positive values. Solving Eq. (8.19), the vector \( r \) can be used to determine both the location and the orientation of the receiver. Location can be determined using trilateration [72]. However, with the available information, orientation and then location can be obtained using closed-form formulae.

To determine orientation, three vectors expressed in both the local and global coordinate systems are required as described in Section 8.3.1.1. From Fig. 8.1, using \( r \) obtained from solving Eq. (8.19), we have

\[ \bar{d}_{ij} = r_i \bar{u}_i - r_j \bar{u}_j, \{i, j\} \subset \{1, ..., 3\}, i \neq j. \]  \hspace{1cm} (8.21)

The global counterparts of the vectors in Eq. (8.21) are directly calculated as

\[ d_{ij} = \frac{b_j - b_i}{\| b_j - b_i \|}. \]  \hspace{1cm} (8.22)

Given (8.21) and (8.22) the rotation matrix can be determined in a similar way to (8.14) as

\[ A = \bar{D} \bar{D}^{-1} \]  \hspace{1cm} (8.23)
8.4 SIMULATION RESULTS

where \( D \) and \( \bar{D} \) are \( 3 \times 3 \) matrices defined as

\[
D = \begin{bmatrix} d_{12} & d_{23} & d_{31} \end{bmatrix}^T,
\bar{D} = \begin{bmatrix} \bar{d}_{12} & \bar{d}_{23} & \bar{d}_{31} \end{bmatrix}^T.
\] (8.24)

Alternatively, if \( D \) is singular due to the beacons lying on a plane \( z = \) constant, the LS solution for \( A \) is given by (8.17) and (8.18) with \( Q, q_x, q_y \) and \( q_z \) defined as

\[
q_x = \begin{bmatrix} \bar{d}_{12, x} \\ \bar{d}_{23, x} \\ \bar{d}_{31, x} \end{bmatrix}, \quad q_y = \begin{bmatrix} \bar{d}_{12, y} \\ \bar{d}_{23, y} \\ \bar{d}_{31, y} \end{bmatrix}, \quad q_z = \begin{bmatrix} \bar{d}_{12, z} \\ \bar{d}_{23, z} \\ \bar{d}_{31, z} \end{bmatrix}
\] (8.25)

The receiver’s location, \( \mathbf{m} \), can now be determined using the relationship

\[
\bar{r}_k = r_k \bar{u}_k = A r_k = A (b_k - \mathbf{m}), \quad k = 1, \ldots, 3
\] (8.26)

where \( r_k \) and \( \bar{r}_k \) are the LOS vectors corresponding to beacon \( k \) in the global and local coordinate system, respectively. Eq. (8.26) can be solved for \( \mathbf{m} \) for each value of \( k \) resulting in three estimates for \( \mathbf{m} \). In a noisy scenario, these three estimates will generally constitute a triangle whose centroid can be taken as the final location estimate. In other words, \( \mathbf{m} \) can be estimated from (8.26) using

\[
\mathbf{m} = \frac{1}{3} \sum_{k=1}^{3} b_k - A^T \bar{r}_k
\] (8.27)

where \( A^T \) has replaced \( A^{-1} \) due to the fact that \( A \) is unitary.

8.4 Simulation Results

The two solutions for location-orientation presented in the previous section were simulated in Matlab. In simulation, the beacons were situated on the \( z = 0 \) plane with \( b_1 = [0, 0, 0]^T, b_2 = [5, 0, 0]^T \) and \( b_3 = [0, 5, 0]^T \) (all in meters). The receiver was located on the \( z = -3 \) plane and allowed to move
on the square plane whose edge points are \([0, 0, -3]^T\), \([5, 0, -3]^T\), \([0, 5, -3]^T\) and \([5, 5, -3]^T\). The orientation of the receiver was fixed with the rows of \(A\) set equal to \(\cos[45^\circ, 45^\circ, 0^\circ]^T\), \(\cos[135^\circ, 45^\circ, 0^\circ]^T\) and \(\cos[90^\circ, 90^\circ, 0^\circ]^T\) respectively. With this setting, the true angles (\(\alpha_k\) and \(\beta_k\) in degrees) were contaminated with added error. The error followed a normal distribution with zero mean and unity variance (in squared degrees).

Figs. 8.3 and Fig. 8.4 show location error for solution 1 and solution 2 respectively. Fig. 8.5 and Fig. 8.6 plot orientation error for these solutions, respectively. The resolution of the plots is 0.2 m for both the \(x\) and \(y\) coordinates. Location error was calculated as the distance between the estimated location and the true location, while orientation error, for each axis, was calculated as the angle between the estimated direction for that axis and the true direction. Each figure shows the average of 20 simulation runs. The figures show a general trend of increase in error at the corners with the furthermore corner from the beacons receiving the highest error, which reflects algorithmic non-convergence at that spot. However, the error is acceptable around the center of the test plane. Such a spatial distribution of the error can be attributed to the so-called geometric dilution of precision (GDOP) \([90]\), that is the effect of the beacon arrangement. Apparently, this effect can be reduced by adding a fourth beacon at the fourth corner. It can also be seen that the error in the orientation of the \(z\)-axis is always greater than that associated with the other two axes. It is clear that is due to error accumulation in Eq. (8.18).

From the figures, it can also be seen that the performance of solution 2 is apparently better than that of solution 1 in determining location, and is worse in determining orientation. This is also reflected in Table 8.1, which summarizes the results in terms of the overall mean of location and orientation errors. However, this apparent difference in performance is found to be only due to the contributions from extreme locations. In other words, the two solutions behave differently at their own sets of unfavored locations where they result in large errors (mainly due to algorithmic divergence). However, the performance in locations where the two solutions produce meaningful results is found to be similar. This conclusion will be investigated more thoroughly in Chapter 9 where experimental results will be presented.
Figure 8.3: 3-D location error surface for solution 1.

Figure 8.4: 3-D location error surface for solution 2.
Figure 8.5: Orientation error surface for solution 1: a) x-axis b) y-axis c) z-axis.

Figure 8.6: Orientation error surface for solution 2: a) x-axis b) y-axis c) z-axis.
Table 8.1: Mean 3-D location and orientation errors.

<table>
<thead>
<tr>
<th></th>
<th>solution 1</th>
<th>solution 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location</td>
<td>0.28 m</td>
<td>0.26 m</td>
</tr>
<tr>
<td>x orient.</td>
<td>1.3°</td>
<td>2.6°</td>
</tr>
<tr>
<td>y orient.</td>
<td>1.4°</td>
<td>2.1°</td>
</tr>
<tr>
<td>z orient.</td>
<td>1.7°</td>
<td>2.9°</td>
</tr>
</tbody>
</table>

8.5 Summary

In this chapter, we presented formulations and two solutions for the receiver location-orientation problem using only TDOA/AOA information. It has been demonstrated that, both location and orientation can be determined using as few as three non-collinear beacons and two pairs of orthogonal sensors. The performance of the two proposed approaches was assessed via simulation.
In this chapter, issues related to the implementation of a location and orientation estimation system using the methods proposed in the preceding chapters, are discussed. Descriptions of a system prototype and experiments conducted to test the final system are given, and the results obtained are reported.

9.1 Introduction

In this chapter, an indoor positioning system that utilizes transmission of FHSS signals is described. The system uses only AOA information to determine both the location and orientation of the receiver. This chapter integrates the methods proposed in the previous chapters in order to realize a complete system prototype. The chapter presents a general approach for the design of such a system. The impact of different design parameters is highlighted. A specific design based on the available equipment is described. The design is meant to be the simplest possible in terms of the amount of hardware to be used.

The main functions that the system is required to perform is location and orientation estimation. However, prior to that, the system must acquire the hopping patterns (HPs) of the in-range beacons and synchronizes with them. These acquisition and synchronization aspects of the systems were covered in Chapter 4 together with the relevant experimental tests. In this chapter, it will be assumed that synchronization has been achieved in the manner described in Chapter 4, and the chapter will therefore focuses on the location and orientation estimation part of the system.

This chapter is organized as follows. Section 9.2 details the FHSS design parameters chosen for the final system. The discussion is based on Chapter 3. Sec-
tion 9.3 employs one of the 1-D AOA estimation methods presented in Chapter 7 to estimate the direction vectors corresponding to the beacons. First, the choice of one particular 1-D AOA method is motivated, then a receiver structure of three co-planar sensors is proposed to extend that method to the 2-D AOA case. In section 9.4, the experimental setup of the system is described and results are discussed. Finally, Section 9.5 concludes this chapter by giving a brief summary of the contents of the chapter and highlighting the main points.

9.2 The Proposed FHSS Design

The transducers available for the prototype provide a bandwidth of 15 kHz in the range 35-50 kHz. According to the design criteria proposed in Chapter 3 and Chapter 4, the usable bandwidth is divided into two subbands—one for preamble transmission and one for payload. The payload band is divided into 14 separate channels each of 750 Hz. The channels start from channel 1 at a frequency of 36 kHz and step by 750 Hz until channel 14 at 45.75 kHz. The remaining frequency band up to 50 kHz is left for preamble transmission in the manner described in chapter 4. The hop duration is chosen to be 3.06 ms. Note that the chosen number of payload hops (14) coincides with one of the optimal choices, as in Chapter 4. However, it is suboptimal in terms of preamble processing requirement. The choice here is dictated by bandwidth limitation and the desired clearance distance. After each transmission cycle (3.06 ms/hop × 14 hops ≈ 43 ms), the transmitter has to stay silent for a period of $RT - 43$ ms before starting the next cycle, where $RT$ is the room reverberation time. Note that the silence period includes the preamble part of the transmission, which is not of interest herein.

The 750 Hz channel width allows for a hopping rate of 750 hop/s when no modulation is applied. That is equivalent to a 1.33 ms hop duration. However, to improve the orthogonality of signals in the reverberant indoor environment, and also to allow for a sufficient guard against Doppler shifts, the hop duration is increased to 3.06 ms reducing the hopping rate to approximately 327 hops/s. The corresponding clearance distance is approximately 0.5 m. This design allows for Doppler shifts up to 423 Hz to be detected, which is approximately equivalent to movements at speeds of 4.0 m/s down to 3.2 m/s for the payload frequency range. Although the experiments discussed in this chapter did not include movement, it was found that having such a guard is particularly useful.
in the reverberant environment. This design is subject to auto-contamination (see Chapter 3) when the transmitter(s) and/or the receiver is located within 0.5 m from a reflector. It was found empirically that decreasing the hop duration further is more harmful than useful, at least in the test space of consideration.

9.3 AOA Estimation

A number of AOA estimation methods were discussed in Chapter 7. The discussion was confined to the 1-D case. For receiver's location and orientation estimation, 2-D AOA information is required in order to obtain the direction vectors required for location and orientation estimation. The most straightforward way to obtain 2-D AOA information is to utilize a pair of orthogonal sensors, as presented in Chapter 8. In this case, the estimates need to be unambiguous. Therefore, depending on which method is used, the configuration may need to be extended so that disambiguation can be performed. Disambiguation of the estimates from each of the orthogonal sensor pairs can be carried out independently, and the 2-D AOA process amounts to two independent 1-D AOA estimation including disambiguation.

The 1-D AOA estimation methods discussed in Chapter 7 are based on two different disambiguation approaches, namely, spatial-diversity-based and frequency-diversity-based disambiguation. The chapter conclusion was that the grid-search methods are the most robust and deliver good estimation performance in the real environment. For the final system implementation, the frequency-diversity-based grid-search method is favoured over the spatial-diversity-based method. The decision is based on the following two points:

- Despite the slightly improved performance, the spatial-diversity-based method has pitfalls in some regions of space, namely, around the boresight area.

- The frequency-diversity-based method does not require extra sensors, which results in a simpler receiver structure and also lower computational cost.

Due to these points, the grid-search method using the frequency-diversity-based (FDB-GS) approach was adopted as the AOA estimation method (the 1-D case) for the final system.
9.3.1 Direction Vectors Estimation

To obtain 2-D AOA information, at least three non-collinear sensors are required. Direction vectors carry the same information as 2-D AOA information. To obtain the direction vector corresponding to each source, the final system utilizes the three-sensor configuration depicted in Fig. 9.1. The process works as follows. The AOA estimates pertaining to source \( k \), \( \hat{\theta}_{12,k} \), \( \hat{\theta}_{23,k} \) and \( \hat{\theta}_{31,k} \) are estimated from the three pairs of sensors using the FDB-GS method. From these angles we obtain a vector, 
\[
v_k \triangleq \begin{bmatrix} \cos(\hat{\theta}_{12,k}) \\ \cos(\hat{\theta}_{23,k}) \\ \cos(\hat{\theta}_{31,k}) \end{bmatrix}^T.
\]

The direction vector, \( \bar{u}_k \), corresponding to source \( k \) is calculated from the three angles as follows. First, a local coordinate system is assumed. The x-axis and y-axis of the local coordinate system assumed herein are shown in Fig. 9.1, where the centroid of the triangle is adopted as the origin of the local coordinate system. Based on this coordinate system, the direction vectors, \( \bar{d}_{12} \), \( \bar{d}_{23} \) and \( \bar{d}_{32} \), of the three triangle sides are calculated. Note that each of these three direction vectors has zero z-component since the sensors are assumed to be located in the \( \bar{z} = 0 \) plane. The direction vector, \( \bar{u}_k \), makes the angles given by the elements of the vector \( v_k \) with each of \( \bar{d}_{12} \), \( \bar{d}_{23} \) and \( \bar{d}_{32} \), respectively. Applying the dot product rule, results in the following system of equations

\[
Q_k q_k = v_k \tag{9.1}
\]

where

\[
Q_k \triangleq \begin{bmatrix} \bar{d}_{12,x} & \bar{d}_{12,y} \\ \bar{d}_{23,x} & \bar{d}_{23,y} \\ \bar{d}_{31,x} & \bar{d}_{31,y} \end{bmatrix} \quad \text{and} \quad q_k \triangleq \begin{bmatrix} \cos(\alpha_k) \\ \cos(\beta_k) \end{bmatrix}.
\]

Eq. (9.1) has the least squares solution

\[
q_k = \left( Q_k^T Q_k \right)^{-1} Q_k^T v_k \tag{9.2}
\]

whereby \( \cos(\alpha_k) \) and \( \cos(\beta_k) \) are obtained. The third component of \( \bar{u}_k \), \( \cos(\gamma_k) \), is calculated in the manner described in chapter 8. This way, the direction vectors corresponding to the beacons in range that are required for location and orientation estimation, are obtained. Next, the two solutions discussed in Chapter 8 are applied to obtain the receiver’s location and orientation, with the location of the receiver defined as the 3-D coordinates of the local origin, and the orientation of the receiver defined as the local-to-global coordinate rotation matrix (see Chapter 8).
9.3 AOA ESTIMATION

Figure 9.1: The proposed sensor configuration for 2-D AOA estimation.

Figure 9.2: Experimental setup: the grid indicates the locations of the receiver.
9.4 Experiments

9.4.1 System Components

A general layout of the system used for testing the proposed location-orientation estimation system is shown in Fig. 9.3. The system consists of the following units:

- A DSP board.
- Transmitters.
- Transmitter interfacing circuits.
- Sensors.
- Sensor interfacing circuits.
- A PC.

The DSP board used was the SMT361A from Sundance [124]. The Board contained an analog-to-digital converter (ADC), a digital-to-analog converter (DAC), a field-programmable gate array (FPGA) and a digital signal processor (DSP).

Each transmitter was connected to a separate interfacing circuit to form a separate transmitter unit. A total of three transmitter units were used in the tests. The transmitters used were the E-152/40 wideband ultrasonic transducers from Massa [125]. This type of transducer has a bandwidth of 15 kHz (the
9.4 EXPERIMENTS

band 35–50 kHz) and a total beam angle of 75°. The main function of the sensor interfacing circuits was to amplify the sensor output voltages from a few millivolts to a peak-to-peak voltage of up to 4 volts, which was the maximum level of the ADC in use.

All the sensors and their interfacing circuits were contained in a single receiver unit. The receiver contained three sensors configured in a triangle as shown in Fig. 9.1. The triangle was equilateral with a side of 21 mm. The sensors were the SPM0204UD5 from Knowles [126]. These have a frequency range of 10–65 kHz and are omnidirectional. The interfacing circuits on the transmitter side were used to boost the output of the DAC from a maximum of 5 volts (peak-to-peak) to about 20 volts (peak-to-peak). An additional function of the interfacing circuits on both the transmitter and receiver sides was to filter out any signals outside the frequency range of interest. Analog bandpass filters with cut-off frequencies of 20 and 60 kHz were used for this purpose. This ensured that all noise in the audible range was be removed. The interfacing circuits were designed, implemented and tested before integration with the rest of the system.

The transmitted signals were stored in files on the PC. Diamond’s 3L IDE was used as the interface between the DSP board and the user/PC. The environment was used to create a software that read the signals from the PC and sent them to the DAC, where they were used to drive the ultrasonic transducers in the transmitter units with the appropriate voltage levels. At the same time, the ADC recorded the signals received by the sensors after the were amplified and filtered. The recorded signals were passed to the to the DSP, which was programmed to store them in the PC to be processed off-line. Matlab software was used to carry out the processing task.

9.4.2 Experimental Setup

The (three) sensors contained on the receiver were arranged in a triangle, as depicted in Fig. 9.1. The triangle was equilateral with a side of 21 mm.

The experimental setup was mainly dictated by the type of ultrasonic transducers used. It was found that despite their good properties in terms of bandwidth and beamwidth, the Massa E152/40 were very limited in range. For an application that uses the entire 15 kHz bandwidth, as was the case for this work, problems occur for ranges greater than 2 m (at least for the space where the tests were conducted). This limited the size of the space where signals
could be received with good quality. To reduce this effect, the beacons were situated at rather a low height.

The test space is depicted in Fig. 9.2. A global coordinate system was assumed to center at one of the corners of the test space as shown in Fig. 9.2. The locations of the beacons were \( \mathbf{b}_1 = [13.0, 154.0, 193]^T \), \( \mathbf{b}_2 = [35.0, 5.0, 193]^T \) and \( \mathbf{b}_3 = [290.0, 7.0, 193]^T \), with all values measured in centimeters. The equal z-coordinates of the beacons indicate that the beacons were co-planar. In all cases, the locations (of the beacons or the receiver) were measured to within 0.5 cm of error for each coordinate. The reverberation time of the test space at the frequency band of interest was measured to be approximately \( RT = 90 \text{ ms} \). Therefore, beacons had to go silent for approximately 47 ms in every 90 ms.

The choice was to have the whole period attached to the end of each hopping cycle (rather than intermittent and interleaved small silence periods). This choice was suitable for the static (no movement) experimental setup.

During the tests, the location of the receiver was varied over a grid of 25 locations. The grid step was 30 cm. The locations are marked in Fig. 9.2. The receiver’s location and orientation were calculated from the received signals and the results were compared to the ground truth. The process was repeated ten times at each location to make sure that the results were consistent. The final results presented herein—the root mean squared errors (RMSEs)—were calculated from the set of ten measurements at each location. Each measurement was calculated from a complete hopping cycle, that is 14 hops for each beacon. Note that the beacons were not synchronized. Therefore, the AOA measurements corresponding to the three beacons were not always concurrent; most of the time there were a slight timing differences. This means that location and orientation were calculated only every \( RT = 90 \text{ ms} \), which corresponds to an update rate of approximately 11 Hz.

To facilitate comparison, the z-coordinate of the receiver was maintained fixed at 81 cm and the receiver (local) coordinate axes were maintained parallel to the corresponding global coordinate axes, resulting in the rotation matrix

\[
\mathbf{A} = \begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{bmatrix}.
\]  

(9.3)

The following notes on the experimental setup are worth mentioning before discussing results:
• The transmitters were located such that signal autocontamination (see Chapter 3) was possible over the entire test space.

• All walls close to the transmitters were non-rigid (however consisted of small metallic parts), except for transmitter 1, which was neighbored by a rigid wall.

• The top row and the bottom row in Fig. 9.2 are, respectively, 30 and 35 cm, from the neighboring walls.

• The test space also contained other reflecting objects, e.g., chairs, computers, and other equipment scattered on the desks and other locations. Almost all the locations at the edges of the test space (about 16 locations) were neighbored by reflectors at distances less than the clearance distance.

9.4.3 Results

Figs. 9.4–9.6 plot the raw AOA estimation performance (This is for the angles observed by the three pairs of sensors). The left columns show the RMSE for each of the 25 locations. The right columns show samples of the cumulative distribution of the RMSE over the 25 locations. The figures clearly visualize the occurrence of outliers in certain locations, more specifically at the edges of the test space. The outliers are a manifestation of the failure of the AOA estimation process. This phenomenon can be attributed to two types of system impairment:

• Reverberation: These locations are closer to the walls where the effect of reverberation is more pronounced.

• Attenuation: In most of the cases, large errors occur at locations that are far from the location of the source. Measurements show that the SNR at these locations degrades to as low as -3 dB for some frequencies in some cases.

It can be said that the outer locations of the test space are, in most cases, inadequate for good AOA estimation. On the other hand, for the inner locations, the results show, in all the cases, no occurrence of outliers.

The RMSE distribution graphs implicitly indicate the accuracy achievable in central locations. For example, it can be seen that an RMSE of less than 1° is
achievable in some of these locations, e.g., the angle $\theta_{12,1}$ has been measured with such an accuracy in more than 50% of the cases. For the same angle, 80% of the locations had an RMSE of less than $5^\circ$. This indicates that about 20% of the locations are giving unreliable results. A conclusion that can be derived from the distribution graphs is that, where the location is not severely affected by the effects of reverberation and/or low signal strength, the error in AOA estimation is bounded by $5^\circ$ or less.

Figs. 9.7–9.9 show the error in direction vector estimation. The figures show the individual errors associated with the three angles comprising each vector ($\alpha$, $\beta$ and $\gamma$) and the total deviation of each vector from the ground truth. The figures show an increase in the locations with outlying results compared to Figs. 9.4–9.6. However, the accuracy in the central locations remains around the same level; improving in some cases and declining in other cases. This depends on the way the error accumulates in the final estimates. The number of locations with outlying measurements increases due to the influence of the outlying measurements from the raw angle estimates. Simply, incorporating three estimates to calculate the vector components increases the chance of obtaining outliers. Roughly speaking, the total number of locations with outlying estimates is at least equivalent to the number of locations where outlying measurements occur for any of the raw AOA estimates. The same phenomenon is also true for the angles $\gamma_k, k = 1, 2, 3$; they may have better accuracy than the corresponding $\alpha_k$ and $\beta_k$ that they are calculated from. In conclusion with respect to Figs. 9.7–9.9, the accumulation of error is graceful only for some of the locations, and the corresponding direction vector components estimation accuracy is comparable to that of the raw AOAs. On the other hand, the error in estimating the whole vector (consolidating the three components as one entity) shows the effect of accumulation of the error from the three basic components. This accumulation is important since it affects the final location and orientation accuracy. Note that this accumulation is inevitable regardless of the raw AOA error model, simply because the errors pertaining to the three components are naturally cumulative and cannot cancel or reduce.

In Fig. 9.10, the errors associated with the inter-vector angles are depicted. The errors are generally worse than the direction vector errors depicted in Figs. 9.7–9.9 and the effect of outliers is more pronounced. However, for some locations an improved accuracy is obtained. Again, this can be explained as a graceful accumulation of error at these particular locations.
Figs. 9.11–9.14 present the results pertaining to location and orientation estimation using the two solutions discussed in Chapter 8. From the error distribution plots, a conclusion can be reached that the two solutions offer the same performance for both location and orientation determination. However, contemplating the RMSE plots for location estimation, it can be seen that the occurrence of outliers is more notable for solution 1 than for solution 2. This effect is not reflected in the error distribution graphs since these are limited to errors that are less than 100 cm. Since an accuracy worse than that limit is useless, the fact that solution 2 produces outliers with smaller error values than solution 1 does not favour the former on the latter from a practical perspective. For orientation, the role is exchanged and solution 1 offers slightly better performance in terms of outlier effect. Both location and orientation results confirm that the proposed methods are capable of delivering reasonable location accuracy, when the receiver is sufficiently far from reflecting objects and receives adequate signal power. The accuracy is in the range of 30 cm for location estimation. The accuracy for orientation estimation varies for each of the three axes of the device. The best is for the x-axis which is bounded by 5° for the region of interest and may improve to 3° for some locations. For the y-axis the accuracy bound is 10°, and for the z-axis the accuracy lies somewhere between 10° and 15°.

The variation in orientation performance for each axis is related to the sensor (array) configuration. The sensors are planar and situated on the x-y plane of the local coordinate system of the receiver. All information pertaining to the z-dimension is obtained indirectly by transformation of the information obtained from that plane. Therefore, it is expected that the final error associated with the z-axis orientation be the least accurate among the three axes. Considering the other two axes, the configuration is not symmetric; for example, the x-axis is parallel to the baseline connecting sensor 2 and sensor 3 (and so the angles $\theta_{23,k}$ will coincide with the angles $\alpha_k$), while the y-axis is not parallel to any sensor baseline. This means that the transformation from raw AOA estimates to direction vector estimates takes place in different ways that affect the orientation estimates for each axis in a different way. In this regard, a more symmetric configuration, e.g., a circular array, can be more helpful. However, the goal was to deploy the simplest possible system—a circular array requires more sensors than the triangular configuration used herein.

The convergence of the iterative part of each of the two solutions is analyzed in Fig. 9.15. In the implementation of these iterative algorithms, the algorithms
were set to iterate for a maximum of 20 iterations. Therefore, the cases where an algorithm required 20 iterations are suspected to coincide with algorithmic divergence. Fig. 9.15 indicates that divergence takes place in very few cases. These cases coincide with at least one outlying AOA measurement.

Figure 9.4: AOA estimation performance: source 1.
Figure 9.5: AOA estimation performance: source 2.
Figure 9.6: AOA estimation performance: source 3.
Figure 9.7: Direction vector estimation performance: source 1.
Figure 9.8: Direction vector estimation performance: source 2.
Figure 9.9: Direction vector estimation performance: source 3.
Figure 9.10: Inter-vector angle estimation performance.
Figure 9.11: Location estimation performance: solution 1.
Figure 9.12: Location estimation performance: solution 2.
Figure 9.13: Orientation estimation performance: solution 1.
Figure 9.14: Orientation estimation performance: solution 2.
9.5 Summary

This chapter described an AOA-based ultrasonic 3-D location and orientation estimation system. The system used FHSS as the transmission scheme. Three co-planar sensors were used to estimate the AOA of the signals received from three co-planar beacons. The sensors were widely-spaced and AOA were estimated based on a phase-difference estimation and disambiguation method. The whole process of AOA estimation was carried out in the frequency domain where time-delays were estimated from the phase of the non-averaged cross-power spectra of the received signals. Special attention was paid to signal design to overcome indoor signal impairments, in particular, reverberation.

The results presented in this chapter serves as a proof-of-concept. Due to some hardware limitations (ultrasonic transducer range and bandwidth), the system performance at some locations was not satisfactory. However, the results demonstrate that the system can offer reasonable performance in the scenarios for which the system has been designed.

Despite the good AOA estimation accuracy in many cases, the final location and orientation results were always less accurate. The proposed location and
orientation algorithms showed some error accumulation effects. The algorithms consisted of a number of stages that propagated error to one another, sometimes resulting in relatively large location and orientation errors. Recommendations for improvement of the system are found in Chapter 10.
This thesis described an ultrasonic 3-D location-orientation approach based on AOA estimation. The approach is suitable for privacy-aware indoor location systems. The major advantage of the proposed approach is that it does not require RF (or any other type of) synchronization between system units. Instead, the receiver device determines its own position based on the received ultrasonic signals. This results in reduced system complexity, reduced system cost and easier deployment. The results obtained from the prototype, as presented in this thesis, serve as a proof-of-concept for the proposed approach (and its underlying methods). Yet further work is required to enhance performance and improve robustness towards a more practical system. The results show that the current approach is capable of providing an accuracy of about 30 cm for location, 5° for the x-axis orientation, 10° for the y-axis orientation and 15° for the z-axis orientation, at an 11 Hz update rate.

The proposed approach is based on solving a number of underlying problems. First of all, the signal design that best suits the operational environment and is realizable using the available resources, was identified. The choice of FHSS was motivated by previously published work. The signaling scheme proved suitable for the proposed approach. The sparsity and orthogonality of FHSS signals were exploited for multi-source AOA estimation.

A FHSS acquisition and synchronization approach based on DBPSK transmission was devised. A communication system for transmitting the IDs of the beacons was designed and tested experimentally. The whole system was implemented digitally. The approach was found to be promising and can be used for ultrasonic data communications.

The thesis dealt with the phase-difference ambiguity problem associated with AOA estimation. The problem was addressed and mathematically well-grounded solutions were provided. The solutions were evaluated both ana-
lytically and in simulation. The solutions allowed for a number of AOA estimation approaches. These approaches were evaluated and compared using a real experimental setup. Alas, in a difficult indoor environment, very few approaches gave satisfactory performance. Two algorithms exhibited good performance, one of which was adopted for the final system. The final system setup employed a coplanar array of three sensors arranged in an equilateral triangle.

The thesis also proposed solutions to the AOA based receiver 3-D location-orientation estimation problem. The solutions were iterative. The solutions were first tested in simulation. No convergence problems were detected. These solutions together with the selected AOA approach formed the final system that was tested through a prototype. The prototype was implemented using off-the-shelf components.

System performance was affected by two factors. The first factor was the signal design, which unavoidably, made accurate AOA estimation prohibitive in some regions of the space covered by the beacons, namely, close to reflectors. The second factor is related to the signal strength of the ultrasonic transducers in use. Due to the short range of the transducers used in the tests, the SNR was too low at extreme locations. These two factors have seriously affected the final location and orientation results. Both problems could be alleviated through the use of more powerful ultrasonic transducers.

Ultrasonic transducers with larger bandwidth could be helpful in increasing the total number of hops, and at the same time reducing the hop duration. This could improve performance significantly. Increasing the bandwidth also helps improve signal orthogonality in different ways. For example, the guard frequency can be increased, and hopping patterns with better properties can be designed. Also, transducers with greater range are desirable. The two requirements—greater bandwidth and greater range—together may not be satisfied by the off-the-shelf components. Therefore, one needs to either consider the fabrication of new hardware that meets the requirements, or contemplate more efficient signal design approaches using the available components.

For more accurate and robust AOA estimation, one may consider using the same methods with a larger number of sensors. The results presented in this thesis were obtained by using the minimum number of sensors (three sensors). Increasing the number of sensors gives one the liberty of arranging the sensors in different ways. In particular, a circular array can be more attractive to overcome the apparent asymmetry in the orientation estimation results for
the different axes. Since we are dealing with small devices, the array should not be made too large.

For improving location and orientation results, increasing the number of beacons might also be helpful. Practically, this is limited by the bandwidth available for FHSS transmission. However, a robust signal design may help circumvent this difficulty. This, again, stresses the importance of signal design.

In addition to the above ways of improving location and orientation performance, the location and orientation algorithms themselves can be questioned. The algorithms are computationally efficient but they exhibit an increased amount of error, even when the raw AOA error is not too large. The effect of error propagation in different stages seems significant. In this regard, more compact solutions that do not consist of multiple stages should be sought. These solutions may have different sensor and/or beacon configuration requirements.

Performance improvement could also be introduced by using a device such as an inclinometer for determining the receiver orientation with respect to gravity. These devices give very good accuracy that location error (from the proposed approach) can be significantly reduced. The introduction of such a device would maintain almost all the advantages of the proposed approach and improve performance, albeit for an increased receiver device cost.

Finally, it should be mentioned that although the approach and the underlying methods discussed in this thesis were originally designed for ultrasonic technology, they are also suitable for any single-medium technology (i.e., a technology that uses one type of signal), for example, RF.
A.1 Derivation of the Probabilities $P_{s,m}, P_{k_1}$ and $P_{k_2}$

First, let us start with the probability $P(|a + \epsilon| < |a_i + \epsilon|)$, where $\epsilon$ is a random variable with Gaussian distribution and CDF $\Phi_\epsilon(x)$, and $a$ and $a_i$ are real constants, $a \neq a_i$. This probability can be stated as

$$P(|a + \epsilon| < |a_i + \epsilon|) = P(\epsilon > \Omega_i), \text{ for } a < a_i$$

$$= P(\epsilon < \Omega_i), \text{ for } a > a_i$$

(A.1)

where $\Omega_i \equiv -(a + a_i)/2$.

Now, consider the probability $P = P(|a + \epsilon| < |a_1 + \epsilon| \& |a + \epsilon| < |a_2 + \epsilon|)$, $a \neq a_1 \neq a_2$. Based on A.1, four different cases can be recognized as follows:

case 1: $a < a_1$ and $a < a_2$

$$P = P(\epsilon > \Omega_1 \ & \ \epsilon > \Omega_2)$$

$$= P[\epsilon > \max(\Omega_1, \Omega_2)]$$

$$= 1 - \Phi_\epsilon[\max(\Omega_1, \Omega_2)]$$

(A.2)

case 2: $a > a_1$ and $a > a_2$

$$P = P(\epsilon < \Omega_1 \ & \ \epsilon < \Omega_2)$$

$$= P[\epsilon < \min(\Omega_1, \Omega_2)]$$

$$= 1 - \Phi_\epsilon[-\min(\Omega_1, \Omega_2)]$$

(A.3)
case 3: $a < a_1$ and $a > a_2$

$$P = P(\varepsilon > \Omega_1 \& \varepsilon < \Omega_2)$$
$$= P(\varepsilon < \Omega_2) - P(\varepsilon \leq \Omega_1)$$
$$= 1 - \Phi_\varepsilon(\Omega_1) - \Phi_\varepsilon(-\Omega_2), \text{ for } \Omega_1 < \Omega_2$$
$$= 0, \text{ for } \Omega_1 > \Omega_2 \quad \text{(A.4)}$$

case 4: $a > a_1$ and $a < a_2$

$$P = P(\varepsilon < \Omega_1 \& \varepsilon > \Omega_2)$$
$$= P(\varepsilon < \Omega_1) - P(\varepsilon \leq \Omega_2)$$
$$= 1 - \Phi_\varepsilon(-\Omega_1) - \Phi_\varepsilon(\Omega_2), \text{ for } \Omega_1 > \Omega_2$$
$$= 0, \text{ for } \Omega_1 < \Omega_2 \quad \text{(A.5)}$$

Hence, for $P = P_{s,m}$ given by (5.41), the above procedure can be used by setting $a = \tau_{12}(k_o) + \gamma_{k_o}, a_1 = \tau_{12}(k_1) + \gamma_{k_1}$ and $a_2 = \tau_{12}(k_2) + \gamma_{k_2}$. To evaluate $P_{k_1}$, set $a = \tau_{12}(k_1) + \gamma_{k_1}, a_1 = \tau_{12}(k_o) + \gamma_{k_o}$ and $a_2 = \tau_{12}(k_2) + \gamma_{k_2}$; and to evaluate $P_{k_2}$, set $a = \tau_{12}(k_2) + \gamma_{k_2}, a_1 = \tau_{12}(k_1) + \gamma_{k_1}$ and $a_2 = \tau_{12}(k_o) + \gamma_{k_o}$.

### A.2 Derivation of $\Gamma_f$

Similar to Eqs. (5.45) and (5.46), the contribution of the failure cases to the overall MSE of the whole disambiguation method is the resultant of three different cases. The general expression for $\Gamma_f$ takes the form

$$\Gamma_f = P\{\Psi[k_{12} + f(\varepsilon - \varepsilon_{12}) + f\gamma_{k_o}] \neq k_{12}\} P_{s,m} \Gamma_{f,m}$$
$$+ P\{\Psi[k_{12} + f(\varepsilon - \varepsilon_{12}) + f\gamma_{k_1} + q_{12}\mu_{\rho} ] \neq k_{12}\} P_{k_1} \Gamma_{f,1}$$
$$+ P\{\Psi[k_{12} + f(\varepsilon - \varepsilon_{12}) + f\gamma_{k_2} + q_{22}\mu_{\rho} ] \neq k_{12}\} P_{k_2} \Gamma_{f,2} \quad \text{(A.6)}$$

where $\Gamma_{f,m}$, $\Gamma_{f,1}$ and $\Gamma_{f,2}$ represent the MSEs associated with the three cases of failure that correspond to the probabilities $P_{s,m}$, $P_{k_1}$ and $P_{k_2}$, respectively. By exploiting Eq. (5.48), Eq. (A.6) can be written as
A.2 DERIVATION OF $\Gamma_f$

\[
\Gamma_f = \left\{ 1 - \left[ \Phi_{\epsilon - \epsilon_{12}} \left( \frac{0.5}{f} + \nu_{k_0} \right) - \Phi_{\epsilon - \epsilon_{12}} \left( -\frac{0.5}{f} + \nu_{k_0} \right) \right] \right\} P_{s,m} \Gamma_{f,m} \\
+ \left\{ 1 - \left[ \Phi_{\epsilon - \epsilon_{12}} \left( 0.5 + \frac{q_1 \mu_p}{f} + \nu_{k_1} \right) - \Phi_{\epsilon - \epsilon_{12}} \left( -0.5 + \frac{q_1 \mu_p}{f} + \nu_{k_1} \right) \right] \right\} P_{k_1} \Gamma_{f,1} \\
+ \left\{ 1 - \left[ \Phi_{\epsilon - \epsilon_{12}} \left( 0.5 + \frac{q_2 \mu_p}{f} + \nu_{k_2} \right) - \Phi_{\epsilon - \epsilon_{12}} \left( -0.5 + \frac{q_2 \mu_p}{f} + \nu_{k_2} \right) \right] \right\} P_{k_2} \Gamma_{f,2}. \tag{A.7}
\]

To obtain the expressions for $\Gamma_{f,m}$, $\Gamma_{f,1}$ and $\Gamma_{f,2}$, the corresponding errors need to be identified. These errors are given by $\hat{k}_{12} - k_{12}$, where $\hat{k}_{12}$ is the output of the RO step. In the case of failure, the difference $\hat{k}_{12} - k_{12}$ has a minimum absolute value of unity. In the general case, the range of values taken by $\hat{k}_{12} - k_{12}$ is determined by the probability distribution of the random variable $\epsilon - \epsilon_{12}$, which suggests that $\hat{k}_{12} - k_{12}$ can be any value in $\mathbb{Z}$ excluding zero. Therefore, each of $\Gamma_{f,m}$, $\Gamma_{f,1}$ and $\Gamma_{f,2}$ can be expressed as an infinite sum of the products of the MSEs corresponding to infinite integer values for $\hat{k}_{12} - k_{12}$, and the probability of obtaining each integer. However, to simplify the analysis, we will ignore all the cases where the rounding function $\Psi(.)$ produces an integer output that has an absolute difference from the noise-free output that is greater than unity. In other words, it will be assumed that the effect of the noise term $f(\epsilon - \epsilon_{12})$ in the erroneous cases, will be confined to changing the noise-free round-off result by plus or minus one. For practical values of $\sigma_p^2$, this assumption was found to be reasonably accurate. Based on this assumption, we can write

\[
\Gamma_{f,m} = \sigma_p^2 + \frac{1}{f^2}. \tag{A.8}
\]

To obtain $\Gamma_{f,i}$, $i = 1, 2$; the corresponding outputs of the RO stage will be writ-
ten as

\[ k_{12,i} = \Psi [k_{12} + f(\epsilon - \epsilon_{12}) + f v_{ki} + q_i \mu_p] \]
\[ = k_{12} + \Psi (f v_{ki} + q_i \mu_p) \]
\[ + \Psi \{ [f v_{ki} + q_i \mu_p - \Psi (f v_{ki} + q_i \mu_p)] + f (\epsilon - \epsilon_{12}) \} \]

(A.9)

where \(-0.5 \leq [f v_{ki} + q_i \mu_p - \Psi (f v_{ki} + q_i \mu_p)] < 0.5\).

Now define the probabilities

\[
P_{-1,i} \approx P \left\{ \Psi \left\{ [f v_{ki} + q_i \mu_p - \Psi (f v_{ki} + q_i \mu_p)] + f (\epsilon - \epsilon_{12}) \right\} = -1 \right\}
\]
\[ = P \left\{ [f v_{ki} + q_i \mu_p - \Psi (f v_{ki} + q_i \mu_p)] + f (\epsilon - \epsilon_{12}) \leq -0.5 \right\}
\]
\[ = \Phi_{\epsilon - \epsilon_{12}} \left\{ -0.5 - f v_{ki} - q_i \mu_p + \Psi (f v_{ki} + q_i \mu_p) \right\} \]

\[
P_{1,i} \approx P \left\{ \Psi \left\{ [f v_{ki} + q_i \mu_p - \Psi (f v_{ki} + q_i \mu_p)] + f (\epsilon - \epsilon_{12}) \right\} = 1 \right\}
\]
\[ = P \left\{ [f v_{ki} + q_i \mu_p - \Psi (f v_{ki} + q_i \mu_p)] + f (\epsilon - \epsilon_{12}) \geq 0.5 \right\}
\]
\[ = \Phi_{\epsilon - \epsilon_{12}} \left\{ -0.5 + f v_{ki} + q_i \mu_p - \Psi (f v_{ki} + q_i \mu_p) \right\} \]

\[
P_{0,i} \approx 1 - P_{-1,i} - P_{1,i}. \tag{A.10}
\]

From the above assumption on the rounding, the contribution of the rounding to the overall MSE can be approximated as

\[
\Gamma_{r,i} \approx \left\{ P_{-1,i} \left[ \Psi (f v_{ki} + q_i \mu_p) - 1 \right]^2 + P_{0,i} \left[ \Psi (f v_{ki} + q_i \mu_p) \right]^2 \right\} / f^2
\]
\[ + P_{1,i} \left[ \Psi (f v_{ki} + q_i \mu_p) + 1 \right]^2 \] / f^2

(A.11)

and \(\Gamma_{f,i}\) can be expressed in the form

\[
\Gamma_{f,i} \approx \sigma_p^2 + \Gamma_{r,i}. \tag{A.12}
\]

Finally, by substituting for \(\Gamma_{f,m}, \Gamma_{f,1}\) and \(\Gamma_{f,2}\) in Eq. (A.7), the final expression for \(\Gamma_f\) is obtained.
A.3 Derivation of the variance $\sigma_p^2$ as a Function of the SNR

From the definition of $|\gamma(\omega)|^2$ in (5.51)

$$
\frac{1 - |\gamma(\omega)|^2}{|\gamma(\omega)|^2} = \frac{1 - \frac{G^2_\omega(\omega)}{[G_\omega(\omega) + G_{nn}(\omega)]^2}}{\frac{G^2_\omega(\omega)}{[G_\omega(\omega) + G_{nn}(\omega)]^2}} = \frac{[G_{ss}(\omega) + G_{nn}(\omega)]^2 - G^2_{ss}(\omega)}{G^2_{ss}(\omega)} = \left[ 1 + \frac{G_{nn}(\omega)}{G_{ss}(\omega)} \right]^2 - 1. \quad (A.13)
$$

However, for white noise, the total power ($P_n = \sum_{<\omega>} G_{nn}(\omega)$), where L is the CPS length, is equally divided among the frequency bins. Noise power at any frequency, assuming a single-sided spectral case, is usually given by $\sigma^2 = P_s/L$. Therefore, the ratio $G_{nn}(\omega)/G_{ss}(\omega)$ can be expressed as

$$
\frac{G_{nn}(\omega)}{G_{ss}(\omega)} = \frac{P_n}{\alpha(\omega)P_s} = \frac{P_n}{\alpha(\omega)L_p} = \frac{1}{\alpha(\omega)L\Lambda} \quad (A.14)
$$

where $P_s = \sum_{<\omega>} G_{ss}(\omega)$ is the total signal power; $\alpha(\omega) \triangleq G_{ss}(\omega)/P_s$; and $\Lambda$ is the linear SNR. By inserting (A.14) into (A.13), and inserting the latter back into (5.50) and manipulating, (5.52) is obtained.
B.1 Derivation of the Probabilities $P_{s,m}$, $P_{k_1}$ and $P_{k_2}$

First, let us start with the probability $P(|a + \epsilon| < |a_i + \epsilon|)$, where $\epsilon$ is a random variable with Gaussian distribution and CDF $\Phi_{\epsilon}(x)$; and $a$ and $a_i$ are real constants, $a \neq a_i$. This probability can be stated as

\[
P(|a + \epsilon| < |a_i + \epsilon|) = P(\epsilon > \Omega_i), \text{ for } a < a_i
\]

\[
= P(\epsilon < \Omega_i), \text{ for } a > a_i
\]

(B.1)

where $\Omega_i \triangleq -(a + a_i)/2$.

Now, consider the probability $P = P(|a + \epsilon| < |a_1 + \epsilon| \& |a + \epsilon| < |a_2 + \epsilon|)$, $a \neq a_1 \neq a_2$. Based on B.1, four different cases can be recognized as follows:

case 1: $a < a_1$ and $a < a_2$

\[
P = P(\epsilon > \Omega_1 \& \epsilon > \Omega_2)
\]

\[
= P[\epsilon > \max(\Omega_1, \Omega_2)]
\]

\[
= 1 - \Phi_{\epsilon}[\max(\Omega_1, \Omega_2)]
\]

(B.2)

case 2: $a > a_1$ and $a > a_2$

\[
P = P(\epsilon < \Omega_1 \& \epsilon < \Omega_2)
\]

\[
= P[\epsilon < \min(\Omega_1, \Omega_2)]
\]

\[
= 1 - \Phi_{\epsilon}[-\min(\Omega_1, \Omega_2)]
\]

(B.3)
case 3: \(a < a_1\) and \(a > a_2\)

\[
P = P(\epsilon > \Omega_1 \quad \& \quad \epsilon < \Omega_2) = P(\epsilon < \Omega_2) - P(\epsilon \leq \Omega_1) = 1 - \Phi_\epsilon(\Omega_1) - \Phi_\epsilon(-\Omega_2), \quad \text{for} \quad \Omega_1 < \Omega_2
\]

\[
= 0, \quad \text{for} \quad \Omega_1 > \Omega_2 \quad \text{(B.4)}
\]

case 4: \(a > a_1\) and \(a < a_2\)

\[
P = P(\epsilon < \Omega_1 \quad \& \quad \epsilon > \Omega_2) = P(\epsilon < \Omega_1) - P(\epsilon \leq \Omega_2) = 1 - \Phi_\epsilon(-\Omega_1) - \Phi_\epsilon(\Omega_2), \quad \text{for} \quad \Omega_1 > \Omega_2
\]

\[
= 0, \quad \text{for} \quad \Omega_1 < \Omega_2 \quad \text{(B.5)}
\]

Hence, for \(P = P_{s,m}\) given by (6.23), the above procedure can be used by setting \(a = \phi_u(k_{uv}), a_1 = \phi_u(k_1), a_2 = \phi_u(k_2)\) and \(\epsilon = \epsilon_{uv}\). Likewise, to evaluate \(P_{k_1}\), we set \(a = \phi_u(k_1), a_1 = \phi_u(k_{uv})\) and \(a_2 = \phi_u(k_2)\); and to evaluate \(P_{k_2}\), we set \(a = \phi_u(k_2), a_1 = \phi_u(k_1)\) and \(a_2 = \phi_u(k_{uv})\).

\subsection*{B.2 Derivation of \(\Gamma_f\)}

Similar to Eqs. (6.27) and (6.28), the MSE in the case of failure of the whole disambiguation method, \(\Gamma_f\), is the resultant of three different cases, which dictates the following expression for \(\Gamma_f\):

\[
\Gamma_f = P \left[ \psi \left( k_u + \frac{\epsilon_{uv} - \epsilon_u}{2\pi} \right) \neq k_u \right] P_{s,m} \Gamma_{f,m} + P \left[ \psi \left( k_u + \frac{\epsilon_{uv} - \epsilon_u + 2\pi q_1 \mu_{uv}}{2\pi} \right) \neq k_u \right] P_{k_1} \Gamma_{f,1} + P \left[ \psi \left( k_u + \frac{\epsilon_{uv} - \epsilon_u + 2\pi q_2 \mu_{uv}}{2\pi} \right) \neq k_u \right] P_{k_2} \Gamma_{f,2} \quad \text{(B.6)}
\]

where \(\Gamma_{f,m}, \Gamma_{f,1}\) and \(\Gamma_{f,2}\) represent the MSEs associated with the three cases of failure that correspond to the probabilities \(P_{s,m}, P_{k_1}\) and \(P_{k_2}\), respectively. Benefiting from Eq. (6.30), Eq. (B.6) can be written in the form

\[212\]
\[ \Gamma_f = \{1 - [\Phi_{\varepsilon_{uv}}(\pi) - \Phi_{\varepsilon_{uv}}(-\pi)]\} P_{s,m} \Gamma_{f,m} + \{1 - [\Phi_{\varepsilon_{uv}}(\pi + 2\pi q_1 \mu_{uv}) - \Phi_{\varepsilon_{uv}}(-\pi + 2\pi q_1 \mu_{uv})]\} P_{k_1} \Gamma_{f,1} + \{1 - [\Phi_{\varepsilon_{uv}}(\pi + 2\pi q_2 \mu_{uv}) - \Phi_{\varepsilon_{uv}}(-\pi + 2\pi q_2 \mu_{uv})]\} P_{k_2} \Gamma_{f,2}. \] 

(B.7)

To obtain the expressions for \( \Gamma_{f,m}, \Gamma_{f,1} \) and \( \Gamma_{f,2} \), the corresponding errors need to be identified. These errors are given by \( \hat{k}_u - k_u \), where \( \hat{k}_u \) is the outcome of the RO step. In the case of failure, the difference \( \hat{k}_u - k_u \) has a minimum absolute value of unity. In the general case, the range of values taken by \( \hat{k}_u - k_u \) is determined by the probability distribution of the random variable \( \varepsilon_{uv} - \varepsilon_u \), which suggests that \( \hat{k}_u - k_u \) can be any value in \( \mathbb{Z} \) excluding zero. Therefore, each of \( \Gamma_{f,m}, \Gamma_{f,1} \) and \( \Gamma_{f,2} \) can be expressed as an infinite sum of the products of the MSEs corresponding to infinite integer values for \( \hat{k}_u - k_u \) and the probability of obtaining each integer. However, to simplify the analysis, we will ignore all the cases where the rounding function \( \Psi(.) \) produces an integer output that has an absolute difference from the noise-free output that is greater than unity. In other words, it will be assumed that the effect of the noise term \( (\varepsilon_{uv} - \varepsilon_u)/2\pi \) in the erroneous cases, will be confined to changing the noise-free round-off result by plus or minus one. For practical values of \( \sigma_u^2 \) and \( \sigma_v^2 \), this assumption is found reasonably accurate. Based on this assumption and (6.20), we can write

\[ \Gamma_{f,m} = \frac{\sigma_u^2 + (2\pi)^2}{\omega_u^2}. \] 

(B.8)

To obtain \( \Gamma_{f,i}, i = 1, 2 \); the corresponding outputs of the RO stage will be written as

\[ \hat{k}_{ui} = \Psi\left(k_u + \frac{\varepsilon_{uv} - \varepsilon_u + 2\pi q_i \mu_{uv}}{2\pi}\right) = k_u + \Psi(q_i \mu_{uv}) + \Psi\left[q_i \mu_{uv} - \Psi(q_i \mu_{uv}) + \frac{\varepsilon_{uv} - \varepsilon_u}{2\pi}\right]. \] 

(B.9)

where \(-0.5 \leq [q_i \mu_{uv} - \Psi(q_i \mu_{uv})] \leq 0.5\).
Now, define the probabilities

\[
P_{-1,i} \approx P \left\{ \Psi \left\{ [q_i \mu_{uv} - \Psi(q_i \mu_{uv})] + \frac{\epsilon_{uv} - \epsilon_u}{2\pi} \right\} = -1 \right\}
\]

\[
= P \left\{ [q_i \mu_{uv} - \Psi(q_i \mu_{uv})] + \frac{\epsilon_{uv} - \epsilon_u}{2\pi} \leq -0.5 \right\}
\]

\[
P_{1,j} \approx P \left\{ \Psi \left\{ [q_i \mu_{uv} - \Psi(q_i \mu_{uv})] + \frac{\epsilon_{uv} - \epsilon_u}{2\pi} \right\} = 1 \right\}
\]

\[
= P \left\{ [q_i \mu_{uv} - \Psi(q_i \mu_{uv})] + \frac{\epsilon_{uv} - \epsilon_u}{2\pi} \geq 0.5 \right\}
\]

\[
P_{0,i} \approx 1 - P_{-1,i} - P_{1,j}.
\]

(B.10)

From the above assumption on the rounding, the contribution of the rounding to the overall MSE of the estimated phase can be approximated as

\[
\Gamma_{r,i} \approx (2\pi)^2 \{P_{-1,i}[\Psi(q_i \mu_{uv}) - 1]^2 + P_{0,i}[\Psi(q_i \mu_{uv})]^2 + P_{1,i}[\Psi(q_i \mu_{uv}) + 1]^2\}
\]

(B.11)

and \(\Gamma_{f,j}\) can be obtained as

\[
\Gamma_{f,j} \approx \frac{\sigma_u^2 + \Gamma_{r,i}}{\omega_u^2}.
\]

(B.12)

Finally, by substituting for \(\Gamma_{f,m}\), \(\Gamma_{f,1}\) and \(\Gamma_{f,2}\) in Eq. (B.7), the final expression for \(\Gamma_f\) can be obtained.


