Distributed Detection and Estimation in Wireless Sensor Networks

Mohammad Fanaei

Dissertation submitted to the
Benjamin M. Statler College of Engineering and Mineral Resources
at West Virginia University

in partial fulfillment of the requirements
for the degree of

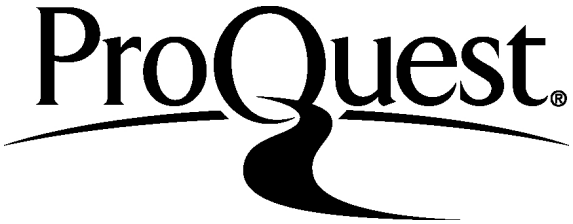
Doctor of Philosophy
in
Electrical Engineering

Matthew C. Valenti, Ph.D., Chair
Yaser P. Fallah, Ph.D.
Natalia A. Schmid, Ph.D.
Vinod K. Kulathumani, Ph.D.
Mark V. Culp, Ph.D.

Lane Department of Computer Science and Electrical Engineering

Morgantown, West Virginia
2016

Keywords: Distributed detection and classification, M-ary hypothesis testing, distributed estimation, best linear unbiased estimator (BLUE), maximum likelihood (ML), expectation maximization (EM), source localization, spatial collaboration, power allocation, Lloyd algorithm, limited feedback, spatial randomness, fusion center, wireless sensor network.



Abstract

Distributed Detection and Estimation in Wireless Sensor Networks

by

Mohammad Fanaei
Doctor of Philosophy in Electrical Engineering

West Virginia University

Matthew C. Valenti, Ph.D., Chair

Wireless sensor networks (WSNs) are typically formed by a large number of densely deployed, spatially distributed sensors with limited sensing, computing, and communication capabilities that cooperate with each other to achieve a common goal. In this dissertation, we investigate the problem of distributed detection, classification, estimation, and localization in WSNs. In this context, the sensors observe the conditions of their surrounding environment, locally process their noisy observations, and send the processed data to a central entity, known as the fusion center (FC), through parallel communication channels corrupted by fading and additive noise. The FC will then combine the received information from the sensors to make a global inference about the underlying phenomenon, which can be either the detection or classification of a discrete variable or the estimation of a continuous one.

In the domain of distributed detection and classification, we propose a novel scheme that enables the FC to make a multi-hypothesis classification of an underlying hypothesis using *only binary* detections of spatially distributed sensors. This goal is achieved by exploiting the relationship between the *influence fields* characterizing different hypotheses and the accumulated noisy versions of local binary decisions as received by the FC. In the realm of distributed estimation and localization, we make four main contributions: (a) We first formulate a general framework that estimates a vector of parameters associated with a deterministic function using spatially distributed noisy samples of the function for both analog and digital local processing schemes. (b) We consider the estimation of a scalar, random signal at the FC and derive an optimal power-allocation scheme that assigns the optimal local amplification gains to the sensors performing analog local processing. The objective of this optimized power allocation is to minimize the L^2-norm of the vector of local transmission powers, given a maximum estimation distortion at the FC. We also propose a variant of this scheme that uses a limited-feedback strategy to eliminate the requirement of perfect feedback of the instantaneous channel fading coefficients from the FC to local sensors through infinite-rate, error-free links. (c) We propose a linear spatial collaboration scheme in which sensors collaborate with each other by sharing their local noisy observations. We derive the optimal set of coefficients used to form linear combinations of the shared noisy observations at local sensors to minimize the total estimation distortion at the FC, given a constraint on the maximum average cumulative transmission power in the entire network. (d) Using a novel performance measure called the *estimation outage*, we analyze the effects of the spatial randomness of the location of the sensors on the quality and performance of localization algorithms by considering an energy-based source-localization scheme under the assumption that the sensors are positioned according to a uniform clustering process.

To My Parents
Rahim Fanaei and Mehrangiz Irannezhad
from whom I learned the meaning of selflessness
and in whom I saw the embodiment of sacrifice and dedication!

Acknowledgments

The last several years have been full of learning, personal and professional development, inspiration, as well as challenges. A lot of people have played an instrumental role in making these years an amazing experience for me, and I would like to seize this opportunity to thank them and show my sincere appreciations.

First, I would like to express my deepest gratitude to my amazing advisor and committee chair, Dr. Matthew C. Valenti. If it were not because of his support and his extending my research assistantship offer three times, I would not be in Morgantown in the first place, and this dissertation would not exist. He has been an excellent teacher, a great leader, and a perfect role model for me. When I joined the Ph.D. program in the Lane Department of Computer Science and Electrical Engineering at West Virginia University, I could not imagine how much growth and development it had in store for me, and I owe most of it to Dr. Valenti. Like any other student, I learned a lot of technical material from him both in his classes and in the one-on-one interactions that I had with him throughout the last several years. However, I learned a lot more invaluable lessens not from his words but from his actions, lessons such as patience and understanding, dedication to student development and growth, work ethics, optimistic open-mindedness, professionalism, and commitment to excellence. It would be a great challenge for me to strive to liken myself to him, a challenge that I believe will remain with me for the rest of my life. I truly feel honored and fortunate to have such an amazing role model and mentor and will always remember his contributions to my learning and personal development. I would like to thank him very much for giving me the opportunity to work with and learn from him. This dissertation would not be possible without his constant guidance, support, and encouragement.

Many thanks are due to Dr. Yaser. P. Fallah for his kind support and for accepting me in his research team over the last two years. I learned a lot from him in technical material in the field of automated-driving systems, lab management, and the art of forging and sustaining

close collaboration with industry. The work that I started with him has shaped my future research area as a new faculty member. Dr. Natalia A. Schmid played an influential role in the technical development of this dissertation. I learned a lot from her classes and always felt welcome in her office to discuss the technical roadblocks that I would encounter in this journey. Dr. Vinod K. Kulathumani deserves special thanks as the seed of my interest in the subject of wireless sensor networks was planted in his class in the first semester that I joined WVU. In fact, our first research paper in this subject was a result of his suggestion and encouragement. I would also like to thank Dr. Mark V. Culp for his time and suggestions in the final stages of the development of this dissertation. I am also grateful to Dr. Abbas Jamalipour at the University of Sydney, Australia for hosting me as a researcher under the National Science Foundation's East Asia and Pacific Summer Institute (EAPSI) fellowship program in summer 2013. Moreover, special thanks are due to Dr. Gaurav Bansal at Toyota InfoTechnology Center for his collaboration in our research in the field of congestion control in vehicular networks over the last two years.

Next, I would like to thank all past and current members of the wireless communications research lab with whom I have had the pleasure of working. In particular, I would like to thank my colleague Terry Ferrett, who has been a great help to me over the last few years. Moreover, special thanks are due to Dr. Xingyu Xiang, Dr. Salvatore Talarico, Chandana Nannapaneni, Aruna Sri Bommagani, Marwan Alkhweldi, and Veeru Talreja with whom I had a lot of technical and cultural discussions. Furthermore, I would like to express my gratitude to all past and current members of the cyber-physical systems lab. Especially, I am grateful to Amin Tahmasbi-Sarvestani for patiently answering my programing and NS3 questions and for being available day and night to work on our shared papers, Dr. Ehsan Moradi-Pari and Mir Mohammad Navidi for being available for occasional chat about daily life and future, S.M. Osman Gani for always smiling and being ready to help me debug my NS3 programs, Hadi Kazemi, Fereshteh Mahmoodzadeh, Hossein Mahjoub, Mehdi Iranmanesh, and Ahmad Jami for the deep discussions that we had on a wide range of subjects from scientific topics to political arguments.

Living in Morgantown provided me with a unique opportunity to make new friends and create many fun memories. A lot of friends made this journey memorable and joyful. I am going to try to name them to the best of my ability in the order that I can remember. First and foremost, I am grateful to my close friend, Sobhan Soleymani, who has always been available and open to discuss any personal or technical subject and to share his wisdom and

deep understanding. I was always assured that I can count on Sobhan to provide thoughtful, supportive input on any question or problem that I would encounter. I would also like to thank Amir Hossein Houshmandyar for always being present when I needed his help, Mehrzad Zahabi for creating so many memories of laughter and adventure, Dr. Mehrdad Shahabi, Dr. Hadi Rashidi, Reza Shisheie, Saeed Motiiyan, Ali Dehghan, and Behnam Khaki for creating so many memories of card games and fun, outdoor activities, and Yuan Li for his words of support and encouragement. I also owe a lot of thanks and appreciations to three of my roommates, Max Hunt, Caleb Stanley, and Mohammad Nasr Azadani. Max, Caleb, and Mohammad were a lot more than roommates and I consider them lifelong friends. I will never forget the lessons that they taught me.

Being thousands of miles away from family has changed the meaning of connectedness and has elevated the appreciation of their presence for me. My uncles, Mohammadreza and Sadegh Irannezhad, and their families have been extremely caring and supportive throughout the entire time that I have lived in the United States. Their presence and constant check-in and contact have made it a lot less foreign to me. Knowing that they would always be available to lend their hand has made all the difference in how I feel living here. I believe it would be very unlikely for me to decide to move to the United States had it not been because of them. My sister, Afsaneh, has been a role model in kindness from when I was a kid. She was my refuge as long as I remember. Anytime I would face a problem or would become confused, she was the one to listen and propose suggestions to relieve stress and confusion. My big brother, Hamidreza, has taught me how one can love you in deeds rather than the words. I was always confident that I can count on his help and guidance at the time of need. My parents, Rahim and Mehrangiz, are the "definition" of selflessness and have shown me the ultimate meaning of sacrifice for the comfort, well-being, and growth of another human being. Words do not even get close to do the justice to express how much I owe them, how much I love them, and how much I appreciate them. Without them, no success or growth would ever be possible for me. I dedicate this dissertation to my parents and would express my sincere gratitude and appreciation for their unbounded support, unconditional love, and all that they have done for me throughout my life.

Table of Contents

Dedication **iii**

Acknowledgments **iv**

List of Tables **vii**

List of Figures **viii**

List of Abbreviations **x**

Notation **xii**

1 Introduction **1**
1.1 System Model . . . 3
1.1.1 Local Observation Model . . . 4
1.1.2 Local Sensor Processing . . . 4
1.1.3 Communication Channel Model . . . 6
1.1.4 Fusion Center Processing . . . 7
1.2 Main Contributions and Dissertation Organization . . . 7

2 Low-Complexity Channel-aware Distributed M-ary Classification Using Binary Local Decisions **12**
2.1 Introduction . . . 12
2.2 Related Works to Distributed Binary Hypothesis-Testing Problem . . . 14
2.2.1 Network Topology . . . 18
2.2.2 Local Processing Schemes . . . 19
2.2.3 Resource Constraints . . . 20
2.2.4 Channel Capacity Constraints . . . 21
2.2.5 Wireless Channels Between Sensors and Fusion Center . . . 22

2.2.6 Correlated Local Observations . . . 24
2.3 Related Works to Distributed M-ary Hypothesis-Testing Problem . . . 25
2.4 System Model . . . 29
2.5 Derivation of the Optimal Fusion Rule . . . 32
2.6 Numerical Analysis . . . 37
2.6.1 Network Setup . . . 38
2.6.2 Effects of Observation and Channel signal-to-noise ratio (SNR) on Classification Performance . . . 39
2.6.3 Effect of the Number of Sensors K on Classification Performance . . . 41
2.7 Conclusions . . . 43

3 Distributed Parameter Estimation Using Non-Linear Observations 45
3.1 Introduction . . . 45
3.2 Related Works . . . 46
3.3 System Model and Problem Statement . . . 49
3.4 ML Estimation of $\boldsymbol{\theta}$ with Analog Local Processing . . . 51
3.5 ML Estimation of $\boldsymbol{\theta}$ with Digital Local Processing . . . 53
3.6 Linearized EM Solution for Digital Local Processing . . . 55
3.7 Case Study and Numerical Analysis . . . 60
3.7.1 Simulation Setup, Parameter Specification, and Performance-Measure Definition . . . 61
3.7.2 Effects of K, and Observation and Channel SNR on the Performance of Distributed Estimation Framework . . . 64
3.7.3 Effects of M on the Performance of Distributed Linearized EM Estimation Framework . . . 66
3.8 Conclusions . . . 69

4 Channel-aware Power Allocation for Distributed BLUE Estimation: Full and Limited Feedback of CSI 70
4.1 Introduction . . . 70
4.2 System Model . . . 72
4.3 Optimal Power Allocation with Minimal L^2-Norm of Transmission-Power Vector 74
4.4 Limited Feedback for Adaptive Power Allocation . . . 79
4.5 Codebook Design Using Lloyd Algorithm . . . 80
4.6 Numerical Analysis . . . 83
4.7 Conclusions . . . 86

5 Linear Spatial Collaboration for Distributed BLUE Estimation 87
5.1 Introduction . . . 87
5.2 Related Works . . . 88
5.3 System Model . . . 90
5.4 Derivation of Optimal Power Allocation . . . 93
5.5 Numerical Analysis . . . 97

5.6 Conclusions 98

6 Effects of Spatial Randomness on Source Localization with Distributed Sensors 100
6.1 Introduction 100
6.2 System Model 102
6.3 Derivation of ML Estimator and Cramér-Rao lower bound (CRLB) 105
6.3.1 Derivation of ML Estimator 106
6.3.2 Derivation of CRLB 107
6.3.3 Performance-Assessment Metric for Localization Schemes 108
6.3.4 Derivation of Optimal Local Quantization Thresholds 109
6.4 Numerical Performance Assessment 109
6.5 Spatial Dependence of Source Localization 111
6.5.1 Effect of Sensor Exclusion Zones on Source Localization 113
6.5.2 Effect of the Closest Sensors to Source on Localization 114
6.6 Conclusions 116

7 Summary and Future Work 117
7.1 Summary 117
7.2 Future Work 120
7.2.1 Extension of Chapter 4 120
7.2.2 Extension of Chapter 5 121
7.2.3 Extension of Chapter 6 121

References 123

List of Tables

2.1 Performance improvement due to globally optimizing local decision thresholds. 42

List of Figures

1.1 System model of a typical WSN used to make an inference about an underlying phenomenon θ. 4

2.1 System model of the proposed multi-hypothesis classification system. 29
2.2 An example network setup used in the numerical analysis. $K = 15$ sensors are distributed over an area with size $S = 15$. The sizes of the influence fields characterizing the non-null hypotheses are $A_1 = 5$ (shown by a dotted-line ellipse) and $A_2 = 15$ (shown by a solid-line rectangle). Assuming uniform distribution of the sensors within the observation environment, $K_1 = 5$ and $K_2 = 15$ are the average number of sensors that can be in the influence field of θ_1 and θ_2, respectively. The first $K_1 = 5$ sensors can be inside the influence field of either of the non-null hypotheses. The other $K - K_1 = 10$ sensors can only be inside the influence field of hypothesis θ_2. 38
2.3 Optimized average probability of correct classification at the FC versus observation SNR (ψ) for different values of channel SNR (η). There are $K = 15$ sensors randomly distributed in the observation environment. The local decision thresholds are optimized either globally (solid lines) or locally (dotted lines). 39
2.4 Optimized average probability of correct classification at the FC versus channel SNR (η) for different values of observation SNR (ψ). There are $K = 15$ sensors randomly distributed in the observation environment. The local decision thresholds are optimized either globally (solid lines) or locally (dotted lines). 41
2.5 Optimized average probability of correct classification at the FC versus the number of distributed sensors in the observation environment (K) for different values of the observation and channel SNRs. The local decision thresholds are optimized either globally (solid lines) or locally (dotted lines). 43

3.1 System model of a typical WSN for distributed estimation of a vector parameter $\boldsymbol{\theta}$. 50
3.2 Integrated mean-squared error (IMSE) versus the number of distributed sensors in the observation environment (K) for different values of the observation SNR (ψ) and channel SNR (η). 65

3.3 IMSE versus the observation SNR (ψ) for different values of the number of distributed sensors in the observation environment (K) and channel SNR (η). 67
3.4 IMSE versus the number of quantization levels at local sensors for linearized EM estimation at the FC based on *digital* local processing for different values of the number of distributed sensors in the observation environment (K), observation SNR (ψ), and channel SNR (η). 68

4.1 System model of a WSN in which the FC finds an estimate of θ. 73
4.2 Average energy efficiency versus the target estimation distortion D_0 for the proposed adaptive power-allocation scheme and the equal power-allocation strategy. 84
4.3 Average energy efficiency of the proposed power-allocation scheme versus the target estimation distortion D_0 for different values of the number of feedback bits L, when there are $K = 50$ sensors in the network. 85

5.1 System model of a WSN with error-free inter-sensor collaboration in which the FC finds an estimate of $\boldsymbol{\theta} \stackrel{\text{def}}{=} [\theta_1, \theta_2, \ldots, \theta_K]^T$. 91
5.2 Total estimation distortion at the FC versus the average cumulative transmission power for different degrees of spatial collaboration within two random network realizations. 98

6.1 The network topology of an example WSN consisting of $K = 50$ sensors denoted by '$\times$', whose objective is to localize a source target denoted by '$*$' and located at $(x_{\mathsf{T}}, y_{\mathsf{T}}) = (5, 10)$. The sensors are randomly placed in the circular surveillance region with radius $R = 50$ and centered at the origin according to a uniform clustering process. Each sensor is surrounded by an exclusion zone with radius $R_{\mathsf{ex}} = 5$, shown by a dotted circle around the sensor. A dashed circle with radius $R_{\mathsf{T}} = 14$ is depicted around the source target, within which there is $K_{\mathsf{T}} = 1$ sensor enclosed. 103
6.2 Functional system model of a WSN in which the FC localizes a source of energy. 104
6.3 Empirical RMSE of the source-location estimation, shown by solid lines, and its corresponding CRLB, shown by dashed lines, vs channel SNR in dB for three different random realizations of the network geometry, when the observation SNR is $\psi = 40$ dB. 110
6.4 Complementary cumulative distribution function (CCDF) of the empirical root mean-squared error (RMSE) of the source-location estimation and its corresponding CRLB vs the outage threshold γ for different settings of the network geometry. The observation SNR and channel SNR are fixed at $\psi = 40$ dB and $\eta = 0$ dB, respectively. 113
6.5 CCDF of the empirical RMSE of the source-location estimation vs the outage threshold γ for different values of R_{T} and K_{T}, when $R_{\mathsf{ex}} = 0$. The network geometries are generated without considering any sensor exclusion zones. The observation SNR and channel SNR are $\psi = 40$ dB and $\eta = 0$ dB, respectively. 115

List of Abbreviations

AWGN Additive white Gaussian noise

BLUE Best linear unbiased estimator

CCDF Complementary cumulative distribution function

CDF Cumulative distribution function

CRLB Cramér-Rao lower bound

CSI Channel state information

EGC Equal-gain combiner

EM Expectation maximization

FC Fusion center

FIM Fisher information matrix

GLE Geometric location-estimation error

i.i.d. Independent and identically distributed

IMSE Integrated mean-squared error

KKT Karush-Kuhn-Tucker conditions

LMMSE Linear minimum mean-squared error

MAC Multiple access channel

MEMS Micro-electromechanical systems

MGF Moment-generating function

ML Maximum likelihood

MMSE Minimum mean-squared error

MRC Maximum-ratio combining

MSE Mean-squared error

OOK On-off keying

PBPO Person-by-person optimization

pdf Probability density function

RF Radio frequency

RMSE Root mean-squared error

SNR Signal-to-noise ratio

WSN Wireless sensor network

Notation

Throughout this dissertation, the following notations and symbols are used:

$\mathrm{diag}(\cdot)$	This operator manipulates the diagonal elements of a matrix or puts a column vector into a diagonal matrix.
$\mathrm{Tr}\left[\cdot\right]$	Matrix trace operator
$\det(\cdot)$	Matrix determinant operator (for a square matrix)
$(\cdot)^T$	Matrix transpose operator
$(\cdot)^{-1}$	Matrix inverse operator
$(\cdot)^H$	Matrix complex-conjugate transpose operator
$\|\cdot\|$	Euclidean norm
$\|\cdot\|$	Cardinality of a set
$\|\cdot\|$	Absolute value of a complex number
$\delta[\cdot]$	Discrete Dirac delta function
$\mathbb{E}[\cdot]$	Expectation operator
$F_X(x)$	Cumulative distribution function (CDF) of a random variable X
$f_X(x)$	Probability density function (pdf) of a random variable X
$l(\cdot)$	Log-likelihood function (proportional to probability)
$\mathcal{N}\left(\mu, \sigma^2\right)$	Gaussian distribution with mean μ and variance σ^2
$\mathcal{CN}\left(\mu, \sigma^2\right)$	Circular complex Gaussian distribution with mean μ and variance σ^2
$Q(\cdot)$	Complementary distribution function of the standard Gaussian random variable defined as $Q(x) \stackrel{\text{def}}{=} \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp\left(-\frac{t^2}{2}\right) \mathrm{d}t$.

Unless otherwise stated, boldface upper-case letters denote matrices, boldface lower-case letters denote column vectors, and standard lower-case letters denote scalars.

Chapter 1

Introduction

Recent developments in integrated electronics, micro-electromechanical systems (MEMS), microprocessors, radio-frequency (RF) technologies, and ad hoc networking protocols have enabled low-power, low-cost sensors to collaborate with each other in order to perform complicated tasks. Wireless sensor networks (WSNs) are typically formed by a large number of densely deployed, spatially distributed sensors with limited sensing, computing, and communication capabilities that cooperate with each other to achieve a common goal [1, 2]. A few of the typical applications of WSNs are military surveillance [3], space explorations, habitat and environmental monitoring [4, 5], precision agriculture [6, 7], remote sensing, monitoring of industrial processes, tele-medicine and healthcare [8–11], traffic flow analysis [12], and underwater WSNs for marine environment monitoring and undersea explorations [13].

Among the most important functionalities of WSNs are *distributed detection*, *classification*, and *estimation*, which enable a wide range of applications such as event detection, classification, localization, system identification, and target tracking [14–26]. In a WSN performing distributed detection, classification, or estimation, spatially distributed sensors observe the conditions of their surrounding environment, where their noisy observations are correlated with an unknown phenomenon. The noisy observations are then locally processed and sent to a central entity, known as the fusion center (FC), for global processing and ultimate inference. The transmission of the sensors' processed data to the FC is performed through communication channels corrupted by fading and additive noise.[1] The FC will then

[1]Note that in practice, interference will also be available in the shared communication channel among sensors

combine the received information from spatially distributed sensors to make the ultimate global inference about the underlying phenomenon, which can be either the detection or classification of a discrete variable or the estimation of a continuous one. The local processing at the sensors can be in one of the following forms: (*a*) *Analog local processing* in which each sensor uses an amplify-and-forward strategy and sends an amplified version of its local analog noisy observations to the FC [17–20], and (*b*) *digital local processing* in which each sensor sends a quantized version of its local noisy observations to the FC [20–26].

In this dissertation, we investigate the problem of distributed detection, classification, and estimation in WSNs. In the domain of distributed detection and classification, we propose a novel scheme that enables the FC to make a multi-hypothesis classification of an underlying hypothesis using *only binary* decisions of spatially distributed sensors. This goal is achieved by exploiting the relationship between the *influence fields* characterizing different hypotheses and the accumulated noisy versions of local binary decisions as received by the FC, where the influence field of a hypothesis is defined as the spatial region in its surrounding in which it can be sensed using some sensing modality [27].

In the realm of distributed estimation, we make four main contributions:

(1) We first formulate a general framework that estimates a vector of parameters associated with a deterministic function using spatially distributed noisy samples of the function. We consider both analog and digital local processing schemes in our analyses.

(2) We consider the estimation of a scalar, random signal at the FC and derive an optimal power-allocation scheme that assigns the optimal local amplification gains to the sensors performing analog local processing. The objective of this optimized power allocation is to minimize the L^2-norm of the vector of local transmission powers, given a maximum estimation distortion at the FC. We also propose a variant of this scheme that uses a limited-feedback strategy to eliminate the requirement of perfect feedback of the instantaneous channel fading coefficients from the FC to local sensors through infinite-rate, error-free links.

---

unless the channels are orthogonal. Throughout this dissertation, the channels between local sensors and the FC are assumed to be orthogonal and therefore, we do not consider the effects of the interference in our analyses. If the channel is not orthogonal, the interference can be absorbed into the noise term.

(3) We propose a linear spatial collaboration scheme in which sensors collaborate with each other by sharing their local noisy observations. We derive the optimal set of coefficients used to form linear combinations of the shared noisy observations at local sensors to minimize the total estimation distortion at the FC, given a constraint on the maximum average cumulative transmission power in the entire network.

(4) We analyze the effects of spatial randomness of sensor locations on the performance of a recently proposed energy-based source-localization algorithm under the assumption that the sensors are positioned according to a uniform clustering process.

In the remainder of this chapter, a generic system model of the WSN studied in this dissertation is introduced in Section 1.1. This section defines the components of the network and sets the scene for the next chapters. In Section 1.2, a list of the main contributions of this dissertation and its organization are presented, and the results of different chapters are separately summarized.

1.1 System Model

The system model described in this section is a generic functional model for the WSNs studied in this dissertation. Unless otherwise stated, this system model is applicable to all chapters, and its details specific to each chapter that are different from the following specifications will be discussed in the corresponding chapter accordingly.

Consider a set of K spatially distributed sensors forming a WSN as shown in Figure 1.1. The ultimate goal of the WSN is to make an inference about an underlying phenomenon θ. In a *detection* or *classification* application, as considered in Chapter 2, the underlying phenomenon θ is a binary or M-ary variable, respectively, which can be either random or deterministic. In an *estimation* or *localization* application, as considered in Chapters 3–6, it is a continuous random or deterministic variable, which can be either a scalar, as considered in Chapter 4, or a vector, as considered in Chapters 3, 5, and 6. In the following subsections, different components of the system model shown in Figure 1.1 will be introduced.

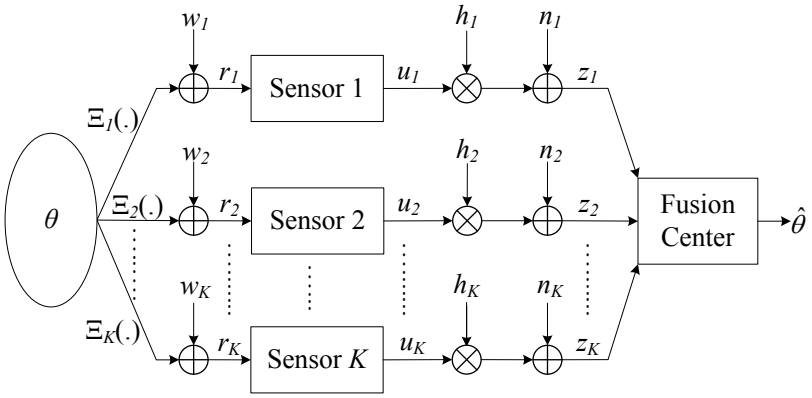


Figure 1.1: System model of a typical WSN used to make an inference about an underlying phenomenon θ.

1.1.1 Local Observation Model

Assume that each sensor makes a noisy observation that is correlated with the underlying phenomenon θ as

$$r_i = \Xi_i(\theta) + w_i, \qquad i = 1, 2, \ldots, K, \tag{1.1}$$

where r_i is the local noisy observation at the ith sensor, $\Xi_i(\cdot)$ is a *known* function of the underlying phenomenon observed at the ith sensor, and w_i is the spatially uncorrelated (except in Chapter 5 in which the observation noise at different sensors can be spatially correlated), zero-mean additive white Gaussian noise with known variance σ_i^2, i.e., $w_i \sim \mathcal{N}(0, \sigma_i^2)$.

Note that the observation model introduced in Equation (1.1) covers a large number of different applications. For example, it is a generalization of the well-studied linear observation model of vector $\boldsymbol{\theta} \stackrel{\text{def}}{=} [\theta_1, \theta_2, \ldots, \theta_p]^T$ in which $r_i = \mathbf{g}_i^T \boldsymbol{\theta} + w_i$, where $\mathbf{g}_i \stackrel{\text{def}}{=} \left[g_{i_1}, g_{i_2}, \ldots, g_{i_p}\right]^T$ is the vector of local observation gains at sensor i (considered, for example, by Ribeiro and Giannakis [24]). Furthermore, it is also a generalized version of the observation model considered by Niu and Varshney [28] and Ozdemir et al. [29].

1.1.2 Local Sensor Processing

Throughout this dissertation, except in Chapter 5, it is assumed that there is no inter-sensor communication and/or collaboration among spatially distributed sensors. Hence, each

sensor i processes its local noisy observations based on its local processing rule $\gamma_i(\cdot)$. The result of the local processing at sensor i, denoted by $u_i = \gamma_i(r_i)$, is sent to a FC through an impaired communication channel. In general, two local processing schemes can be considered for distributed sensors:

Analog Local Processing: In the analog local processing scheme, each sensor acts as a (pure) relay and uses an *amplify-and-forward* scheme to transmit an amplified version of its analog local noisy observations to the FC as

$$u_i = \alpha_i r_i = \alpha_i \Xi_i(\theta) + \alpha_i w_i, \qquad i = 1, 2, \ldots, K, \tag{1.2}$$

where u_i is the signal transmitted from sensor i to the FC, and α_i is the local amplification gain at sensor i. This scheme is considered in Chapters 3–5.

Digital Local Processing: In the digital local processing scheme, each sensor quantizes its local observations and sends their quantized version to the FC using a digital modulation format. This scheme is considered in Chapters 2, 3, and 6.

Suppose that sensor i quantizes its local noisy observation r_i to $b_i \stackrel{\text{def}}{=} \log_2 M_i$ bits, where M_i is the number of quantization levels at sensor i. Let $\mathcal{L}_i \stackrel{\text{def}}{=} \{\beta_i(0), \beta_i(1), \ldots, \beta_i(M_i)\}$ be the set of quantization thresholds at sensor i, where $\beta_i(\ell)$ is the ℓth quantization threshold of the ith sensor, $\beta_i(0) = -\infty$, and $\beta_i(M_i) = \infty$ for $i = 1, 2, \ldots, K$. The local processing rule at sensor i is then defined as a function $\gamma_i : \mathbb{R} \longmapsto \{0, 1, \ldots, M_i - 1\}$, whose values are determined as

$$u_i = \ell \Longleftrightarrow \beta_i(\ell) \leq r_i < \beta_i(\ell + 1), \qquad \ell = 0, 1, \ldots, M_i - 1 \text{ and } i = 1, 2, \ldots, K. \tag{1.3}$$

As it was mentioned, in all chapters except Chapter 5, the local processing is performed on *only* the local observations, which implicitly means that there is *no* inter-sensor communication and/or collaboration. However, the system model in Chapter 5 considers the case in which subsets of sensors can collaborate with each other by sharing their local noisy observations, where the local processing is performed on all of the observations available at each sensor. More details will be provided in Chapter 5.

1.1.3 Communication Channel Model

Suppose that all locally processed observations are transmitted to the FC over *parallel*, *independent* (*orthogonal*) fading channels. The channel between each sensor and the FC is assumed to be corrupted by fading and additive Gaussian noise. Assume that the received signal from sensor i at the FC is

$$z_i = h_i u_i + n_i, \qquad i = 1, 2, \ldots, K, \tag{1.4}$$

where h_i is the spatially independent multiplicative fading coefficient of the parallel channel between sensor i and the FC, and n_i is the spatially uncorrelated (except in Chapter 5 in which the noise at the communication channels between different sensors and the FC can be spatially correlated), zero-mean additive white Gaussian noise with known variance τ_i^2, i.e., $n_i \sim \mathcal{N}(0, \tau_i^2)$ (or $n_i \sim \mathcal{CN}(0, \tau_i^2)$ in Chapter 6).

Note that the above channel model implicitly makes the following assumptions:

(1) Each sensor is *only* synchronized with the FC. There is *no* need for any kind of time synchronization among spatially distributed sensors (as required in a coherent multiple-access channel).

(2) The distance-dependent path-loss in the communication channels between local sensors and the FC is fully compensated for all sensors using an appropriate power-control scheme [30]. Such power control makes the location of the FC irrelevant to the analyses. It should, however, be noted that sensors that are farther from the FC will deplete their energy resources faster.

In Chapters 2–5, the channel fading coefficients are assumed to be *completely known* at the FC. This assumption can be satisfied using any channel-estimation technique such as the transmission of pilot sequences from local sensors to the FC. In Chapter 6, it is assumed that the amplitude of the channel fading coefficients has a Rayleigh distribution and therefore, the random variable h_i is assumed to be spatially independent, zero-mean complex Gaussian with unit power, i.e., $h_i \sim \mathcal{CN}(0, 1)$. However, it is assumed that the FC does *not* have access to the instantaneous channel fading coefficients and that it *only* knows their distribution along with their first- and second-order statistics.

1.1.4 Fusion Center Processing

Upon receiving the vector of locally processed observations communicated through orthogonal channels by distributed sensors $\mathbf{z} \stackrel{\text{def}}{=} [z_1, z_2, \ldots, z_K]^T$, the FC combines them to make an inference about the underlying phenomenon θ, which is either its detection, classification, or estimation. For example, in Chapter 2, the FC adds all of the received information from the sensors to form a decision metric as the linear combination $\chi \stackrel{\text{def}}{=} \sum_{i=1}^{K} z_i$, based on which it classifies the underlying hypothesis θ using Bayesian decision theory. In Chapter 3, the FC finds the maximum likelihood (ML) estimate of a vector of deterministic parameters using the vector of the received information form spatially distributed sensors. In Chapters 4 and 5, the FC finds the best linear unbiased estimator (BLUE) of a scalar and vector, random signal, respectively, using the vector of the received data from local sensors $\mathbf{z}$.

1.2 Main Contributions and Dissertation Organization

In this section, a brief overview of the main contributions of this dissertation along with its organization will be summarized. Details of each contribution will be covered in a separate chapter.

Low-Complexity Channel-aware Distributed M-ary Classification Using Binary Local Decisions [31]: In Chapter 2, we investigate the problem of distributed multi-hypothesis classification of an underlying hypothesis at the FC of a WSN using local *binary* decisions. The binary decisions at spatially distributed sensors are made based on their noisy observations and sent to the FC through parallel additive white Gaussian noise (AWGN) channels. The FC uses the received noisy versions of local decisions to perform a global classification. In contrast with other approaches in the literature for multi-hypothesis classification based on combined binary decisions, our scheme exploits the relationship between the *influence fields* characterizing different hypotheses and the accumulated noisy versions of local binary decisions as received by the FC, where the influence field of a hypothesis is defined as the spatial region in its surrounding in which it can be sensed using some sensing modality [27]. The main contribution of Chapter 2 is the formulation of local and

fusion decision rules that maximize the probability of correct global classification at the FC, along with an algorithm for channel-aware global optimization of the decision thresholds at local sensors and the FC. The performance of the proposed classification system is investigated through studying practical scenarios. The results of numerical performance analyses show that the proposed approach simplifies decision making at the sensors while achieving acceptable performance in terms of the global probability of correct classification at the FC.

Main Publication

M. Fanaei, M.C. Valenti, N.A. Schmid, and V.K. Kulathumani, "Channel-aware distributed classification using binary local decisions," in *Proceedings of SPIE Signal Processing, Sensor Fusion, and Target Recognition XX*, (Orlando, FL), May 2011.

Distributed Parameter Estimation Using Non-Linear Observations [20]: In Chapter 3, we investigate the problem of estimating a vector of unknown parameters associated with a *deterministic* function at the fusion center of a wireless sensor network, based on the noisy samples of the function. The samples are observed by spatially distributed sensors, processed locally by each sensor, and communicated to the FC through *parallel* channels corrupted by *coherent* fading and additive white Gaussian noise. In our analyses, two local processing schemes at the sensors, namely *analog* and *digital*, will be considered. In the analog local processing scheme, each sensor acts as a pure relay and transmits an amplified version of its raw analog noisy observations to the FC. In the digital local processing method, each sensor quantizes its local noisy observations and sends the quantized samples to the FC using a digital modulation format. The FC combines all of the received locally processed observations and estimates the vector of unknown parameters. The main contribution of Chapter 3 is a generalized formulation of distributed parameter estimation in the context of WSNs, where local observations are *not* (necessarily) linearly dependent on the underlying parameters to be estimated and no specific observation model has been considered in the analyses.

Main Publication

M. Fanaei, M.C. Valenti, N.A. Schmid, and M.M. Alkhweldi, "Distributed parameter estimation in wireless sensor networks using fused local observations," in *Proceedings*

of SPIE Wireless Sensing, Localization, and Processing VII, vol. 8404, (Baltimore, MD), May 2012.

Channel-aware Power Allocation for Distributed BLUE Estimation – Full and Limited Feedback of CSI [32, 33]: In Chapter 4, we investigate the problem of finding the optimal local amplification gains in a distributed estimation framework in which the sensors use amplify-and-forward local processing. We propose an optimal, adaptive power-allocation strategy that minimizes the L^2-norm of the vector of local transmission powers, given a maximum estimation distortion defined as the variance of the BLUE estimator of a scalar, random signal at the FC. This approach prevents the assignment of high transmission power to sensors by putting a higher penalty on them, which results in the increased lifetime of the WSN compared to similar approaches that are based on the minimization of the sum of the local transmission powers.

The limitation of the proposed power-allocation scheme is that the optimal local amplification gains found based on it depend on the instantaneous fading coefficients of the channels between the sensors and FC. Therefore, the FC must feed the exact channel fading coefficients back to sensors through infinite-rate, error-free links, which is not a practical requirement in most applications of large-scale WSNs. In the remainder of Chapter 4, we propose a *limited-feedback strategy* to eliminate this requirement. The proposed approach is based on designing an optimal codebook using the generalized Lloyd algorithm with modified distortion metrics, which is used to quantize the space of the optimal power-allocation vectors used by the sensors to set their local amplification gains. Based on this scheme, each sensor amplifies its analog noisy observations using a quantized version of its optimal amplification gain determined by the designed optimal codebook.

Main Publications

M. Fanaei, M.C. Valenti, and N.A. Schmid, "Limited-feedback-based channel-aware power allocation for linear distributed estimation," in *Proceedings of Asilomar Conference on Signals, Systems, and Computers*, (Pacific Grove, CA), November 2013.

M. Fanaei, M.C. Valenti, and N.A. Schmid, "Power allocation for distributed BLUE estimation with full and limited feedback of CSI," in *Proceedings of IEEE Military*

Communications Conference (MILCOM), (San Diego, CA), November 2013.

Linear Spatial Collaboration for Distributed BLUE Estimation [34]: In Chapter 5, we investigate the problem of linear spatial collaboration for distributed estimation in a context in which each sensor can collaborate with a subset of other sensors by sharing its local noisy (and potentially spatially correlated) observations with them through error-free, low-cost links. A binary adjacency matrix defines the connectivity of the network and the pattern by which local sensors share their noisy observations with each other. The goal of the WSN is for a FC to estimate the vector of unknown signals observed by individual sensors. Each one of the sensors that is connected to the FC forms a linear combination of the noisy observations to which it has access and sends the result of this analog local processing to the FC through an orthogonal communication channel corrupted by fading and additive Gaussian noise. The FC combines the received data from spatially distributed sensors to find the BLUE estimator of the vector of unknown signals observed by individual sensors. The main novelty of Chapter 5 is the derivation of an optimal power-allocation scheme in which the set of coefficients or weights used to form linear combinations of shared noisy observations at the sensors connected to the FC is optimized. Through this optimization, the total estimation distortion at the FC (defined as the sum of the estimation variances of the BLUE estimators for different signals observed by individual sensors) is minimized, given a constraint on the maximum average cumulative transmission power in the entire network. Numerical results show that even with a moderate connectivity across the network, spatial collaboration among sensors significantly reduces the estimation distortion at the FC.

Main Publication

M. Fanaei, M.C. Valenti, A. Jamalipour, and N.A. Schmid, "Optimal power allocation for distributed BLUE estimation with linear spatial collaboration," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing* (ICASSP), (Florence, Italy), May 2014.

Effects of Spatial Randomness on Source Localization with Distributed Sensors [35]: The problem of estimating the location of a point source in WSNs has extensively been studied in the literature. Most of these studies assume that the source location

is estimated using the energy measurements of a set of spatially distributed sensors, whose locations are fixed. Because these sensors can randomly be distributed in the observation environment, both their observation quality and the performance of the localization algorithm depend on the realization of their random locations. Motivated by this fact, Chapter 6 analyzes the effects of spatial randomness of sensor locations on the performance of a recently proposed, energy-based source-localization algorithm under the assumption that the sensors are positioned according to a uniform clustering process. By introducing a novel performance measure called the estimation outage, it is investigated how the localization performance is affected by the parameters related to the network geometry such as the distance between the source and the closest sensor to it, the number of sensors within a region surrounding the source, as well as the existence and size of the exclusion zones around each sensor and the source.

Main Publication

M. Fanaei, M.C. Valenti, and N.A. Schmid, "Effects of spatial randomness on locating a point source with distributed sensors," in *Proceedings of IEEE International Conference on Communications Workshop on Advances in Network Localization and Navigation* (ICC–ANLN), (Sydney, Australia), June 2014.

Chapter 2

Low-Complexity Channel-aware Distributed M-ary Classification Using Binary Local Decisions

2.1 Introduction

One of the most important applications of wireless sensor networks (WSNs) is distributed detection and classification of an object, event, or phenomenon, also called an underlying hypothesis, which is the first step in a wider range of applications such as estimation, identification, and tracking [14]. Note that the presence of an object must first be ascertained before its attributes, such as location, movement pattern, heading, and velocity, can be estimated. Moreover, for WSNs that monitor infrequent events, the detection and classification of the event may be the main expected functionality. Furthermore, in some of the most important and widespread applications of WSNs, such as wireless surveillance, the detection of an intruder and its classification is the sole purpose.

The distributed detection and classification schemes usually consist of three main components: local processing of noisy observations, wireless communication of the locally processed data, and final data fusion. In a WSN performing distributed detection and classification, local sensors observe the conditions of their surrounding environment, process their local noisy observations, and send their processed data to a fusion center (FC), which makes the

ultimate global decision. The detection and classification at the FC must be performed despite the presence of faults in both sensor decisions and communication channels between local sensors and the FC. Different aspects of this problem have attracted a lot of interest in the research community throughout the last three decades.

In this chapter, we investigate the problem of distributed multi-hypothesis classification of an underlying hypothesis at the FC of a WSN using local binary decisions. The binary decisions at spatially distributed sensors are made based on their noisy observations and sent to the FC through parallel additive white Gaussian noise (AWGN) channels. The FC uses the received noisy versions of local decisions to perform a global classification. In contrast with other approaches in the literature for multi-hypothesis classification based on combined binary decisions, our scheme exploits the relationship between the *influence fields* characterizing different hypotheses and the accumulated noisy versions of local binary decisions as received by the FC, where the influence field of a hypothesis is defined as the spatial region in its surrounding in which it can be sensed using some sensing modality [27]. The main contribution of this chapter is the formulation of local and fusion decision rules that maximize the probability of correct global classification at the FC, along with an algorithm for channel-aware global optimization of the decision thresholds at local sensors and the FC. The performance of the proposed classification system is investigated through studying practical scenarios. The results of numerical performance analyses show that the proposed approach simplifies decision making at the sensors while achieving acceptable performance in terms of the global probability of correct classification at the FC.

The rest of this chapter is organized as follows: In Section 2.2, we present a detailed literature review on the field of distributed binary hypothesis-testing problem in which the number of underlying hypotheses is two. Section 2.3 concentrates on a through literature review on the distributed M-ary hypothesis-testing problem in which the number of underlying hypotheses is M. Having described the weaknesses of current distributed M-ary classification schemes, we will then analyze this problem from a new perspective. Section 2.4 describes the model of the distributed parallel fusion WSN that is considered in our analyses. In Section 2.5, the network is analyzed and the optimal parameters of a specifically defined decision rule at the FC are derived. Moreover, different methods of local versus global opti-

mization of sensor decision rules are discussed. Section 2.6 presents the numerical results of the analytical performance evaluations of the proposed classification system and studies the effects of different parameters of the classification network on its performance. Finally, we conclude our discussions and summarize the main achievements of this work in Section 2.7.

2.2 Related Works to Distributed Binary Hypothesis-Testing Problem

In the realm of distributed detection and classification in WSNs, most of the attention has been given to the binary hypothesis-testing problem in which the FC is designed to detect the presence or absence of an underlying hypothesis based on local binary (and potentially faulty) decisions received from spatially distributed sensors. In recent years, this problem has been considered for a practical case of non-ideal communication channels between local sensors and the FC in which the decisions of the sensors may *not* be reliably received at the FC (See [15, 36] and references therein for a survey on recent developments in this research area).

Decentralized detection with fusion was an active research area during the 1980s and early 1990s, following the ground-breaking work of Tenney and Sandell [37]. The main application of this research was *distributed radar*. To be more precise, it was assumed that K radars observing the same event were spatially distributed at different locations and that their decisions needed to be fused at a command center. At the time, the high cost of raw data transfer from local radars to the command center motivated researchers to propose novel approaches for local quantization and compression of data before transmitting the information to the FC; hence, the decentralized aspect of the problem arose. The goal of this research was to design the sensors and FC to detect the event as accurately as possible, subject to an alphabet-size constraint on the messages transmitted by each sensor. Note that aside from (potentially faulty) local processing in the above-described framework, the local decisions of distributed sensors are typically assumed to be reliably available at the FC in a *decentralized* scheme (in contrast with a *distributed* framework in which the locally processed

data is sent to the FC through impaired communication channels). A survey on the early works in the area of decentralized detection and classification can be found in [38–40] and references therein.

More recently, the applications of distributed detection and classification in WSNs have gained a lot of attention. Due to stringent resource constraints of wireless sensors, one should have a deep understanding of the interplay between local data processing, data compression, resource allocation, communication cost and reliability, and overall performance of the wireless sensor networks to be able to design an efficient distributed detection and classification approach. Classical results on inference problems in general, and on decentralized detection and classification in particular, can be extended to gain insight into and to form a basis for the efficient design of WSNs used in solving the distributed version of these problems.

In a *centralized* detection and classification system, all of the sensor observations are available at the FC *without any distortion* [15]. In the Bayesian problem formulation, the probability of error or misclassification at the FC is to be minimized, whereas in the Neyman–Pearson problem formulation for a binary detection system, the *probability of miss* (type–II error) is to be minimized, subject to a constraint on the maximum *probability of false alarm* (type–I error) [41, Chapter 2].

In classical decentralized detection and classification systems, spatially distributed sensors observe the state of their surrounding environment, represented by random variable θ. Based on its local noisy observation, sensor i selects one of D_i possible messages and sends it to the FC via a *dedicated* channel. Perfect reception of sensor outputs at the FC is typically assumed in a decentralized framework. The FC then produces an estimate of the state of the observation environment by selecting one of the possible hypotheses after reliably receiving all local data. Resource constraints in the classical decentralized detection and classification framework are addressed by fixing the number of sensors and imposing a finite-alphabet constraint on the output of each sensor. These constraints limit the amount of information available at the FC. It can be concluded that a decentralized sensor network in which every sensor sends a partial summary of its own observations to the FC is suboptimal compared to a centralized sensor network in which the FC has access to the observations of all sensors without any distortion [15]. Nevertheless, practical factors such as cost, spectral bandwidth

limitations, and complexity may justify the use of compression algorithms at distributed sensors. Furthermore, in systems with a large number of sensors, unprocessed information can flood and overwhelm the FC, and a centralized implementation of the optimal detection rule may simply be infeasible [15]. It should be noted that once the structure of the information provided by each sensor is fixed and known, the FC should solve a standard problem of statistical inference [38, 42]. Therefore, a *likelihood-ratio test* on the received data from the sensors will minimize the probability of error at the FC for a binary hypothesis-testing problem, and a minimum mean-square estimator will minimize the mean-squared error for an estimation problem [15].

One of the most important accomplishments in classical decentralized detection for binary hypothesis-testing problem is the demonstration that likelihood-ratio tests at the sensors are optimal when spatially distributed observations are conditionally independent, given each hypothesis [38]. This property drastically reduces the search space for an optimal collection of local quantizers and makes the resulting problem analytically tractable. The significance of this result is appreciated by the fact that the majority of research on classical decentralized detection assumes that local observations are conditionally independent and identically distributed (i.i.d.), given any hypothesis. In general, it is reasonable to assume conditional independence across sensors *only* if inaccuracies at local sensors are responsible for the noisy observations. However, if the observed process is itself stochastic or if the sensors are subject to external noise, this assumption may fail. Without the assumption of conditional independence, the task of finding an optimal solution to the classical decentralized detection problem is computationally intractable [43]. Even under the assumption of conditional independence, finding optimal quantization levels for distributed sensors is, in most cases, a difficult task [43]. This optimization problem is known to be tractable only under restrictive assumptions regarding the observation space and the topology of the underlying network. The solution does not scale well with the number of sensors except in some special cases, and it is not robust with respect to prior probabilities of the observation statistics.

A popular heuristic method to design decentralized detection systems is the *person-by-person optimization* (PBPO) technique [39]. In this approach, the decision rules are optimized for one sensor at a time while the local decision rules of the remaining sensors are

kept fixed. The index of the sensor being optimized is changed at every step. It is guaranteed that the overall performance of the detection rule at the FC is improved, or at least is not worsen, with every iteration of the PBPO algorithm. To be more precise, in a Bayesian setting for example, the probability of error at the FC will be a monotonically decreasing function of the number of PBPO iterations. One of the main disadvantages of this algorithm is that it does not necessarily result in a globally optimal solution and may only lead to a locally optimal one. There are several other important heuristic techniques for designing a decentralized detection system such as the saddle-point approximation method [44] and techniques based on empirical risk minimization and marginalized kernels [45]. In contrast with the majority of the works on decentralized detection and classification, the kernel method addresses system design for situations in which only a collection of empirical samples is available, i.e., the conditional joint distributions of the the sensor observations, given different hypotheses, are not needed to be known.

For networks with a small number of sensors used in decentralized detection and classification applications, the intuition regarding an optimal solution may be misleading. Consider a scenario in which local observations of different sensors are conditionally i.i.d. The symmetry in the problem suggests that the decision rules at the sensors should be identical, and identical local decision rules are, indeed, frequently assumed in many situations. However, counterexamples for which non-identical decision rules are optimal have been identified [38, 46, 47]. It is worth mentioning that identical decision rules are optimal, in terms of error exponent, for decentralized binary hypothesis-testing problem in the *asymptotic* regime, i.e., when the number of active sensors and (possibly) the area covered by these sensors increase to infinity [48]. In other words, it is proved that for any reasonable collection of transmission strategies of local decisions to the FC, the probability of error at the FC goes to zero *exponentially fast* as the number of the sensors K goes to infinity. Therefore, it is sufficient to compare different transmission strategies based on their exponential rate of convergence to zero, which is defined as follows [15]:

$$\lim_{K\to\infty} \frac{\log P_e\left(\mathcal{G}_K\right)}{K},$$

where $\mathcal{G}_K$ is a system configuration that contains K sensors, and $P_e\left(\cdot\right)$ denotes the probability

of error for a given system configuration.

The classical decentralized detection and classification framework cannot directly be applied to modern WSNs since it does not adequately take into account important features of sensor technology and of wireless links between local sensors and the FC. Particularly, as explained in [15]:

- Finite-alphabet restrictions on the sensor outputs do not adequately capture the resource constraints on spectral bandwidth and energy in WSNs.
- The assumption that sensor decisions are received reliably at the FC ignores the intrinsic characteristics of wireless links, particularly the fading effects.
- The emphasis of the research on the classical decentralized detection and classification problem has been on *optimal* solutions rather than *scalable* ones.

Many recent developments in the field of decentralized detection and classification in WSNs have been obtained by studying the classical problem while incorporating more realistic system assumptions in the problem definition. In particular, the following main assumptions have been considered in the problem formulation: network topology, local processing schemes, resource constraints of WSNs, channel capacity constraints, wireless fading channels between local sensors and the FC, and correlated local observations. The remainder of this section is devoted to describing the consequences of incorporating these assumptions in the problem of distributed detection and classification in WSNs.

2.2.1 Network Topology

Different network architectures for distributed sensor networks have been considered including the following main topologies:

- Carefully deployed WSNs usually form a *tree*, i.e., a network where nodes form a connected graph with no cycles [38]. In a tree structure, the information propagates from sensors to the FC in a straightforward manner, following a unique deterministic path. Therefore, the communication overhead is minimal.

- *Parallel* architecture is a subclass of the tree structure in which each sensor communicates directly with the FC. This configuration has received much of the attention in the distributed detection and classification literature since it is more analytically tractable. All WSNs considered throughout this dissertation are formed in the parallel structure.

- Distributed sensor systems can also form a *self-configuring* WSN in which the sensors are positioned in an observation environment *randomly* and then cooperate with each other to form a dynamic communication infrastructure. The price paid for the greater flexibility of self-configuring networks is a much more complicated communication mechanism with substantial overhead. The challenges that should be addressed in self-configuring WSNs include topology management, clustering, node identification, distributed synchronization, and the choice of routing policies. In these networks, nodes successively play the roles of sensors, relays, and routers. A reasonable assumption for distributed detection and classification using WSNs is that the sensors local to an event of interest are used for sensing and transmit their information to the FC using a single hop or multiple hops. The other sensors in the system may be used as relays or routers. The FC is responsible for final decision making and further relaying of the information across the network if necessary [15].

2.2.2 Local Processing Schemes

In-network signal processing can combine the information from neighboring sensors to improve the reliability of the local observations and reduce the amount of traffic in the network. On the other hand, the exchange of additional information among local sensors can potentially result in better decisions [49, 50]. For example, D'Costa et al. [50] have assumed that the local observations possess a correlating structure that extends only to a limited distance. If this assumption holds, the WSN can be partitioned into disjoint spatial coherence regions over which the signals remain strongly correlated. Local observations from different regions are assumed to be approximately conditionally independent. The resulting partitioning imposes a structure on the optimal decision rule that is suited to the communication constraints

of the network. Information is exchanged locally to improve the reliability of the measurements while compressed data is exchanged among coherence regions. Under mild conditions, the probability of error for the proposed classification scheme is found to decay exponentially to zero as the number of independent sensor measurements increases to infinity.

2.2.3 Resource Constraints

A problem formulation that better accounts for the physical resource constraints of a typical WSN is needed for accurate performance evaluations. As discussed earlier, wireless sensors often have very stringent power requirements. A limited spectral bandwidth and a bound on the total cost of the system may impose further constraints on system design. A flexible and appropriate solution to distributed sensing should account for these important factors. It can be proved that under the assumption of conditionally i.i.d. observations, using identical local processing rules for all sensors becomes asymptotically optimal as a global resource budget for the WSN goes to infinity, when the resource budget (instead of the number of sensors in classical detection and classification systems) is the fundamental design limitation [51]. Examples of the resource budget for a WSN include sum-rate constraint, total power requirement, a bound on the system cost, or a combination of them. A necessary condition for this result to hold is that the number of sensors must approach infinity as the actual resource budget goes to infinity. In this formulation, the resource budget (instead of the number of sensors) forms the fundamental constraint on the sensor system. Asymptotic analysis is performed based on an appropriate *local* metric, which decouples the optimization across sensors. The appropriate local metric is the *normalized Chernoff information*[1] in

[1]Suppose that r is a random (and potentially noisy) observation at a sensor that is to be classified into one of the two possible hypotheses θ_0 and θ_1. Let $f_0(r) \stackrel{\text{def}}{=} f_{R|\theta_0}(r|\theta_0)$ and $f_1(r) \stackrel{\text{def}}{=} f_{R|\theta_1}(r|\theta_1)$ be the conditional probability distributions of the local observation r, given hypothesis θ_0 and θ_1, respectively. The Chernoff information between the two probability distributions f_0 and f_1 is defined as [52, Section 11.9]

$$C(f_0; f_1) \stackrel{\text{def}}{=} -\min_{0\leq\lambda\leq 1} \log_2 \int_{-\infty}^{\infty} (f_0(r))^{\lambda} (f_1(r))^{1-\lambda} \, \mathrm{d}r.$$

Note that the Chernoff information is *not* symmetric, i.e., $C(f_0; f_1) \neq C(f_1; f_0)$. The normalization factor to find the normalized Chernoff information is the average transmission power of the sensor.

the Bayesian problem formulation and the *normalized relative entropy*[2] in Neyman–Pearson variant of the detection problem [51]. When the local observations are not conditionally i.i.d., these metrics can no longer be shown to be the right metrics for optimizing the local processing rules. However, even in this case, the asymptotic results can be used as some justification to decouple the optimization across sensors and to choose the local processing rule at each sensor to maximize the normalized Chernoff information (for the Bayesian problem formulation) or the relative entropy (for the Neyman–Pearson problem formulation).

2.2.4 Channel Capacity Constraints

The information-theoretic capacity or more generally, the admissible rate-region of a multiple-access channel is determined by its bandwidth, the signal power, and the noise power spectral density. Specifying these quantities is equivalent to fixing the *sum-rate* of the corresponding multiple-access channel. An initial approach to the capacity-constrained distributed detection and classification problem is to limit the sum-capacity of the multiple-access channel available to the sensors. More specifically, suppose that a multiple-access channel is only able to carry R bits of information per channel use. Then, the new design problem is converted to the selection of the number of active sensors K and the number of quantization levels at each sensor D_i, $i = 1, 2, \ldots, K$, with an objective to optimize the system performance at the FC, subject to the following capacity constraint:

$$\sum_{i=1}^{K} \lceil \log_2 D_i \rceil \leq R,$$

where $\lceil \cdot \rceil$ denotes the ceiling operation. It can be shown that for a binary hypothesis-testing problem in this framework, an identical binary quantization scheme at sensors is asymptotically optimal if there exists a binary quantization function whose Chernoff information exceeds half of the information contained in an unquantized observation [51]. This result

[2]The relative entropy or Kullback–Leibler distance between two probability distributions $f_0(r)$ and $f_1(r)$ is defined as [52, Section 2.3]

$$D\left(f_0(r) \parallel f_1(r)\right) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} f_0(r) \log_2 \left(\frac{f_0(r)}{f_1(r)}\right) \mathrm{d}r.$$

Note that the relative entropy is *not* symmetric, i.e., $D\left(f_0(r) \parallel f_1(r)\right) \neq D\left(f_1(r) \parallel f_0(r)\right)$. The normalization factor to find the normalized relative entropy is the average transmission power of the sensor.

explains the intuitive expectation that if the contribution of the first bit of the quantized data to the Chernoff information exceeds half of the Chernoff information offered by an unquantized observation, then using binary sensors is optimal. Note that the immediate consequence of this result is that having $K = R$ identical binary sensors is asymptotically optimal, i.e., the gain offered by having more sensors outperforms the benefits of getting detailed information from each sensor.

The above conclusion can be generalized to very important results that seem to be valid for a wide range of detection and classification problems. First, in most detection and classification configurations, the number of bits necessary to capture most of the information contained in each observation of each sensor appears to be very small [51]. In other words, for detection and classification purposes, most of the information provided by an observation can be found in the first few bits of the compressed data [53–56]. However, it is shown that the performance loss due to quantization decays very rapidly as the number of quantization levels increases. Therefore, message compression only plays a limited role in the overall system performance. A second result is that for conditionally i.i.d. observations, the diversity obtained by using multiple sensors more than offsets the performance degradation associated with receiving only coarse (one-bit) quantized data from each sensor [51].

2.2.5 Wireless Channels Between Sensors and Fusion Center

Most of the early results on classical decentralized detection and classification assume that each sensor produces a finite-valued function of its noisy observation, which is reliably received by the FC. In a WSN, this latter assumption of reliable transmission may fail since the information is transmitted over noisy channels corrupted by attenuation and fading. If sensors are randomly scattered in an observation environment, it can be assumed that their communication channels experience different mean path losses, with certain sensors possibly having much better connections than others. Furthermore, changes in the environment, interference, and motion of the sensors can produce time-dependent variations in the instantaneous quality of the wireless channels. The limitations posed by the wireless fading channels are made even worse by the fact that most detection and classification problems are

subject to stringent delay constraints. Therefore, the use of powerful error-correcting codes at the physical layer is prevented.

Most recent works on distributed detection incorporate the effects of the wireless channel on the transmission of local decisions to the FC. At least two distinct cases can be considered: The channel state information may or may not be available at local sensors.

If the channel state information is available at sensors, they can use adaptive transmission strategies, where a sensor decides what type of information to send to the FC based on the current quality of the communication channel. If the wireless channel is unreliable, then most of the available resources should be devoted to transmitting critical information. On the other hand, when the channel is in a good state, the sensor can potentially use a more complicated transmission scheme. In [55], Chen et al. have modified the classical decentralized detection problem by incorporating a fading channel between each sensor and the FC. They have derived a likelihood-ratio-based fusion rule for fixed decisions of the sensors. This optimal fusion rule requires perfect knowledge of the local performance metrics (e.g., the probability of detection and the probability of false alarm at each sensor) and the state of the communication channels over which messages are sent. Alternative fusion schemes that do not require as much side information are also proposed. A decision rule based on maximum-ratio combining (MRC) and a two-stage approach inspired by the Chair-Varshney decision rule [42] are analyzed. These concepts are further investigated by Niu et al. [57] for the scenario in which instantaneous channel state information is not available at the FC. They have proposed a fusion rule that requires *only* the knowledge of the channel statistics. At low signal-to-noise ratios (SNRs), the proposed fusion rule reduces to a statistic in the form of an equal-gain combiner (EGC), whereas at high SNRs, the proposed fusion rule is equivalent to the Chair-Varshney decision procedure.

As it was mentioned earlier, in distributed detection and classification, the most significant bit of a quantized observation seems to carry most of the information for the purpose of decision making. Therefore, it should be given more protection against noise and errors. This intuition is supported by the fact that sending a one-bit message outperforms schemes in which two bits of information are transmitted at a low SNR [51]. Conversely, multiple bits of quantized data can be transferred to the FC at high SNRs. The energy allocated

to each bit of the quantized observation is different for different channel conditions. At low SNRs, most of the energy is given to the most significant bit, whereas at higher SNRs, energy is split between the bits. For encoded systems, this requirement is equivalent to using error-correcting codes with unequal bit protection. Therefore, channel state information at the sensors increases the overall performance of the distributed detection and classification by allowing for the adaptation of the signaling schemes at individual sensors based on the fading level of their wireless communication channels.

Although specific system assumptions must be made to analyze WSNs with fading channels, several universal guidelines can be derived from their analyses. For example, it can be shown that the overall performance of the detection and classification system is optimized when the system uses as many independent sensors as possible, giving each sensor a minimal amount of system resources [51]. Furthermore, it can be observed that analog local processing schemes in which sensors transmit an amplified version of their local noisy observations to the FC perform better at low observation SNRs. However, digital local processing schemes in which sensors send a quantized version of their local noisy observations to the FC exhibit a superior performance above a certain threshold on the observation SNR [51].

2.2.6 Correlated Local Observations

When sensors in a WSN are densely packed in a finite area, their observations are likely to be correlated. Therefore, the results derived based on the popular assumption of conditionally independent local observations at the sensors do not necessarily hold. Different approaches have been employed to study the problem of distributed detection and classification under the assumption of correlated observations at sensors [58,59]. In [60], the binary quantization of a pair of dependent Gaussian random variables is analyzed. The results of this analysis indicate that even in this simple setting, an optimal detector may exhibit very complicated behavior. In [61], the structure of an optimal fusion rule is analyzed for a more general scenario of multiple binary sensors that observe conditionally dependent random variables. The structure of an optimal detector of weak signals from dependent observations is investigated in [46] by Blum and Kassam. They have also considered distributed detection

from dependent observations under a constant false-alarm rate in [62]. An adaptive fusion algorithm is proposed in [63] for an environment in which the observations and local decisions are dependent from one sensor to another. Blum [64] has presented a discussion on locally optimal detectors for correlated observations based on ranks. The numerical results contained in this work suggest that distributed detection schemes based on ranks and signs are less sensitive to the exact noise statistics compared to optimal schemes based directly on local observations. The theory of large deviations can also be employed to assess the performance of WSNs with correlated observations [65, 66].

Generally speaking, it can be noted that the correlation between local observations degrades the overall performance of the distributed detection and classification systems. It is interesting to note that the performance of the system with correlated observations improves with an increase in the sensor density. Note that in general, there are situations in which the performance does not necessarily improve with an increase in the sensor density. For example, in a scenario in which sensors try to detect the presence of a stochastic signal in Gaussian noise, the performance improves with sensor density only up to a certain point. Beyond this threshold, the performance starts to decay [65].

2.3 Related Works to Distributed M-ary Hypothesis-Testing Problem

The problem of testing M hypotheses using sensory data in WSNs has been investigated in some capacity (for instance, see [67]). In general, decisions made by the sensors in this case are sent to the FC using at least $\lceil \log_2 M \rceil$ information bits, where M is the number of hypotheses to be classified. However, two main constraints of WSNs make this approach undesirable: First, the processing power of the sensors is limited. Therefore, they may not be able to distinguish between different hypotheses. Second, the bandwidth and energy resources of WSNs are limited. Therefore, it is desired to send the sensor decisions to the FC with as few bits as possible. These requirements motivated us to design a distributed M-ary classification WSN in which the sensors make binary (rather than M-ary) decisions

and send them to the FC. The FC uses the local decisions collectively and makes a global inference about the underlying hypothesis based on the known influence fields of different hypotheses.

Few recent studies in the literature have investigated distributed M-ary hypothesis testing in WSNs using local binary decisions. In [68], a fault-tolerant distributed multi-hypothesis classification fusion approach is proposed based on *binary error-correcting codes*. In this approach, an error-correcting code matrix is designed in which each row forms a codeword that corresponds to one of the M hypotheses to be classified. Moreover, each column of the code matrix corresponds to the binary decision rule of a sensor. When sensor i detects hypothesis θ_j, it sends the binary element in the jth row and ith column of the code matrix to the FC through an orthogonal channel. The FC makes a final M-ary decision on the underlying hypothesis based on the received local binary decisions using the decoding criterion of *minimum Hamming distance*, where the Hamming distance between two binary vectors is defined as the number of distinct positions between the vectors. The performance of this multi-hypothesis classification WSN depends on the minimum Hamming distance of the designed binary code matrix. Note that in the classification algorithm proposed by [68], sensors still need to make an M-ary classification. Having made that classification, each sensor sends a binary decision to the FC. Therefore, this approach addresses the constraints of WSNs related to limited bandwidth and energy resources. However, it does not alleviate the requirement of high processing capability at local sensors.

The approach proposed in [68] does not consider the impact of fading channels between distributed sensors and the FC. In fact, this approach has assumed that the communication channels are binary symmetric channels. This weakness has been addressed in [69], which has a similar problem statement with a different decoding rule that is robust to flat-fading channels with phase coherent reception at the FC. Another proposed enhancement in [69] compared to [68] is that it allows the sensors to send multi-level D-ary (rather than binary) decisions to the FC, if needed, while the FC still uses a fixed binary code matrix for all values of D. The FC in this architecture uses a soft-decision decoding rule to measure the distance between a received multi-level local decision vector and a codeword in the given binary code matrix. It is shown in [69] that when more bits of local decision information

are sent to the FC, the classification performance can be improved while the total energy consumed by each sensor is fixed. In [70], the ideas presented in [69] for a binary code matrix are extended by using a D-ary code matrix with $D > 2$, when $\log_2 D$ bits of local decision information are used at the FC. In [71], the approach presented in [69] is further refined in a multiple-observation scenario while the sensor complexity is kept low. In this three-dimensional M-ary coded classification scheme, each sensor makes D (rather than one) independent observations and then sends D bits to the FC as the result of its local decisions for D observations, i.e., one bit for each observation rather than D bits for one observation as in [69]. Each hypothesis is, therefore, represented by a two-dimensional codeword, and the binary code matrix becomes three dimensional.

In [72], the problem of M-ary hypothesis classification in WSNs using local binary decisions is solved through modeling each sensor by a set of M transition probabilities that specify the probability that the sensor sends a binary message to the FC for different underlying hypotheses. Moreover, the FC is modeled by a set of M conditional misclassification probabilities, given any hypothesis. The authors of [72] have developed conditions for which the average probability of misclassification at the FC asymptotically goes to zero as the number of sensors goes to infinity. Moreover, they have used a genetic-algorithm-based approach to find the optimal local decision thresholds.

Zhang and Varshney [73] have considered the fusion of binary tree classifiers in a multi-hypothesis classification WSN. Binary decision trees make a sequence of binary decisions in a hierarchical manner, are easy to design, and are very efficient. In [73], this hierarchical tree structure is used to break the complex M-ary hypothesis-testing problem into a set of much simpler problems of binary decision fusion. Each sensor uses a binary decision tree to make its decision and sends it to the FC through an ideal communication channel. The FC combines the local decisions to make the global inference about the underlying hypothesis. Since each set of the received local decisions corresponds to a unique path from the root node to a terminal node of the binary decision tree, it can be encoded as a sequence of binary decisions made by all the sensors in the corresponding path. Detailed analysis of designing the binary decision trees for the sensors and for the FC, designing the decision rules at the internal nodes of the binary decision trees at the sensors, designing the optimal fusion rule,

and designing the system topology including communication structure of the WSN has been presented in [73].

In all of the aforementioned references, the conditional observations at different sensors, given any underlying hypothesis, are assumed to be independent. In [74], the problem of distributed Bayesian detection in WSNs with M hypotheses is considered, when sensors make conditionally dependent observations. Each sensor is modeled as a quantizer that makes a D-ary decision based on its observation and sends it to a FC through an ideal communication channel. The FC makes a decision on the actual hypothesis based on the local decisions it receives from spatially distributed sensors so that the average probability of misclassification is minimized. It is shown that due to the conditional dependence between sensor observations, the threshold-based decision rules (or likelihood ratios) at sensors are no longer optimal. The same problem has been considered in a more general form in [75] with conditionally correlated observations, given any underlying hypothesis, perfect communication channels between local sensors and the FC, and D-ary local decisions. The PBPO algorithm has been used to optimize the decision rules of local sensors and the FC, iteratively.

In some applications involving distributed M-ary hypothesis testing in WSNs, local knowledge of sensors may not be sufficient for making an M-ary decision, or it may be very costly to have sensors capable of doing such a classification. As an example, consider a surveillance system consisting of a densely deployed WSN whose ultimate goal is to detect and classify an intruder, which can be an person carrying a magnetic object, a motorcycle, or a vehicle. Suppose that the sensors are simple magnetometers that can only measure the strength of a magnetic field in their limited surrounding region. Since all three hypotheses may have the same magnetic field at a sensor location, the sensors may not be able to distinguish between these hypotheses based only on their local observations. In other words, they can only detect the absence or existence of a magnetic field in their surrounding, i.e., a local binary hypothesis problem. On the other hand, if the FC has access to local binary decisions made at all of the distributed sensors, it can make a global inference about the underlying hypothesis based on, for example, the number of sensors that have detected a magnetic field of one of the three hypotheses. More precisely, the number (and possibly location) of sensors that detect the presence of the magnetic field of an object determines

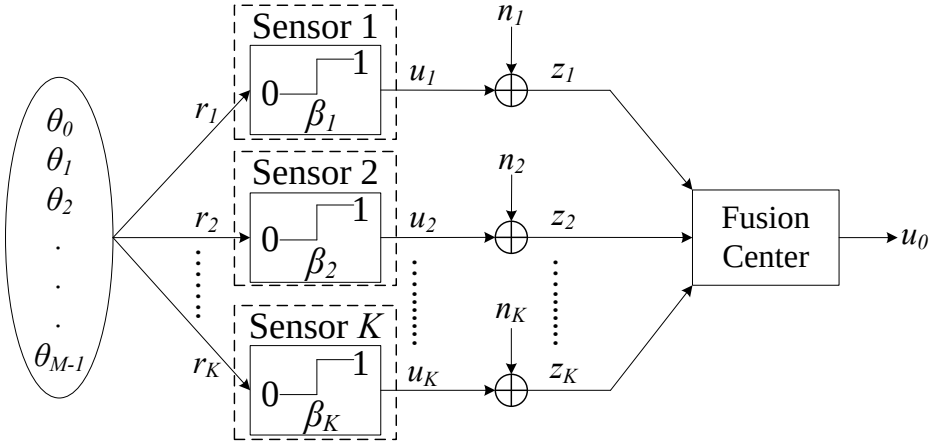


Figure 2.1: System model of the proposed multi-hypothesis classification system.

the coverage area of the object's influence field. This intuition motivated our work to design a distributed multi-hypothesis classification strategy for WSNs that uses local binary inferences to make its final optimal decision based on the knowledge of the influence fields of different underlying hypotheses.

Bapat et al. [76] have previously used the idea of influence fields for classification of objects in a large-scale WSN. The objective of [76] was to obtain requirements on the density of the underlying WSN to ensure accurate classification in the presence of false decisions at sensors, channel fading, and channel contention. However, our objective in this chapter is to obtain conditions on the decision thresholds at both local sensors and the FC for a *given density* of sensor deployment in order to maximize the classification accuracy at the FC.

2.4 System Model

Consider a WSN deployed as a parallel distributed classification system as shown in Figure 2.1. As described in Section 1.1, the system is formed by a FC and K sensors spatially distributed in an environment with area S. There are M independent and mutually exclusive hypotheses θ_0, θ_1, ..., and θ_{M-1}, $M \geq 2$, with the following known prior probabilities:

$$p_j \stackrel{\text{def}}{=} \mathbb{P}\left[\theta = \theta_j\right], \qquad j = 0, 1, \ldots, M-1,$$

where θ is a discrete random variable representing the underlying hypothesis. Note that θ_0 is the null or rejection hypothesis, and its existence means that none of the other $M-1$ hypotheses has occurred. Each non-null hypothesis is associated with a known *influence field* defined as the spatial region in its surrounding in which it can be sensed using some sensing modality [27]. As an example, suppose that the sensors are simple magnetometers and that the non-null hypotheses define the presence of a motorcycle or a vehicle. The regions in which the motorcycle or a vehicle can be sensed by the magnetometers are called their influence fields. The influence field of hypothesis θ_j is denoted by A_j, $j = 1, 2, \ldots, M-1$. It is assumed that the entire influence field of the underlying hypothesis is inside the observation area S. If the sensors are distributed uniformly within the observation area, the *average* number of sensors that can be placed in the influence field of hypothesis θ_j will be $K_j \stackrel{\text{def}}{=} \lfloor \frac{A_j}{S} K \rfloor$, where $\lfloor \cdot \rfloor$ denotes the floor operation. Throughout this chapter, we assume that the center of the influence field of the underlying hypothesis is known or has reliably been estimated.[3] Assuming that the center of the underlying influence field is known, the set of hypotheses is divided into two disjoint subsets for each sensor i: the set of hypotheses that sensor i cannot be inside their influence fields, denoted by $\mathcal{C}_i^0$, and the set of hypotheses that sensor i can be inside their influence fields, denoted by $\mathcal{C}_i^1$. On the other hand, assuming uniform sensor distribution within the observation environment, for each underlying hypothesis θ_j, *on average* there are K_j sensors that can be inside its influence field and $K - K_j$ sensors that cannot.

Let $\mathbf{r} \stackrel{\text{def}}{=} [r_1, r_2, \ldots, r_K]^T$ be the vector of sensor observations. It is assumed that the conditional observations of different sensors, given any specific underlying hypothesis, are independent. In other words,

$$f_{\mathbf{R}|\theta_j}(\mathbf{r}|\theta_j) = \prod_{i=1}^{K} f_{r_i|\theta_j}(r_i|\theta_j), \qquad j = 0, 1, \ldots, M-1, \tag{2.1}$$

where $f_{R|\theta_j}(r|\theta_j)$ denotes the conditional probability density function (pdf) of random variable R, given hypothesis θ_j. Given an underlying hypothesis θ_j, assume that each sensor i observes only noise if it cannot be inside the influence field characterizing hypothesis θ_j

[3]For more information on distributed estimation in WSNs, an interested reader is referred to the next chapters, [77], and references therein.

(i.e., $\theta_j \in \mathcal{C}_i^0$) and a noisy version of a *constant* intensity or strength of the influence field if it can be inside the influence field characterizing hypothesis θ_j (i.e., $\theta_j \in \mathcal{C}_i^1$). Therefore, the conditional observation function $\Xi_i(\cdot)$ in Equation (1.1) is defined as

$$\Xi_i(\theta_j) \stackrel{\text{def}}{=} \begin{cases} 0, & \text{if } \theta_j \in \mathcal{C}_i^0 \\ \vartheta, & \text{if } \theta_j \in \mathcal{C}_i^1 \end{cases} \qquad i = 1, 2, \ldots, K \text{ and } j = 0, 1, \ldots, M-1,$$

and the conditional noisy observation at the ith sensor, given any hypothesis θ_j, is found as

$$r_i|\theta_j = \begin{cases} w_i, & \text{if } \theta_j \in \mathcal{C}_i^0 \\ \vartheta + w_i, & \text{if } \theta_j \in \mathcal{C}_i^1 \end{cases} \qquad i = 1, 2, \ldots, K \text{ and } j = 0, 1, \ldots, M-1, \tag{2.2}$$

where ϑ is the *constant* intensity or strength of the influence field of any non-null hypothesis that can be sensed by sensors inside its influence field. Note that irrespective of the underlying hypothesis, the sensors observe the same constant intensity and cannot differentiate between different hypotheses that can create such an influence field. In other words, this model implies that the strength of the influence field of all non-null hypotheses is assumed to be the same and constant over the entire influence field. This assumption makes our analysis tractable. Furthermore, it is valid in a lot of applications such as the one mentioned at the beginning of this section.

Each sensor makes a binary decision based on its sensory data. To be specific, assume that the local decision of any sensor u_i is made based on a local binary decision rule as

$$u_i = \gamma_i(r_i) = \begin{cases} 0, & \text{if} \quad r_i < \beta_i \\ 1, & \text{if} \quad r_i > \beta_i \end{cases}, \tag{2.3}$$

where β_i is the optimal local decision threshold for sensor i. Note that this decision rule might not be the optimal local decision rule for our distributed classification system. However, it is very simple and allows us to achieve an acceptable performance in terms of the probability of correct classification at the FC without requiring the sensors to be able to distinguish between different hypotheses. In other words, the sensors are able to distinguish only the occurrence or not occurrence of any of the $M-1$ non-null hypotheses, and it is the FC that makes the final M-ary decision based on the accumulated local binary decisions. This local decision rule has two main advantages in satisfying the stringent processing capability

and bandwidth limitations of WSNs. The first advantage is that the sensors do not need to differentiate between $M-1$ non-null hypotheses and hence, their required local processing is very limited. The second advantage is that the transmission of the local decisions to the FC can be done using a binary scheme and hence, the bandwidth required for this communication is limited.

Let $\mathbf{u} \stackrel{\text{def}}{=} [u_1, u_2, \ldots, u_K]^T$ be the vector of binary decisions made by the sensors. The model of the communication channels between local sensors and the FC is a special case of the one introduced in Subsection 1.1.3, where an AWGN channel is considered with a unit channel gain, i.e., $h_i \equiv 1$. Therefore, each entry of the vector of the received signals from the sensors at the FC (i.e., $\mathbf{z} \stackrel{\text{def}}{=} [z_1, z_2, \ldots, z_K]^T$) can be found as

$$z_i = u_i + n_i, \qquad i = 1, 2, \ldots, K, \tag{2.4}$$

where n_i is spatially independent and identically distributed, zero-mean additive white Gaussian noise with known variance τ^2, i.e., $n_i \sim \mathcal{N}(0, \tau^2)$, $i = 1, 2, \ldots, K$. Note that unlike the channel model introduced in Subsection 1.1.3, the channel noise in this chapter is assumed to be identically distributed across sensors.

The FC has to make the final M-ary decision u_0 about the underlying hypothesis by using noisy versions of the local binary decisions received from spatially distributed sensors. In other words,

$$u_0 = \gamma_0(\mathbf{z}) \in \{0, 1, \ldots, M-1\},$$

where $\gamma_0(\cdot)$ is a multi-variate function. In the next section, we propose a simple yet powerful decision rule at the FC and analyze the performance of the proposed M-ary classification system.

2.5 Derivation of the Optimal Fusion Rule

Suppose that the receiver at the FC is designed to add all of the received signals from the sensors (i.e., z_i) and to form a *decision metric* (or test statistic) as $\chi \stackrel{\text{def}}{=} \sum_{i=1}^{K} z_i$ based on which the final M-ary decision is made. Note that the final decision metric is sought in the form of a linear combination of local noisy decisions. Other more complicated decision metrics can

be considered, but the linear form has the advantages of being simple and computationally efficient. It can be observed that χ is an appropriate yet simple decision metric, which captures differences in the influence fields of different hypotheses. In other words, it is intuitive to assume that χ tends to have a large value if the influence field of the underlying hypothesis is large. This conclusion is based on the fact that when the underlying influence field is large, more sensors can be located inside it and therefore, more sensors make $u_i = 1$ as their decisions. If the areas of the influence fields characterizing different hypotheses are distinct enough and if appropriate thresholds are found for the values of χ associated with different hypotheses, the system can achieve an acceptable performance in terms of the probability of correct classification at the FC.

Let $\Lambda \stackrel{\text{def}}{=} \{\lambda_1, \lambda_2, \ldots, \lambda_{M-1}\}$ be the set of decision thresholds based on which the FC classifies the underlying hypothesis using χ as its decision metric. In other words, assume that the decision rule at the FC is

$$u_0 = j \quad \text{if and only if} \quad \lambda_j \leq \chi < \lambda_{j+1}, \qquad\qquad j = 0, 1, \ldots, M-1, \tag{2.5}$$

where $\lambda_0 = -\infty$ and $\lambda_M = \infty$. The optimal values for this set of decision thresholds at the FC are derived in this section so that the maximum probability of correct classification at the FC can be achieved. Moreover, the effect of channel-aware global optimization of the set of sensors' binary decision thresholds $\{\beta_i\}_{i=1}^{K}$ on the probability of correct classification at the FC is examined.

Based on Equation (2.2) as the model for the conditional local noisy observation of sensor i (i.e., r_i) and Equation (2.3) as the sensor's local binary decision rule, it can be shown that the conditional pdf of the ith sensor's decision u_i, given hypothesis θ_j, is

$$f_{U_i|\theta_j}(u_i|\theta_j) = \begin{cases} Q\left(-\frac{\beta_i}{\sigma_i}\right)\delta[u_i] + Q\left(\frac{\beta_i}{\sigma_i}\right)\delta[u_i - 1], & \text{if } \theta_j \in \mathcal{C}_i^0 \qquad i = 1, 2, \ldots, K \\ Q\left(-\frac{\beta_i - \vartheta}{\sigma_i}\right)\delta[u_i] + Q\left(\frac{\beta_i - \vartheta}{\sigma_i}\right)\delta[u_i - 1], & \text{if } \theta_j \in \mathcal{C}_i^1 \qquad j = 0, 1, \ldots, M-1 \end{cases} \tag{2.6}$$

where $\delta[\cdot]$ is the discrete Dirac delta function, $Q(\cdot)$ is the complementary distribution function of the standard Gaussian random variable defined as

$$Q(x) \stackrel{\text{def}}{=} \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-\frac{t^2}{2}} \, \mathrm{d}t,$$

and $Q(-x) \equiv 1 - Q(x)$.

From Equation (2.4), it can be seen that given any decision at sensor i (i.e., u_i), the corresponding received signal at the FC z_i has a Gaussian distribution with mean u_i and variance τ^2, i.e., $z_i|u_i \sim \mathcal{N}(u_i, \tau^2)$, $i = 1, 2, \ldots, K$. Therefore, the conditional pdf of the received signal at the FC, given the corresponding decision at the sensor, is given by

$$f_{Z_i|U_i}(z_i|u_i) = \frac{1}{\sqrt{2\pi\tau^2}} e^{-\frac{(z_i-u_i)^2}{2\tau^2}}, \qquad i = 1, 2, \ldots, K. \tag{2.7}$$

Since K parallel communication channels are spatially independent, the received signals from different sensors at the FC (i.e., $\{z_i\}_{i=1}^K$) are also independent.

The conditional pdf of z_i, $i = 1, 2, \ldots, K$, given hypothesis θ_j, $j = 0, 1, \ldots, M-1$, is found as

$$f_{Z_i|\theta_j}(z_i|\theta_j) = \int_{-\infty}^{\infty} f_{Z_i|U_i}(z_i|u_i) \times f_{U_i|\theta_j}(u_i|\theta_j)\, \mathrm{d}u_i. \tag{2.8}$$

Substituting $f_{U_i|\theta_j}(u_i|\theta_j)$ and $f_{Z_i|U_i}(z_i|u_i)$ from Equations (2.6) and (2.7) into Equation (2.8) results in

$$f_{Z_i|\theta_j}(z_i|\theta_j) = \begin{cases} \frac{1}{\sqrt{2\pi\tau^2}} \left(e^{-\frac{z_i^2}{2\tau^2}} Q\left(-\frac{\beta_i}{\sigma_i}\right) + e^{-\frac{(z_i-1)^2}{2\tau^2}} Q\left(\frac{\beta_i}{\sigma_i}\right) \right), & \text{if } \theta_j \in \mathcal{C}_i^0 \\ \frac{1}{\sqrt{2\pi\tau^2}} \left(e^{-\frac{z_i^2}{2\tau^2}} Q\left(-\frac{\beta_i-\vartheta}{\sigma_i}\right) + e^{-\frac{(z_i-1)^2}{2\tau^2}} Q\left(\frac{\beta_i-\vartheta}{\sigma_i}\right) \right), & \text{if } \theta_j \in \mathcal{C}_i^1 \end{cases} \tag{2.9}$$

where the *sifting property* of the discrete Dirac delta function is applied. Using the conditional pdf of the received decision from each sensor i at the FC (i.e., z_i) under any underlying hypothesis θ_j, its conditional *moment-generating function* (MGF) is evaluated as

$$\Phi_{Z_i|\theta_j}(\nu) \overset{\text{def}}{=} \mathbb{E}\left[e^{\nu Z_i}\middle|\theta_j\right] = \int_{-\infty}^{\infty} e^{\nu z_i} f_{Z_i|\theta_j}(z_i|\theta_j)\, \mathrm{d}z_i = \mathcal{L}\left\{f_{Z_i|\theta_j}(z_i|\theta_j)\right\}\Big|_{s\to-\nu}, \tag{2.10}$$

where $\mathbb{E}[\cdot]$ and $\mathcal{L}\{\cdot\}$ denote the expectation operation and the Laplace transform of a function, respectively. Note that at the last step, the variable of the Laplace transform s is changed to $-\nu$. Substituting $f_{Z_i|\theta_j}(z_i|\theta_j)$ from Equation (2.9) into Equation (2.10) results in the conditional MGF of z_i, given the underlying hypothesis θ_j, as follows:

$$\Phi_{Z_i|\theta_j}(\nu) = \begin{cases} e^{\frac{\tau^2\nu^2}{2}} \left[Q\left(-\frac{\beta_i}{\sigma_i}\right) + Q\left(\frac{\beta_i}{\sigma_i}\right) e^{\nu} \right], & \text{if } \theta_j \in \mathcal{C}_i^0 \\ e^{\frac{\tau^2\nu^2}{2}} \left[Q\left(-\frac{\beta_i-\vartheta}{\sigma_i}\right) + Q\left(\frac{\beta_i-\vartheta}{\sigma_i}\right) e^{\nu} \right], & \text{if } \theta_j \in \mathcal{C}_i^1 \end{cases} \tag{2.11}$$

Using the MGF of z_i, $i = 1, 2, \ldots, K$, conditioned on hypothesis θ_j, $j = 0, 1, \ldots, M-1$, we can calculate the conditional MGF of the decision metric of the FC (i.e., $\chi \stackrel{\text{def}}{=} \sum_{i=1}^{K} z_i$), given hypothesis θ_j, as

$$\begin{aligned}\Phi_{\chi|\theta_j}(\nu) &\stackrel{\text{def}}{=} \mathbb{E}\left[e^{\nu\chi}\middle|\theta_j\right] = \mathbb{E}\left[e^{\nu\sum_{i=1}^{K} Z_i}\middle|\theta_j\right] = \mathbb{E}\left[\prod_{i=1}^{K} e^{\nu Z_i}\middle|\theta_j\right] \\ &\stackrel{(a)}{=} \prod_{i=1}^{K} \mathbb{E}\left[e^{\nu Z_i}\middle|\theta_j\right] \;=\; \prod_{i=1}^{K} \Phi_{Z_i|\theta_j}(\nu),\end{aligned} \tag{2.12}$$

where (a) is due to the independence of z_i's under a given hypothesis θ_j.

As mentioned in Section 2.4, assuming uniform sensor distribution within the observation environment, for any hypothesis θ_j, $j = 0, 1, \ldots, M-1$, there are *on average* K_j sensors for which $\theta_j \in \mathcal{C}_i^1$ and $K - K_j$ sensors for which $\theta_j \in \mathcal{C}_i^0$. Therefore, substituting $\Phi_{Z_i|\theta_j}(\nu)$ from Equation (2.11) into Equation (2.12), the conditional MGF of the decision metric of the FC, given hypothesis θ_j, can be written as

$$\begin{aligned}\Phi_{\chi|\theta_j}(\nu) = \exp\left(\frac{K\tau^2\nu^2}{2}\right) &\left[\prod_{i=1}^{K_j} Q\left(-\frac{\beta_i - \vartheta}{\sigma_i}\right) + Q\left(\frac{\beta_i - \vartheta}{\sigma_i}\right) e^{\nu}\right] \\ &\times \left[\prod_{i=K_j+1}^{K} Q\left(-\frac{\beta_i}{\sigma_i}\right) + Q\left(\frac{\beta_i}{\sigma_i}\right) e^{\nu}\right],\end{aligned} \tag{2.13}$$

where it is assumed that the sensors are ordered appropriately so that the first K_j sensors are inside the influence field characterizing hypothesis θ_j and the rest of them are outside of it. Note that $\Phi_{\chi|\theta_j}(\nu)$ can be simplified using algebraic manipulations to the final form of

$$\Phi_{\chi|\theta_j}(\nu) = \exp\left(\frac{K\tau^2\nu^2}{2}\right) \sum_{\ell=0}^{K} a_\ell \exp(\ell\nu), \tag{2.14}$$

where a_ℓ, $\ell = 0, 1, \ldots, K$ is the coefficient of the ℓth term in the above summation that can be expressed as a function of appropriate subsets of $Q\left(\frac{\beta_i - \vartheta}{\sigma_i}\right)$, $i = 1, 2, \ldots, K_j$ and $Q\left(\frac{\beta_i}{\sigma_i}\right)$, $i = K_j + 1, K_j + 2, \ldots, K$.

Based on the result of Equation (2.10), the conditional pdf of the decision metric of the FC χ, given hypothesis θ_j, is calculated from its MGF as

$$f_{\chi|\theta_j}(\chi|\theta_j) = \mathcal{L}^{-1}\left\{\Phi_{\chi|\theta_j}(-s)\right\}, \qquad j = 0, 1, \ldots, M-1, \tag{2.15}$$

where $\mathcal{L}^{-1}\{\cdot\}$ denotes the inverse Laplace transform of a function. Substituting $\Phi_{\chi|\theta_j}(\nu)$ from Equation (2.14) into Equation (2.15) results in the conditional pdf of the decision metric of the FC, given hypothesis θ_j, as follows:

$$f_{\chi|\theta_j}(\chi|\theta_j) = \frac{1}{\sqrt{2\pi K\tau^2}} \sum_{\ell=0}^{K} a_\ell \exp\left[-\frac{(\chi-\ell)^2}{2K\tau^2}\right]. \tag{2.16}$$

Based on the results of Bayesian decision theory, the minimum error probability decision rule for the M-ary classification of the underlying hypothesis using the decision metric of the FC χ is

$$\begin{aligned} \widehat{\theta_j} &= \underset{j\in\{0,1,\ldots,M-1\}}{\arg\max}\ f_{\theta_j|\chi}(\theta_j|\chi) \\ &= \underset{j\in\{0,1,\ldots,M-1\}}{\arg\max}\ p_j f_{\chi|\theta_j}(\chi|\theta_j). \end{aligned} \tag{2.17}$$

Therefore, substituting $f_{\chi|\theta_j}(\chi|\theta_j)$ from Equation (2.16) into Equation (2.17) results in the minimum error probability decision rule at the FC, which achieves the maximum probability of correct classification. Moreover, the optimal decision thresholds at the FC (i.e., $\Lambda \overset{\text{def}}{=} \{\lambda_1, \lambda_2, \ldots, \lambda_{M-1}\}$) can be found as the intersection of different conditional a posteriori pdfs $f_{\theta_j|\chi}(\theta_j|\chi)$, $j = 0, 1, \ldots, M-1$. The FC will then classify the underlying hypothesis based on its decision rule summarized in Equation (2.5) by using its decision metric $\chi \overset{\text{def}}{=} \sum_{i=1}^{K} z_i$.

It can be seen from Equation (2.13) that the coefficients a_ℓ, $\ell = 0, 1, \ldots, K$ in $f_{\chi|\theta_j}(\chi|\theta_j)$ are functions of the local binary decision thresholds β_i, $i = 1, 2, \ldots, K$. Based on Bayesian decision theory, the *locally* optimal decision rule of sensor i for a *binary* decision making on whether or not a non-null hypothesis has occurred is in the following form:

$$\mathbb{P}\left[\theta_j \in \mathcal{C}_i^1 | r_i\right] \underset{u_i=0}{\overset{u_i=1}{\gtrless}} \mathbb{P}\left[\theta_j \in \mathcal{C}_i^0 | r_i\right], \tag{2.18}$$

which can be rewritten as

$$\frac{f_{R_i|\{\theta_j\in\mathcal{C}_i^1\}}(r_i|\theta_j \in \mathcal{C}_i^1)}{f_{R_i|\{\theta_j\in\mathcal{C}_i^0\}}(r_i|\theta_j \in \mathcal{C}_i^0)} \underset{u_i=0}{\overset{u_i=1}{\gtrless}} \frac{\mathbb{P}\left[\theta_j \in \mathcal{C}_i^0\right]}{\mathbb{P}\left[\theta_j \in \mathcal{C}_i^1\right]}. \tag{2.19}$$

Considering the conditional observation model of sensor i, given hypothesis θ_j, defined in Equation (2.2), locally optimal decision rule of sensor i derived in Equation (2.19) can be

written as

$$\frac{\exp\left(-\frac{(r_i-\vartheta)^2}{2\sigma_i^2}\right)}{\exp\left(-\frac{r_i^2}{2\sigma_i^2}\right)} \underset{u_i=0}{\overset{u_i=1}{\gtrless}} \frac{\mathbb{P}\left[\theta_j \in \mathcal{C}_i^0\right]}{\mathbb{P}\left[\theta_j \in \mathcal{C}_i^1\right]},$$

which can be simplified to the following form:

$$r_i \underset{u_i=0}{\overset{u_i=1}{\gtrless}} \frac{\vartheta}{2} + \frac{\sigma_i^2}{\vartheta} \ln\left(\frac{\mathbb{P}\left[\theta_j \in \mathcal{C}_i^0\right]}{\mathbb{P}\left[\theta_j \in \mathcal{C}_i^1\right]}\right). \tag{2.20}$$

If we compare the above local decision rule with the one defined in Equation (2.3), the locally optimal binary decision threshold of sensor i can be defined as

$$\beta_{i,\mathsf{Local}} = \frac{\vartheta}{2} + \frac{\sigma_i^2}{\vartheta} \ln\left(\frac{\mathbb{P}\left[\theta_j \in \mathcal{C}_i^0\right]}{\mathbb{P}\left[\theta_j \in \mathcal{C}_i^1\right]}\right). \tag{2.21}$$

It can be observed from Equation (2.21) that $\beta_{i,\mathsf{Local}}$ depends *only* on the constant intensity of the influence field of any non-null hypothesis ϑ and the variance of the additive observation noise σ_i^2. However, this decision threshold might not result in the globally optimized probability of correct classification at the FC. In this work, our goal is to find the *globally* optimal local decision thresholds that result in the maximum probability of correct classification at the FC. It should be noted that these globally optimal local decision thresholds depend on the variances of both observation noise and channel noise (i.e., σ_i^2 and τ^2). In the next section, we present the results of our analysis using a numerical scenario and discuss the effects of such a global optimization of the local binary decision thresholds on the performance of the M-ary classification system compared to their local optimization.

2.6 Numerical Analysis

In this section, the performance of the proposed channel-aware multi-hypothesis classification architecture is evaluated for a typical numerical scenario. First, the parameters of the WSN under analysis are specified. Then, the effects of the observation SNR and channel SNR on the performance of the classification system are investigated. Moreover, the performance enhancement that can be achieved by optimizing sensors' decision thresholds globally rather than locally is discussed. Finally, the effects of the number of distributed sensors on the performance of the proposed classification system are evaluated.

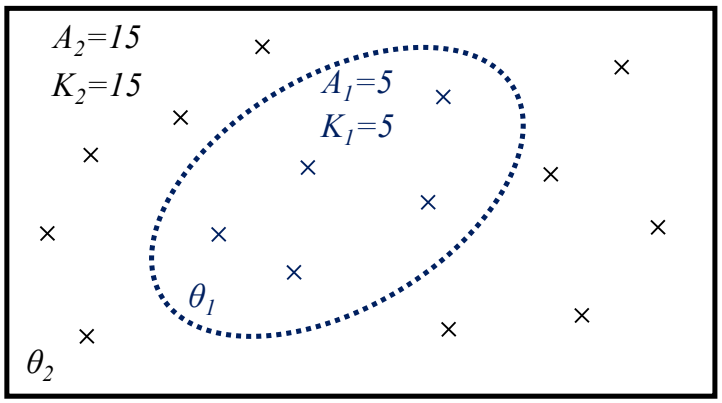


Figure 2.2: An example network setup used in the numerical analysis. $K = 15$ sensors are distributed over an area with size $S = 15$. The sizes of the influence fields characterizing the non-null hypotheses are $A_1 = 5$ (shown by a dotted-line ellipse) and $A_2 = 15$ (shown by a solid-line rectangle). Assuming uniform distribution of the sensors within the observation environment, $K_1 = 5$ and $K_2 = 15$ are the average number of sensors that can be in the influence field of θ_1 and θ_2, respectively. The first $K_1 = 5$ sensors can be inside the influence field of either of the non-null hypotheses. The other $K - K_1 = 10$ sensors can only be inside the influence field of hypothesis θ_2.

2.6.1 Network Setup

Figure 2.2 shows a typical WSN formed by $K = 15$ sensors distributed over an area with size $S = 15$ that is used to analyze the performance of the proposed multi-hypothesis classification system. The goal is to classify an underlying hypothesis using the distributed observed data generated by $M = 3$ hypotheses, θ_0, θ_1, and θ_2, with known prior probabilities $p_0 \stackrel{\text{def}}{=} \mathbb{P}[\theta = \theta_0] = 0.6$, $p_1 \stackrel{\text{def}}{=} \mathbb{P}[\theta = \theta_1] = 0.3$, and $p_2 \stackrel{\text{def}}{=} \mathbb{P}[\theta = \theta_2] = 0.1$. The influence fields of the non-null hypotheses are of size $A_1 = 5$ and $A_2 = 15$. Therefore, assuming uniform distribution of the sensors within the observation environment, $K_1 = 5$ and $K_2 = 15$ are the *average* number of sensors that can be in the influence field of θ_1 and θ_2, respectively. Assume that $\vartheta = 1$ is the normalized strength of the observable influence field of non-null hypotheses. Without loss of generality, suppose that the observation noises are identically distributed across sensors, i.e., $\sigma^2 \stackrel{\text{def}}{=} \sigma_i^2$, $i = 1, 2, \ldots, K$.

Since there are two non-null hypotheses θ_1 and θ_2 in this example, the sensors are divided into two disjoint groups. The first group is composed of $K_1 = 5$ sensors that can be inside

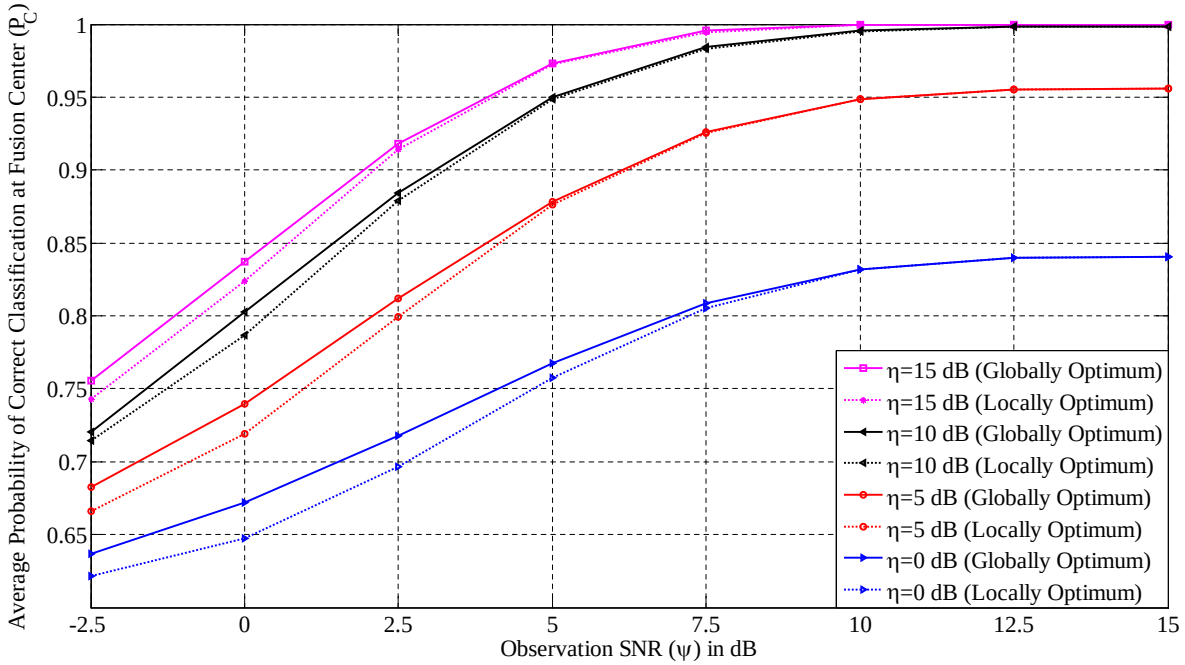


Figure 2.3: Optimized average probability of correct classification at the FC versus observation SNR (ψ) for different values of channel SNR (η). There are $K = 15$ sensors randomly distributed in the observation environment. The local decision thresholds are optimized either globally (solid lines) or locally (dotted lines).

the influence field of either of the non-null hypotheses. The second group is formed by the other $K - K_1 = 10$ sensors that can only be inside the influence field of hypothesis θ_2. It is intuitive to assume that the decision thresholds of the sensors in each one of these two groups are the same. Therefore, the set of local decision thresholds is composed of 15 elements each one of them is associated with one sensor. The first five elements of this set are all equal. Similarly, the last ten elements of this set are all equal. In the rest of this section, we refer to these two local decision thresholds as β_1 and β_2, respectively.

2.6.2 Effects of Observation and Channel SNR on Classification Performance

The average optimized probability of correct classification at the FC, denoted by P_c, versus observation SNR (ψ) is shown in Figure 2.3 for different values of channel SNR (η). The

observation and channel SNRs are defined as

$$\psi \stackrel{\text{def}}{=} \frac{1}{2\sigma^2} \text{ and } \eta \stackrel{\text{def}}{=} \frac{1}{2\tau^2},$$

where σ^2 and τ^2 are variances of the observation noise and channel noise, respectively. Figure 2.4 shows the optimized average probability of correct classification at the FC versus channel SNR for different values of observation SNR. SNR values are given in dB. Solid lines show P_{c} when the local decision thresholds are optimized globally through an exhaustive search over all possible local threshold values between zero and two with step size 0.05. Dotted lines show P_{c} when the local decision thresholds are optimized locally and derived based on Equation (2.21). For the WSN under consideration, $\beta_{1,\mathsf{Local}}$ and $\beta_{2,\mathsf{Local}}$ can be written as

$$\begin{aligned} \beta_{1,\mathsf{Local}} &= \frac{1}{2} + \sigma^2 \ln\left(\frac{p_0}{p_1+p_2}\right) \\ \beta_{2,\mathsf{Local}} &= \frac{1}{2} + \sigma^2 \ln\left(\frac{p_0+p_1}{p_2}\right). \end{aligned} \tag{2.22}$$

As it can be seen from Figures 2.3 and 2.4, the average probability of correct classification approaches one as SNR increases. Moreover, under all SNR regimes, channel-aware classification system, which is based on the globally optimal local decision thresholds, outperforms conventional classification system, which is based on locally optimal local decision thresholds. This crucial point is further demonstrated in detail in Table 2.1. In this table, the values of local decision thresholds derived from both local optimization and global optimization are shown in different columns. Moreover, the corresponding optimized average probability of correct classification at the FC is shown for each case. When the observation SNR ψ is fixed, the locally optimal decision thresholds are also fixed based on Equation (2.22). However, globally optimal decision thresholds change with both observation SNR ψ and channel SNR η. In the last column of the Table 2.1, the percentage improvement in the average probability of correct classification at the FC due to the global optimization of decision thresholds is shown. Note that as SNR increases, the achievable percentage improvement decreases. In other words, global optimization of decision thresholds does not improve the average probability of correct classification at high SNRs. A justification for this conclusion is that for

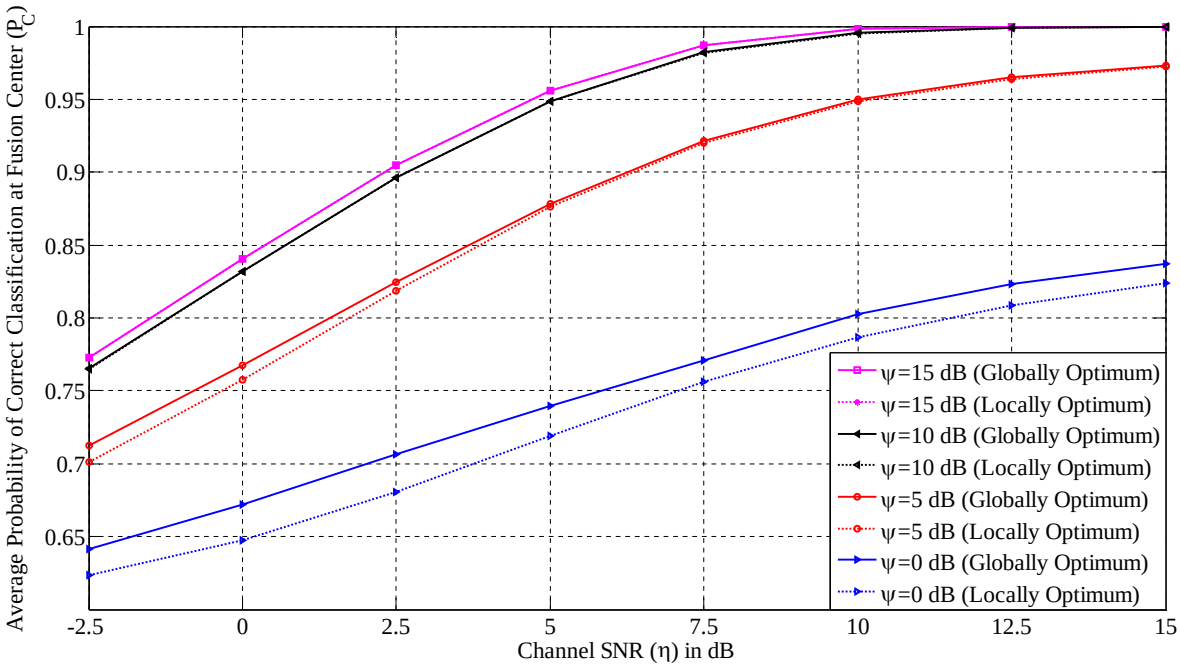


Figure 2.4: Optimized average probability of correct classification at the FC versus channel SNR (η) for different values of observation SNR (ψ). There are $K = 15$ sensors randomly distributed in the observation environment. The local decision thresholds are optimized either globally (solid lines) or locally (dotted lines).

high SNRs, the probability of correct classification at the FC is very high (near one) by itself and cannot be increased further.

2.6.3 Effect of the Number of Sensors K on Classification Performance

The performance of the proposed multi-hypothesis classification system is a function of the number of distributed sensors in the observation environment. In Figure 2.5, the optimized average probability of correct classification at the FC P_c is shown versus the number of distributed sensors in the observation environment K for different values of the observation and channel SNRs. Solid lines are used to indicate that the local decision thresholds are optimized globally through an exhaustive search over all possible threshold values between zero and two with step size 0.05. Dotted lines show P_c when the local decision thresholds are optimized locally and derived based on Equation (2.22). As it can be seen in Figure 2.5,

Table 2.1: Performance improvement due to globally optimizing local decision thresholds.

ψ	η	$\boldsymbol{\beta} \stackrel{\text{def}}{=} \{\beta_1, \beta_2\}$		P_{c}		Percentage Improvement
		Local	Global	Local	Global	
0	0	$\{0.70, 1.6\}$	$\{0.55, 0.65\}$	0.647	0.672	3.82%
	5		$\{0.55, 0.75\}$	0.719	0.740	2.92%
	10		$\{0.5, 1.15\}$	0.787	0.803	1.97%
	15		$\{0.4, 1.3\}$	0.824	0.837	1.59%
5	0	$\{0.56, 0.85\}$	$\{0.5, 0.6\}$	0.757	0.767	1.33%
	5		$\{0.5, 0.8\}$	0.876	0.878	0.27%
	10		$\{0.5, 0.9\}$	0.949	0.951	0.16%
	15		$\{0.5, 0.85\}$	0.972	0.973	0.11%
10	0	$\{0.52, 0.61\}$	$\{0.5, 0.55\}$	0.831	0.832	0.06%
	5		$\{0.5, 0.65\}$	0.948	0.949	0.03%
	10		$\{0.5, 0.75\}$	0.995	0.996	0.06%
	15		$\{0.5, 0.8\}$	0.9996	0.9997	0.02%
15	0	$\{0.51, 0.54\}$	$\{0.5, 0.5\}$	0.840	0.841	$\simeq 0\%$
	5		$\{0.5, 0.55\}$	0.955	0.956	$\simeq 0\%$
	10		$\{0.5, 0.65\}$	0.997	0.998	$\simeq 0\%$
	15		$\{0.5, 0.7\}$	0.999	0.999	$\simeq 0\%$

as the number of the sensors increases, the optimized average probability of correct classification at the FC also increases. Furthermore, our classification system shows an acceptable performance in terms of the average probability of correct classification for moderate number of sensors. Notice that since the proposed classification system works based on the number of sensors that are in the influence field of each hypothesis, if the number of sensors is very small, the number of sensors that can be in the influence field of different hypotheses is almost the same. Therefore, the value of the conditional decision metric under different hypotheses is not distinct enough for the FC to be able to distinguish between them. This problem is more important for low-SNR regimes.

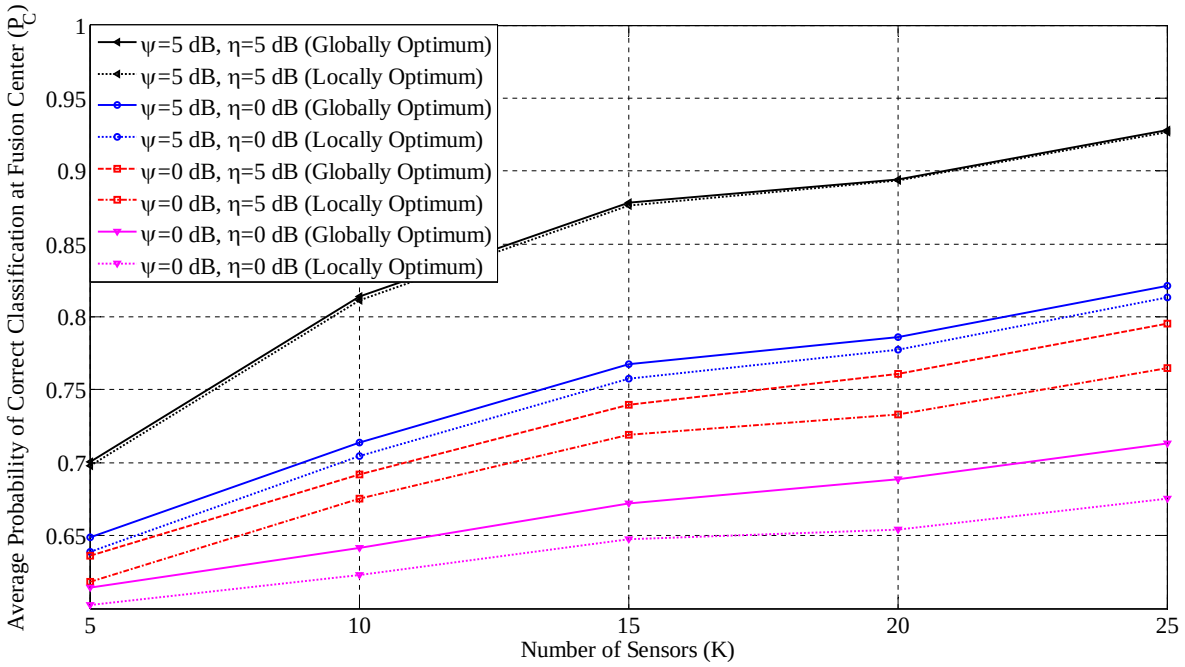


Figure 2.5: Optimized average probability of correct classification at the FC versus the number of distributed sensors in the observation environment (K) for different values of the observation and channel SNRs. The local decision thresholds are optimized either globally (solid lines) or locally (dotted lines).

2.7 Conclusions

In this chapter, we had an extensive literature review on the distributed detection and classification in WSNs and summarized the results of major research accomplishments in this area up to date. In particular, we specified major results in the fields of both binary and M-ary distributed classification in WSNs and identified major solutions and challenges for each class of problems. Furthermore, we designed a method to optimize the performance of a distributed WSN deployed as a multi-hypothesis classification system. The sensors employ a simple binary decision rule and make decisions based on their noisy observations. These binary decisions are sent to the FC through parallel AWGN channels. The FC then forms a decision metric as the linear combination of these local noisy decisions, which will be used to perform a global multi-hypothesis classification based on the known influence fields of different hypotheses. Fusion decision rule was formulated and numerical performance analysis of an example WSN was presented to investigate the effects of the observation and

channel SNR, and the number of distributed sensors, on the classification performance. The results of numerical analyses showed that the proposed approach simplifies decision making at the sensors while achieving an acceptable performance in terms of the global average probability of correct classification at the FC. Furthermore, it was shown that a global optimization of the local decision thresholds improves the probability of correct classification at the FC compared to the case in which local thresholds are only locally optimized.

Chapter 3

Distributed Parameter Estimation Using Non-Linear Observations

3.1 Introduction

In this chapter, we investigate the problem of estimating a vector of unknown parameters associated with a *deterministic* function at the fusion center of a wireless sensor network, based on the noisy samples of the function. The samples are observed by spatially distributed sensors, processed locally by each sensor, and communicated to the fusion center (FC) through *parallel* channels corrupted by *coherent* fading and additive white Gaussian noise. Examples of the parameters to be estimated include attributes associated with the underlying function such as its height, center, variances in different directions, or the weights of its components over a predefined basis set.

Each sensor processes its local noisy observations. In our analyses, two local processing schemes, namely *analog* and *digital*, will be considered. In the analog local processing scheme, each sensor acts as a pure relay and transmits an amplified version of its raw analog noisy observations to the FC. In the digital local processing method, each sensor quantizes its local noisy observations and sends the quantized samples to the FC using a digital modulation format. The FC combines all of the received locally processed observations and estimates the vector of unknown parameters. A major application of this work is in cases where the estimated parameters are then used in detection, classification, localization, and tracking

of the underlying object that has created the observed influence-field intensity function. The main contribution of this chapter is a generalized formulation of distributed parameter estimation in the context of wireless sensor networks (WSNs), where local observations are *not* (necessarily) linearly dependent on the underlying parameters to be estimated and no specific observation model has been considered in the analyses.

The rest of this chapter is organized as follows: In Section 3.2, we will summarize the state of the art on the problem of field estimation and non-linear parameter estimation in general. Section 3.3 describes the model of the distributed parallel fusion WSN that will be considered in our analysis and defines the precise problem that we are considering in this chapter. In Section 3.4, the maximum likelihood (ML) estimate of a vector of unknown parameters associated with a deterministic two-dimensional function is derived for the case of *analog* local processing scheme. Section 3.5 considers the same problem for the case of *digital* signal-processing method. As it will be shown in this section, it is *not* computationally feasible to find the ML estimate in this case. Therefore, a linearized version of the expectation maximization (EM) algorithm will be proposed in Section 3.6 to numerically find the ML estimate of the vector of unknown parameters in the case of digital local processing scheme. Section 3.7 presents the numerical results of our simulations to study the performance of the proposed distributed-estimation framework for a special two-dimensional Gaussian-shaped function, whose associated parameters to be estimated are its height and center. The effects of different parameters of the WSN on the performance of the proposed system will be studied in this section. Finally, we conclude our discussions and summarize the main achievements of this chapter in Section 3.8.

3.2 Related Works

Xiao et al. [18] have considered the *linear* coherent distributed mean-squared error (MSE) estimation of an unknown vector under stringent bandwidth and power constraints, where the local observation model, the compression function at local sensors, and the fusion rule at the FC are all linear. As a result of the bandwidth constraint, each sensor in their proposed framework transmits to the FC a fixed number of real-valued messages per observation.

The power constraint in their model limits the strength of the transmitted signals. Based on their proposed algorithm, each sensor linearly encodes its observations using a linear amplify-and-forward coding strategy. The FC then applies a linear mapping to estimate the unknown vector based on the received messages from local sensors through a coherent multiple-access channel, assuming a perfect synchronization between sensors and the FC so that the transmitted messages from local sensors are coherently combined at the FC. It is shown in [78] that if the sensor statistics are Gaussian (i.e., the parameter to be estimated and the independent observation noises are Gaussian) and the communication channels between local sensors and the FC are standard Gaussian multiple-access channels, a simple uncoded amplify-and-forward scheme in which each sensor's channel input is merely a scaled version of its noisy observation, drastically outperforms the separate source-channel coding approach in the sense of mean-squared error. Therefore, the proposed distributed-estimation algorithm in [18] performs optimally in applications that satisfy the requirements summarized in [78]. In this chapter, a general non-linear model for the distributed local observations is considered and analyzed.

Ishvar et al. [26] have considered the problem of estimation with unreliable communication links and have derived an information-theoretic achievable rate-distortion region characterizing the per-sample sensor bit-rate versus the estimation error. In particular, they have considered two specific cases in their analyses: (a) The case of fully distributed estimation schemes with no inter-sensor collaboration, and (b) the case of localized or collaborative estimation schemes in which it is assumed that the network is divided into a number of clusters, where collaboration is allowed among sensors within the same cluster but not across clusters. Nowak et al. [79] have studied the tradeoffs between the estimation error and energy consumption in a WSN as functions of sensor density. They have proposed practical in-network processing approaches based on hierarchical data handling and multi-scale partitioning methods for the estimation of two-dimensional inhomogeneous fields composed of two or more homogeneous, smoothly varying regions separated by smooth boundaries.

Ribeiro and Giannakis [24] have proposed a distributed bandwidth-constrained ML-estimation method for estimating a scaler deterministic signal in the presence of zero-mean additive white Gaussian observation noise using only quantized versions of the original local

observations, perfectly received at the FC. In a sequel work, they have considered more realistic signal models such as known univariate but generally non-Gaussian noise probability density functions (pdfs), known noise pdfs with a finite number of unknown parameters, completely unknown noise pdfs, and practical generalizations to vector signals and multivariate and possibly correlated noise pdfs [25]. The observation model in this work can in general be a non-linear function of the vector signal to be estimated, but it still ignores the effects of imperfect wireless channels between local sensors and the FC. It is shown in these works that transmitting a few bits (or even a single bit) per sensor can approach the performance of the estimator based on unquantized data under realistic conditions. In this chapter, we consider a general observation model in which local observations are *not* necessarily linear functions of the vector of parameters to be estimated. Furthermore, impaired fading channels between local sensors and the FC will be considered in the analysis.

Niu and Varshney [28] have proposed an intensity- (or energy-) based ML location-estimation scheme to estimate the coordinates of an energy-emitting source using quantized versions of the local noisy observations, which are assumed to be perfectly received by the FC. They have considered an isotropic intensity-attenuation model in which the energy of the signal is assumed to be inversely proportional to the nth exponent of the Euclidean distance from its source. An optimal (but infeasible) as well as two heuristic practical design methods have been proposed to find the local quantization thresholds by minimizing the summation of the estimator variances for the target's two coordinates. In a sequel work, Ozdemir et al. [29] have added the effects of fading and noisy wireless communication channels between local sensors and the FC to their problem formulation and have developed similar results. Maşazade et al. [80] have considered a similar problem in which the quantized version of sensor measurements are transmitted to the FC over error-free channels. They have proposed an iterative, energy-efficient source-localization scheme in which the algorithm begins with a coarse location estimate obtained from measurement data of a set of anchor sensors. Based on the accumulated information at each iteration, the posterior pdf of the source location is approximated using an importance-sampling-based Monte-Carlo method, which is then used to activate a number of non-anchor sensors selected to improve the accuracy of the source-location estimate the most. Distributed compression of mea-

surement data prior to transmission is also employed at the non-anchor sensors to further reduce the energy consumption. It is shown that this iterative scheme reduces the communication requirements by selecting only the most informative sensors and compressing their data prior to transmission to the FC. An iterative, non-linear, least-square received-signal-strength-based location-estimation technique is also proposed in [81] for the joint estimation of unknown location coordinates and distance-power gradient, which is a parameter of radio propagation path-loss model. Salman et al. [82] have proposed a low-complexity version of this approach. Although the illustrative case study for the numerical simulation results in this chapter considers estimating the location of the center of a generic Gaussian function, among its other parameters, the scope of our work is *not* limited to estimating only the location. In other words, we consider a more general framework in this chapter than the ones considered in the aforementioned works, and our analyses are not based on the signal propagation model in the observation environment.

3.3 System Model and Problem Statement

The system model of the WSN considered in this chapter, which is depicted in Figure 3.1, is the same as the one described in Section 1.1. It is assumed that the spatially distributed sensors forming the WSN are located within the domain of a two-dimensional function $g(x, y)$ that is completely known except for a set of unknown deterministic parameters denoted by $\boldsymbol{\theta} \stackrel{\text{def}}{=} [\theta_1, \theta_2, \ldots, \theta_p]^T$. The ultimate goal of the underlying WSN is to reliably estimate the vector of unknown parameters $\boldsymbol{\theta}$ using distributed noisy samples of function $g(x, y)$ provided by local sensors to a FC through parallel, coherent fading channels.

Assume that the ith sensor observes a noisy version of the sample of function $g(x, y)$ at its location. Therefore, the observation function $\Xi_i(\cdot)$ in Equation (1.1) is defined as

$$\Xi_i(\boldsymbol{\theta}) \stackrel{\text{def}}{=} g(x_i, y_i),$$

and the local noisy observation at the ith sensor is found as

$$r_i = g(x_i, y_i) + w_i, \qquad i = 1, 2, \ldots, K, \tag{3.1}$$

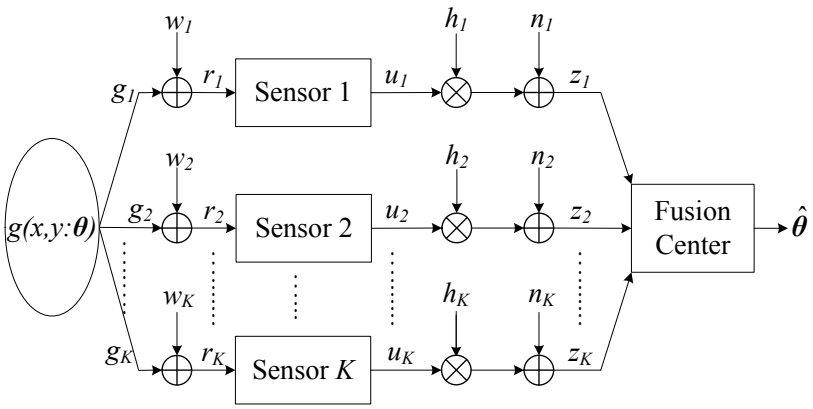


Figure 3.1: System model of a typical WSN for distributed estimation of a vector parameter $\boldsymbol{\theta}$.

where (x_i, y_i) is the location of the ith sensor in the network, and $g_i \stackrel{\text{def}}{=} g(x_i, y_i)$ is the sample of function $g(x, y)$ at location (x_i, y_i). Sensors can be placed on a uniform lattice or can be distributed at random over the observation area covered by the WSN. Throughout the discussions in this chapter, we assume that the locations of distributed sensors are known at the FC.

In this chapter, we consider both *analog* and *digital* local processing schemes described in Subsection 1.1.2. We will formulate detailed analysis of the local analog-processing scheme in Section 3.4 and that of the local digital-processing scheme in Sections 3.5 and 3.6. The model of the communication channels between local sensors and the FC is the same as the one introduced in Subsection 1.1.3. It is assumed that the channel fading coefficients are spatially uncorrelated and *completely known* at the FC.

The FC combines the vector of locally processed observations, communicated through orthogonal channels by distributed sensors (i.e., $\mathbf{z} \stackrel{\text{def}}{=} [z_1, z_2, \ldots, z_K]^T$) to estimate the vector of unknown deterministic parameters $\boldsymbol{\theta}$. In the following three sections, we will formulate the ML estimation of $\boldsymbol{\theta}$ at the FC for the two cases of analog and digital local processing schemes and the EM algorithm to numerically find the ML estimate of $\boldsymbol{\theta}$ at the FC for the digital local processing scheme.

3.4 ML Estimation of $\boldsymbol{\theta}$ with Analog Local Processing

Suppose that the local processing rule at each sensor is the same as the one introduced in Equation (1.2), where the local amplification gain α_i is assumed to be *known* at the FC. Note that in this work, we do not try to optimize the local amplification gains as considered by Banavar et al. [19] for the problem of distributed estimation of a scalar, random signal. The received signal from sensor i at the FC can then be represented as

$$z_i = h_i \alpha_i r_i + n_i = h_i \alpha_i g_i + h_i \alpha_i w_i + n_i, \qquad i = 1, 2, \ldots, K. \tag{3.2}$$

Assuming *complete knowledge* of the local amplification gain α_i and channel fading coefficient h_i, the FC makes a linear transformation on the received signal from each sensor to find

$$\begin{aligned} z'_i &= (h_i \alpha_i)^{-1} z_i \\ &= g_i + w_i + (h_i \alpha_i)^{-1} n_i \qquad i = 1, 2, \ldots, K, \\ &= g_i + v_i \end{aligned} \tag{3.3}$$

where the cumulative noise is defined as $v_i \stackrel{\text{def}}{=} w_i + \frac{1}{h_i \alpha_i} n_i$, which is a zero-mean white Gaussian random variable with variance $\xi_i^2 \stackrel{\text{def}}{=} \sigma_i^2 + \frac{\tau_i^2}{|h_i \alpha_i|^2}$. The estimation process at the FC can then be performed based on the vector of transformed variables $\mathbf{z}' \stackrel{\text{def}}{=} [z'_1, z'_2, \ldots, z'_K]^T$.

The probability density function of the linearly processed received vector of local observations from distributed sensors at the FC is given by

$$f_{\mathbf{z}'}(\mathbf{z}' : \boldsymbol{\theta}) = \frac{1}{\sqrt{(2\pi)^K \det \boldsymbol{\Sigma}}} \exp\left(-\frac{1}{2} (\mathbf{z}' - \mathbf{g})^T \boldsymbol{\Sigma}^{-1} (\mathbf{z}' - \mathbf{g}) \right), \tag{3.4}$$

where $\boldsymbol{\Sigma} \stackrel{\text{def}}{=} \operatorname{diag}(\xi_1^2, \xi_2^2, \ldots, \xi_K^2)$ is the diagonal matrix of cumulative noise variances and $\mathbf{g} \stackrel{\text{def}}{=} [g_1, g_2, \ldots, g_K]^T$ is the vector of the samples of function $g(x, y)$ at sensor locations. The joint log-likelihood function of the vector of unknown parameters $\boldsymbol{\theta}$ is then found as

$$\begin{aligned} l(\boldsymbol{\theta}) &\stackrel{\text{def}}{=} \ln f_{\mathbf{z}'}(\mathbf{z}' : \boldsymbol{\theta}) \\ &\equiv -\left[\ln(\det \boldsymbol{\Sigma}) + (\mathbf{z}' - \mathbf{g})^T \boldsymbol{\Sigma}^{-1} (\mathbf{z}' - \mathbf{g}) \right] \\ &\stackrel{(a)}{\equiv} -(\mathbf{z}' - \mathbf{g})^T \boldsymbol{\Sigma}^{-1} (\mathbf{z}' - \mathbf{g}) \\ &= -\sum_{i=1}^{K} \frac{1}{\xi_i^2} (z'_i - g_i)^2, \end{aligned} \tag{3.5}$$

where (a) is based on the fact that $\det \boldsymbol{\Sigma}$ is independent of $\boldsymbol{\theta}$.

The ML estimate of the vector of unknown parameters $\boldsymbol{\theta}$ is the maximizer of its log-likelihood function given in Equation (3.5). In other words,

$$\begin{aligned} \widehat{\boldsymbol{\theta}}_{\mathsf{ML}} &= \arg \max_{\boldsymbol{\theta}} \, l\left(\boldsymbol{\theta}\right) \\ &= \arg \min_{\boldsymbol{\theta}} \, \left(\mathbf{z}' - \mathbf{g}\right)^T \boldsymbol{\Sigma}^{-1} \left(\mathbf{z}' - \mathbf{g}\right) \\ &= \arg \min_{\boldsymbol{\theta}} \, \sum_{i=1}^{K} \frac{1}{\xi_i^2} \left(z_i' - g_i\right)^2. \end{aligned} \tag{3.6}$$

Based on Equation (3.6), the ML estimate of the vector of unknown parameters $\widehat{\boldsymbol{\theta}}_{\mathsf{ML}}$ is found as the solution of the following system of equations:

$$\boldsymbol{\nabla}_{\boldsymbol{\theta}} \, l\left(\boldsymbol{\theta}\right) \Big|_{\widehat{\boldsymbol{\theta}}_{\mathsf{ML}}} = \mathbf{0},$$

where $\boldsymbol{\nabla}_{\boldsymbol{\theta}} \stackrel{\text{def}}{=} \left[\frac{\partial}{\partial \theta_1}, \frac{\partial}{\partial \theta_2}, \ldots, \frac{\partial}{\partial \theta_p}\right]^T$ is the gradient operator with respect to parameter $\boldsymbol{\theta}$. This system of equations can be simplified in the vector form by substituting $l\left(\boldsymbol{\theta}\right)$ from Equation (3.5) as

$$\frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}} \, \boldsymbol{\Sigma}^{-1} \, \left(\mathbf{z}' - \mathbf{g}\right) = \mathbf{0}, \tag{3.7}$$

where $\frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}}$ is a p-by-K matrix of partial derivatives of the components of vector $\mathbf{g}$ with respect to the components of parameter $\boldsymbol{\theta}$ defined as

$$\frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}} \stackrel{\text{def}}{=} \begin{bmatrix} \frac{\partial g_1}{\partial \theta_1} & \frac{\partial g_2}{\partial \theta_1} & \cdots & \frac{\partial g_K}{\partial \theta_1} \\ \frac{\partial g_1}{\partial \theta_2} & \frac{\partial g_2}{\partial \theta_2} & \cdots & \frac{\partial g_K}{\partial \theta_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_1}{\partial \theta_p} & \frac{\partial g_2}{\partial \theta_p} & \cdots & \frac{\partial g_K}{\partial \theta_p} \end{bmatrix}.$$

The system of equations derived in (3.7) can be rewritten in the scalar format as

$$\sum_{i=1}^{K} \left[\frac{1}{\xi_i^2} \, \left(\frac{\partial g_i}{\partial \theta_j}\right) \, \left(z_i' - g_i\right)\right] = 0, \qquad\qquad j = 1, 2, \ldots, p. \tag{3.8}$$

This system of equations is highly non-linear with respect to the unknown parameters for most practical applications and does not have a closed-form solution. In Section 3.7, this system will be linearized and solved using Newton's method to estimate a vector of unknown parameters associated with a specific two-dimensional Gaussian-shaped function of interest.

3.5 ML Estimation of $\boldsymbol{\theta}$ with Digital Local Processing

Suppose that the local processing rule at each sensor is the same as the one introduced in Subsection 1.1.2 under "digital local processing" and Equation (1.3), where each sensor sends the index of its quantized multi-bit sample to the FC. Based on the local observation model introduced in Equation (3.1), the probability density function of each sensor's quantized output sample is found as

$$f_{U_i}\left(u_i : \boldsymbol{\theta}\right) = \sum_{\ell=0}^{M_i-1} \Delta Q_i\left(\ell\right) \delta\left[u_i - \ell\right], \quad u_i = 0, 1, \ldots, M_i - 1 \text{ and } i = 1, 2, \ldots, K, \tag{3.9}$$

where $\delta\left[\cdot\right]$ denotes the discrete Dirac delta function, $\Delta Q_i\left(\ell\right)$ is defined as

$$\Delta Q_i\left(\ell\right) \stackrel{\text{def}}{=} Q\left(\frac{\beta_i(\ell) - g_i}{\sigma_i}\right) - Q\left(\frac{\beta_i(\ell+1) - g_i}{\sigma_i}\right), \quad \ell = 0, 1, \ldots, M_i - 1 \text{ and } i = 1, 2, \ldots, K, \tag{3.10}$$

and $Q(\cdot)$ is the complementary distribution function of the standard Gaussian random variable defined as

$$Q(x) \stackrel{\text{def}}{=} \frac{1}{\sqrt{2\pi}} \int_x^{\infty} \exp\left(-\frac{1}{2}t^2\right) \, \mathrm{d}t.$$

Similar to the case of analog local processing, assuming *complete knowledge* of the channel fading coefficient h_i, the FC makes a linear transformation on the received signal from each sensor to find

$$\begin{aligned} z_i' &= {h_i}^{-1} \, z_i \\ &= u_i + \frac{1}{h_i} n_i \qquad\qquad i = 1, 2, \ldots, K, \\ &= u_i + v_i \end{aligned} \tag{3.11}$$

where the transformed noise is defined as $v_i \stackrel{\text{def}}{=} \frac{1}{h_i} n_i$, which is a zero-mean Gaussian random variable with variance $\xi_i^2 \stackrel{\text{def}}{=} \frac{\tau_i^2}{|h_i|^2}$. The estimation process at the FC can then be performed based on the vector of transformed variables $\mathbf{z}' \stackrel{\text{def}}{=} \left[z_1', z_2', \ldots, z_K'\right]^T$.

The probability density function of the linearly processed received vector of local obser-

vations from distributed sensors at the FC is given by

$$\begin{aligned} f_{\mathbf{z}'}\left(\mathbf{z}':\boldsymbol{\theta}\right) &\overset{(a)}{=} \prod_{i=1}^{K} f_{Z_i'}\left(z_i':\boldsymbol{\theta}\right) \\ &\overset{(b)}{=} \prod_{i=1}^{K}\sum_{u_i=0}^{M_i-1} f_{Z_i'|U_i}\left(z_i'|u_i\right) f_{U_i}\left(u_i:\boldsymbol{\theta}\right) \\ &\overset{(c)}{=} \prod_{i=1}^{K}\sum_{u_i=0}^{M_i-1}\left[\frac{1}{\sqrt{2\pi\xi_i^2}}\exp\left\{-\frac{\left(z_i'-u_i\right)^2}{2\xi_i^2}\right\}\right]\left[\sum_{\ell=0}^{M_i-1}\Delta Q_i\left(\ell\right)\delta\left[u_i-\ell\right]\right] \\ &= \prod_{i=1}^{K}\sum_{\ell=0}^{M_i-1}\sum_{u_i=0}^{M_i-1}\frac{1}{\sqrt{2\pi\xi_i^2}}\exp\left[-\frac{\left(z_i'-u_i\right)^2}{2\xi_i^2}\right]\Delta Q_i\left(\ell\right)\delta\left[u_i-\ell\right] \\ &\overset{(d)}{=} \prod_{i=1}^{K}\frac{1}{\sqrt{2\pi\xi_i^2}}\sum_{\ell=0}^{M_i-1}\Delta Q_i\left(\ell\right)\exp\left[-\frac{\left(z_i'-\ell\right)^2}{2\xi_i^2}\right], \end{aligned} \tag{3.12}$$

where (a) is based on the fact that the received local observations from distributed sensors are spatially uncorrelated, the summation in (b) is based on the theorem of total probability and over all M_i possible realizations of u_i, (c) is found by replacing the probability density function $f_{U_i}\left(u_i:\boldsymbol{\theta}\right)$ from Equation (3.9), and (d) is based on the fact that the inner summation over u_i can be simplified using the properties of the discrete Dirac delta function as

$$\sum_{u_i=0}^{M_i-1}\exp\left[-\frac{\left(z_i'-u_i\right)^2}{2\xi_i^2}\right]\Delta Q_i\left(\ell\right)\delta\left[u_i-\ell\right] = \exp\left[-\frac{\left(z_i'-\ell\right)^2}{2\xi_i^2}\right]\Delta Q_i\left(\ell\right).$$

Therefore, the joint log-likelihood function of the vector of unknown parameters $\boldsymbol{\theta}$ is found as

$$\begin{aligned} l\left(\boldsymbol{\theta}\right) &\overset{\text{def}}{=} \ln f_{\mathbf{z}'}\left(\mathbf{z}':\boldsymbol{\theta}\right) \\ &\equiv \sum_{i=1}^{K}\ln\left(\sum_{\ell=0}^{M_i-1}\Delta Q_i\left(\ell\right)\exp\left[-\frac{\left(z_i'-\ell\right)^2}{2\xi_i^2}\right]\right). \end{aligned} \tag{3.13}$$

The ML estimate of the vector of unknown parameters $\boldsymbol{\theta}$ is the maximizer of its log-likelihood function given in Equation (3.13). In other words,

$$\widehat{\boldsymbol{\theta}}_{\mathsf{ML}} = \arg\max_{\boldsymbol{\theta}} l\left(\boldsymbol{\theta}\right). \tag{3.14}$$

Therefore, $\widehat{\boldsymbol{\theta}}_{\mathsf{ML}}$ is found as the solution of the following system of equations:

$$\boldsymbol{\nabla}_{\boldsymbol{\theta}}\, l\left(\boldsymbol{\theta}\right)\Big|_{\widehat{\boldsymbol{\theta}}_{\mathsf{ML}}} = \mathbf{0},$$

which can be simplified by the substitution of $l\left(\boldsymbol{\theta}\right)$ from Equation (3.13) as

$$\sum_{i=1}^{K} \frac{\sum\limits_{\ell=0}^{M_i-1} A_{i,j}(\ell) \exp\left[-\frac{\left(z_i'-\ell\right)^2}{2\xi_i^2}\right]}{\sum\limits_{\ell=0}^{M_i-1} \Delta Q_i\left(\ell\right) \exp\left[-\frac{\left(z_i'-\ell\right)^2}{2\xi_i^2}\right]} = 0, \qquad j = 1, 2, \ldots, p, \tag{3.15}$$

where $A_{i,j}(\ell)$ is the partial derivative of $\Delta Q_i\left(\ell\right)$ with respect to θ_j defined as

$$\begin{aligned} A_{i,j}(\ell) &\stackrel{\text{def}}{=} \frac{\partial}{\partial \theta_j}\left[\Delta Q_i\left(\ell\right)\right] \\ &= \frac{1}{\sqrt{2\pi\sigma_i^2}}\left(\frac{\partial g_i}{\partial \theta_j}\right) \exp\left(-\frac{g_i^2}{2\sigma_i^2}\right) B_i(\ell), \end{aligned}$$

where $B_i(\ell)$ is defined as

$$B_i(\ell) \stackrel{\text{def}}{=} \exp\left(\frac{2g_i\beta_i\left(\ell\right) - \beta_i^2\left(\ell\right)}{2\sigma_i^2}\right) - \exp\left(\frac{2g_i\beta_i\left(\ell+1\right) - \beta_i^2\left(\ell+1\right)}{2\sigma_i^2}\right).$$

This system of equations given in (3.15) is highly non-linear with respect to the unknown parameters for most practical applications and does not have a closed-form solution. In practice, an efficient numerical approach has to be devised to find the ML estimate formulated in this section. In the next section, the EM algorithm [83] will be developed as an efficient iterative approach to numerically find the ML estimate of a vector of unknown parameters for the case of digital local processing scheme.

3.6 Linearized EM Solution for Digital Local Processing

Based on the EM algorithm introduced by Dempster et al. [83], we consider the following *complete* and *incomplete* datasets:

Incomplete dataset: The linearly processed received vector of local observations from distributed sensors at the FC $\mathbf{z}'$, defined in Equation (3.11).

Complete dataset: The pair of the vector of local observations $\mathbf{r} \stackrel{\text{def}}{=} [r_1, r_2, \ldots, r_K]^T$, defined in Equation (3.1), and the vector of linearly transformed additive white Gaussian *channel* noise variables $\mathbf{v} \stackrel{\text{def}}{=} [v_1, v_2, \ldots, v_K]^T$, i.e., $\{\mathbf{r}, \mathbf{v}\}$.

The *mapping* $\mathbf{z}' = \mathbf{u} + \mathbf{v}$ relates the incomplete and complete data spaces, where $\mathbf{u} \stackrel{\text{def}}{=} [u_1, u_2, \ldots, u_K]^T$ is the vector of locally quantized observations of distributed sensors based on $u_i = \gamma_i(r_i)$ as the known local quantization rule of sensor i defined in Equation (1.3).

The joint probability density function of the complete dataset, parametrized by the vector of unknown parameters $\boldsymbol{\theta}$ is found as

$$\begin{aligned} f_{\mathsf{CD}}(\mathbf{r}, \mathbf{v} : \boldsymbol{\theta}) &\stackrel{(a)}{=} f_{\mathbf{r}}(\mathbf{r} : \boldsymbol{\theta})\, f_{\mathbf{v}}(\mathbf{v}) \\ &= \left[\frac{1}{\sqrt{(2\pi)^K \det \boldsymbol{\Sigma}_{\mathsf{O}}}} \exp\left(-\frac{1}{2}(\mathbf{r}-\mathbf{g})^T \boldsymbol{\Sigma}_{\mathsf{O}}^{-1}(\mathbf{r}-\mathbf{g})\right)\right] \\ &\quad \times \left[\frac{1}{\sqrt{(2\pi)^K \det \boldsymbol{\Sigma}_{\mathsf{C}}}} \exp\left(-\frac{1}{2}\mathbf{v}^T \boldsymbol{\Sigma}_{\mathsf{C}}^{-1}\mathbf{v}\right)\right], \end{aligned} \tag{3.16}$$

where (a) is based on the fact that $\mathbf{r}$ and $\mathbf{v}$ are independent Gaussian random vectors, $\boldsymbol{\Sigma}_{\mathsf{O}} \stackrel{\text{def}}{=} \text{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_K^2)$ and $\boldsymbol{\Sigma}_{\mathsf{C}} \stackrel{\text{def}}{=} \text{diag}(\xi_1^2, \xi_2^2, \ldots, \xi_K^2)$ are the diagonal matrices of the variances of the observation noise and the transformed channel noise $\xi_i^2 \stackrel{\text{def}}{=} \frac{\tau_i^2}{|h_i|^2}$, respectively. The joint log-likelihood function of the complete dataset is defined as

$$\begin{aligned} l_{\mathsf{CD}}(\mathbf{r}, \mathbf{v} : \boldsymbol{\theta}) &\stackrel{\text{def}}{=} \ln f_{\mathsf{CD}}(\mathbf{r}, \mathbf{v} : \boldsymbol{\theta}) \\ &\equiv -(\mathbf{r}-\mathbf{g})^T \boldsymbol{\Sigma}_{\mathsf{O}}^{-1}(\mathbf{r}-\mathbf{g}) \\ &= -\sum_{i=1}^{K} \frac{1}{\sigma_i^2}(r_i - g_i)^2, \end{aligned} \tag{3.17}$$

where the terms independent of the vector of unknown parameters $\boldsymbol{\theta}$ are omitted.

Let $\widehat{\boldsymbol{\theta}}^{(k)}$ be the estimate of the unknown vector of parameters at the kth iteration of the EM algorithm. To further refine and update the estimates of the unknown parameters, we alternate the *expectation* and *maximization* steps defined as follows:

Expectation Step (E-Step): During the expectation step, the conditional expectation of the joint log-likelihood function of the complete dataset, given the incomplete dataset and

$\widehat{\boldsymbol{\theta}}^{(k)}$, is found as

$$\begin{aligned} F\left(\boldsymbol{\theta} \;\middle|\; \widehat{\boldsymbol{\theta}}^{(k)}\right) &\stackrel{\text{def}}{=} \mathbb{E}\left[l_{\mathsf{CD}}\left(\mathbf{r},\mathbf{v}:\boldsymbol{\theta}\right)\middle|\mathbf{z}',\widehat{\boldsymbol{\theta}}^{(k)}\right] \\ &= \mathbb{E}\left[-\sum_{i=1}^{K}\frac{1}{\sigma_i^2}\left(r_i-g_i\right)^2\middle|\mathbf{z}',\widehat{\boldsymbol{\theta}}^{(k)}\right], \end{aligned} \tag{3.18}$$

where $\mathbb{E}\left[\cdot\right]$ denotes the expectation operation with respect to the conditional pdf of the complete dataset, given the incomplete dataset and the estimate of the vector of parameters at the kth iteration as

$$f_{\mathsf{CD}|\mathsf{ID}}\left(\mathbf{r},\mathbf{v}:\widehat{\boldsymbol{\theta}}^{(k)}\middle|\mathbf{z}'\right) = f_{\mathbf{r}|\mathbf{z}'}\left(\mathbf{r}:\widehat{\boldsymbol{\theta}}^{(k)}\middle|\mathbf{z}'\right) f_{\mathbf{v}|\mathbf{z}'}\left(\mathbf{v}:\widehat{\boldsymbol{\theta}}^{(k)}\middle|\mathbf{z}'\right), \tag{3.19}$$

where $\mathbf{r}$ and $\mathbf{v}$ are two independent Gaussian random vectors, $f_{\mathbf{v}|\mathbf{z}'}\left(\mathbf{v}:\widehat{\boldsymbol{\theta}}^{(k)}\middle|\mathbf{z}'\right)$ is independent of the argument inside the expectation operation, and based on the Bayes' rule,

$$f_{\mathbf{r}|\mathbf{z}'}\left(\mathbf{r}:\widehat{\boldsymbol{\theta}}^{(k)}\middle|\mathbf{z}'\right) = \frac{f_{\mathbf{z}'|\mathbf{r}}\left(\mathbf{z}':\widehat{\boldsymbol{\theta}}^{(k)}\middle|\mathbf{r}\right) f_{\mathbf{r}}\left(\mathbf{r}:\widehat{\boldsymbol{\theta}}^{(k)}\right)}{f_{\mathbf{z}'}\left(\mathbf{z}':\widehat{\boldsymbol{\theta}}^{(k)}\right)}. \tag{3.20}$$

Based on the theorem of total probability,

$$f_{\mathbf{z}'}\left(\mathbf{z}':\widehat{\boldsymbol{\theta}}^{(k)}\right) = \underbrace{\int\cdots\int}_{K} f_{\mathbf{z}'|\mathbf{r}}\left(\mathbf{z}':\widehat{\boldsymbol{\theta}}^{(k)}\middle|\mathbf{r}\right) f_{\mathbf{r}}\left(\mathbf{r}:\widehat{\boldsymbol{\theta}}^{(k)}\right)\,\mathrm{d}\mathbf{r}. \tag{3.21}$$

Note that $f_{\mathbf{z}'}\left(\mathbf{z}':\widehat{\boldsymbol{\theta}}^{(k)}\right)$ is independent of the argument inside the expectation operation in Equation (3.18).

Maximization Step (M-Step): During the maximization step, the next estimate of the vector of unknown parameters is found as the maximizer of the result of the expectation step. In other words,

$$\widehat{\boldsymbol{\theta}}^{(k+1)} = \arg\max_{\boldsymbol{\theta}} F\left(\boldsymbol{\theta} \;\middle|\; \widehat{\boldsymbol{\theta}}^{(k)}\right), \tag{3.22}$$

which can be rewritten as

$$\frac{\partial}{\partial\theta_j}\left\{F\left(\boldsymbol{\theta} \;\middle|\; \widehat{\boldsymbol{\theta}}^{(k)}\right)\right\}\Big|_{\widehat{\boldsymbol{\theta}}^{(k+1)}} = 0, \qquad j=1,2,\ldots,p. \tag{3.23}$$

This system of equations can be specified more precisely by the substitution of $F\left(\boldsymbol{\theta} \;\middle|\; \widehat{\boldsymbol{\theta}}^{(k)}\right)$ from Equation (3.18) as

$$\mathbb{E}\left[\sum_{i=1}^{K}\frac{1}{\sigma_i^2}\left(r_i-g_i\right)\frac{\partial g_i}{\partial\theta_j}\middle|\mathbf{z}',\widehat{\boldsymbol{\theta}}^{(k)}\right]\Bigg|_{\widehat{\boldsymbol{\theta}}^{(k+1)}} = 0, \qquad j=1,2,\ldots,p. \tag{3.24}$$

The conditional expectation in Equation (3.24) can be found using the probability density functions defined in Equations (3.19)–(3.21) as

$$\begin{aligned}
\mathbb{E}\left[\sum_{i=1}^{K}\frac{1}{\sigma_i^2}\left(r_i-g_i\right)\frac{\partial g_i}{\partial\theta_j}\middle|\mathbf{z}',\widehat{\boldsymbol{\theta}}^{(k)}\right] &= \sum_{i=1}^{K}\frac{1}{\sigma_i^2}\mathbb{E}\left[\left(r_i-g_i\right)\frac{\partial g_i}{\partial\theta_j}\middle|z_i',\widehat{\boldsymbol{\theta}}^{(k)}\right]\\
&= \sum_{i=1}^{K}\frac{1}{\sigma_i^2}\iint\left[\left(r_i-g_i\right)\frac{\partial g_i}{\partial\theta_j}\right]f_{\mathsf{CD}|\mathsf{ID}}\left(r_i,v_i:\widehat{\boldsymbol{\theta}}^{(k)}\middle|z_i'\right)\,\mathrm{d}r_i\mathrm{d}v_i\\
&\overset{(a)}{=} \sum_{i=1}^{K}\frac{1}{\sigma_i^2}\int\left[\left(r_i-g_i\right)\frac{\partial g_i}{\partial\theta_j}\right]f_{R_i|Z_i'}\left(r_i:\widehat{\boldsymbol{\theta}}^{(k)}\middle|z_i'\right)\,\mathrm{d}r_i\\
&= \sum_{i=1}^{K}\frac{1}{\sigma_i^2}\left[\frac{1}{f_{Z_i'}\left(z_i':\widehat{\boldsymbol{\theta}}^{(k)}\right)}\right]T_{i,j}^{(k)}(z_i'),
\end{aligned}\tag{3.25}$$

where (a) is based on Equation (3.19) and the independence of v_i from the argument of the expectation operation, and $T_{i,j}^{(k)}(z_i')$ is defined as

$$T_{i,j}^{(k)}(z_i')=\int_{-\infty}^{\infty}\left[\left(r_i-g_i\right)\frac{\partial g_i}{\partial\theta_j}\right]f_{Z_i'|R_i}\left(z_i':\widehat{\boldsymbol{\theta}}^{(k)}\middle|r_i\right)f_{R_i}\left(r_i:\widehat{\boldsymbol{\theta}}^{(k)}\right)\,\mathrm{d}r_i.\tag{3.26}$$

Note that based on Equation (3.11), random variable Z_i', given R_i is Gaussian with mean u_i and variance $\xi_i^2\overset{\text{def}}{=}\frac{\tau_i^2}{|h_i|^2}$, i.e., $Z_i'|R_i\sim\mathcal{N}\left(u_i,\xi_i^2\right)$. Moreover, based on Equation (3.1), the random variable R_i is Gaussian with mean $g_i^{(k)}$ and variance σ_i^2, i.e., $R_i\sim\mathcal{N}\left(g_i^{(k)},\sigma_i^2\right)$, where $g_i^{(k)}=g\left(x_i,y_i:\widehat{\boldsymbol{\theta}}^{(k)}\right)$ is the estimate of the underlying function $g(x,y)$ at location (x_i,y_i) with the vector of unknown parameters $\boldsymbol{\theta}$ replaced by its estimate at the kth iteration of the EM algorithm. Using these two probability density functions, $T_{i,j}^{(k)}(z_i')$ is found as

$$\begin{aligned}
T_{i,j}^{(k)}(z_i') &= \int_{-\infty}^{\infty}\left[\left(r_i-g_i\right)\frac{\partial g_i}{\partial\theta_j}\right]\left[\frac{1}{\sqrt{2\pi\xi_i^2}}\exp\left(-\frac{\left(z_i'-u_i\right)^2}{2\xi_i^2}\right)\right]\left[\frac{1}{\sqrt{2\pi\sigma_i^2}}\exp\left(-\frac{\left(r_i-g_i^{(k)}\right)^2}{2\sigma_i^2}\right)\right]\mathrm{d}r_i\\
&= \frac{1}{2\pi\sqrt{\xi_i^2\ \sigma_i^2}}\sum_{\ell=0}^{M_i-1}\int_{\beta_i(\ell)}^{\beta_i(\ell+1)}\left[\left(r_i-g_i\right)\frac{\partial g_i}{\partial\theta_j}\right]\exp\left(-\frac{\left(z_i'-\ell\right)^2}{2\xi_i^2}\right)\exp\left(-\frac{\left(r_i-g_i^{(k)}\right)^2}{2\sigma_i^2}\right)\mathrm{d}r_i\\
&= \frac{1}{\sqrt{2\pi\xi_i^2}}\sum_{\ell=0}^{M_i-1}\exp\left(-\frac{\left(z_i'-\ell\right)^2}{2\xi_i^2}\right)\left(\frac{\partial g_i}{\partial\theta_j}\right)\Gamma_i^{(k)}\left(\ell\right),
\end{aligned}\tag{3.27}$$

where $\Gamma_i^{(k)}(\ell)$ is defined as

$$\begin{aligned}\Gamma_i^{(k)}(\ell) &= \frac{1}{\sqrt{2\pi\sigma_i^2}}\int_{\beta_i(\ell)}^{\beta_i(\ell+1)}(r_i-g_i)\exp\left[-\frac{\left(r_i-g_i^{(k)}\right)^2}{2\sigma_i^2}\right]\mathrm{d}r_i \\ &= \sqrt{\frac{\sigma_i^2}{2\pi}}\int_{\beta_i(\ell)}^{\beta_i(\ell+1)}\frac{1}{\sigma_i^2}\left(r_i-g_i^{(k)}\right)\exp\left[-\frac{\left(r_i-g_i^{(k)}\right)^2}{2\sigma_i^2}\right]\mathrm{d}r_i \\ &\qquad + \frac{1}{\sqrt{2\pi\sigma_i^2}}\int_{\beta_i(\ell)}^{\beta_i(\ell+1)}\left(g_i^{(k)}-g_i\right)\exp\left[-\frac{\left(r_i-g_i^{(k)}\right)^2}{2\sigma_i^2}\right]\mathrm{d}r_i \\ &= \sqrt{\frac{\sigma_i^2}{2\pi}}\exp\left[-\frac{\left(g_i^{(k)}\right)^2}{2\sigma_i^2}\right]\Lambda_i^{(k)}(\ell)+\left(g_i^{(k)}-g_i\right)\Delta Q_i^{(k)}(\ell),\end{aligned} \tag{3.28}$$

where $\Delta Q_i^{(k)}(\ell) = \Delta Q_i(\ell)\Big|_{g_i=g_i^{(k)}}$, $\Delta Q_i(\ell)$ is defined in Equation (3.10), and $\Lambda_i^{(k)}(\ell)$ is defined as

$$\Lambda_i^{(k)}(\ell) = \exp\left(\frac{2g_i^{(k)}\beta_i(\ell)-\beta_i^2(\ell)}{2\sigma_i^2}\right)-\exp\left(\frac{2g_i^{(k)}\beta_i(\ell+1)-\beta_i^2(\ell+1)}{2\sigma_i^2}\right). \tag{3.29}$$

Combing Equations (3.25) and (3.27)–(3.29) and replacing them in Equation (3.24) will result in a non-linear system of equations in terms of $\widehat{\boldsymbol{\theta}}^{(k+1)}$. In other words, Equation (3.24) can be rewritten as follows to give a new update of the vector of unknown parameters at the $(k+1)$th step of the EM algorithm using its values derived in the kth step:

$$\begin{aligned}&\sum_{i=1}^{K}\frac{1}{\sigma_i^2\sqrt{2\pi\xi_i^2}}\left[\frac{1}{f_{Z_i'}\left(z_i':\widehat{\boldsymbol{\theta}}^{(k)}\right)}\right]\sum_{\ell=0}^{M_i-1}\exp\left(-\frac{(z_i'-\ell)^2}{2\xi_i^2}\right) \\ &\qquad\times\left[\frac{\partial g_i^{(k+1)}}{\partial\theta_j}M_i^{(k)}(\ell)-g_i^{(k+1)}\frac{\partial g_i^{(k+1)}}{\partial\theta_j}\Delta Q_i^{(k)}(\ell)\right]=0,\quad j=1,2,\ldots,p,\end{aligned} \tag{3.30}$$

where $M_i^{(k)}(\ell)$ is defined as

$$M_i^{(k)}(\ell) = \sqrt{\frac{\sigma_i^2}{2\pi}}\exp\left[-\frac{\left(g_i^{(k)}\right)^2}{2\sigma_i^2}\right]\Lambda_i^{(k)}(\ell)+g_i^{(k)}\Delta Q_i^{(k)}(\ell). \tag{3.31}$$

It can easily been seen that the system of equations given in (3.30) is highly non-linear with respect to the components of the vector of unknown parameters to be estimated $\widehat{\boldsymbol{\theta}}^{(k+1)}$. Furthermore, the solution set for this system of equations is generally *non-convex*. Therefore, a numerical solution should be found for this system of equations at each iteration of the EM algorithm. In Section 3.7, this approach will be utilized to numerically find the ML estimate of a vector of unknown parameters associated with a specific two-dimensional Gaussian-shaped function of interest. Newton's method is applied to linearize the system of equations given in (3.30) as briefly described in the following.

Let $\mathbf{f}\left(\boldsymbol{\theta}^{(k+1)}\right) = \mathbf{0}$ be the vector form of the system of equations given in (3.30), where each component of the vector function $\mathbf{f}\left(\cdot\right)$ is a non-linear function of the components of the vector of unknown parameters $\boldsymbol{\theta}^{(k+1)} \stackrel{\text{def}}{=} \left[\theta_1^{(k+1)}, \theta_2^{(k+1)}, \ldots, \theta_p^{(k+1)}\right]^T$. At each step m of the Newton's linearization method, the following system of *linear* equations is solved to find a new update for $\boldsymbol{\theta}_{m+1}^{(k+1)}$ based on the previous estimate of $\boldsymbol{\theta}_m^{(k+1)}$:

$$\mathbf{J}\left(\boldsymbol{\theta}_m^{(k+1)}\right)\left[\boldsymbol{\theta}_{m+1}^{(k+1)} - \boldsymbol{\theta}_m^{(k+1)}\right] = -\mathbf{f}\left(\boldsymbol{\theta}_m^{(k+1)}\right), \qquad m = 0, 1, 2, \ldots, \tag{3.32}$$

where $\mathbf{J}\left(\boldsymbol{\theta}_m^{(k+1)}\right)$ is the p-by-p Jacobian matrix of the vector function $\mathbf{f}\left(\cdot\right)$ evaluated at $\boldsymbol{\theta}_m^{(k+1)}$. The element at the rth row and cth column of the Jacobian matrix $\mathbf{J}\left(\cdot\right)$ evaluated at $\boldsymbol{\theta}_m^{(k+1)}$ is defined as

$$J_{r,c}\left(\boldsymbol{\theta}_m^{(k+1)}\right) = \left.\frac{\partial f_r\left(\boldsymbol{\theta}\right)}{\partial \theta_c}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_m^{(k+1)}}, \qquad r, c = 1, 2, \ldots, p, \tag{3.33}$$

where $f_r(\cdot)$ is the rth component of the vector function $\mathbf{f}\left(\cdot\right)$.

3.7 Case Study and Numerical Analysis

In this section, we will numerically evaluate the performance of the ML-estimation framework developed in the previous sections for both analog and digital local processing schemes for a specific function of interest $g(x, y)$.

3.7.1 Simulation Setup, Parameter Specification, and Performance-Measure Definition

Suppose that a two-dimensional Gaussian-shaped function defined as

$$g(x,y) \stackrel{\text{def}}{=} h \exp\left[-\frac{1}{2}\left(\frac{(x-x_{\mathsf{c}})^2}{\sigma_x^2}+\frac{(y-y_{\mathsf{c}})^2}{\sigma_y^2}\right)\right] \tag{3.34}$$

is being sampled by the WSN under study, where h is the maximum intensity or height of the function, $(x_{\mathsf{c}}, y_{\mathsf{c}})$ is the location of its center, and σ_x^2 and σ_y^2 are the *known* spreads of the function in the x and y directions, respectively. In our simulations, we fix these values to be $\sigma_x^2 = 4$ and $\sigma_y^2 = 1$. Let $\boldsymbol{\theta} \stackrel{\text{def}}{=} [h, x_{\mathsf{c}}, y_{\mathsf{c}}]^T$ be the vector of unknown deterministic parameters associated with function $g(x,y)$. The goal is to estimate these $p = 3$ parameters using distributed (and possibly sparse) noisy samples of the function acquired by a WSN and transmitted to a FC through parallel, coherent fading channels.

Suppose that K sensors are *randomly* distributed in the observation environment, which is assumed to at least cover the area $\mathcal{A} \stackrel{\text{def}}{=} [x_{\mathsf{c}} - 3\sigma_x, x_{\mathsf{c}} + 3\sigma_x] \times [y_{\mathsf{c}} - 3\sigma_y, y_{\mathsf{c}} + 3\sigma_y]$, where $\mathcal{X} \times \mathcal{Y}$ denotes the Cartesian product between two sets $\mathcal{X}$ and $\mathcal{Y}$. This choice of the observation environment guarantees that all of the sensors are within the domain of the underlying function $g(x,y)$, and that almost all of the domain of this function is covered by distributed sensors. In our simulations, we assume that the observations made by distributed sensors are *homogeneous*. In other words, it is assumed that the additive white Gaussian observation noises for all sensors are *identically distributed* and their variances are the same, i.e.,

$$\sigma^2 \stackrel{\text{def}}{=} \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_K^2.$$

We define the *observation signal-to-noise ratio* (SNR) as

$$\psi \stackrel{\text{def}}{=} \frac{1}{2\sigma^2}. \tag{3.35}$$

Note that this definition of ψ is the SNR at a reference distance from the center of the function $(x_{\mathsf{c}}, y_{\mathsf{c}})$ at which the strength of the function samples is unit.

As it was mentioned in Section 3.3, there are two main classes of local processing schemes, namely analog and digital. Both of these schemes are considered in our simulations. For

the analog local processing scheme, it is assumed that the observation amplification gain is absorbed in the channel fading coefficient and normalized to one as described in the following paragraph. For the digital local processing scheme, it is assumed that the quantization rule at all of the sensors is the same. Therefore, all sensors quantize their local analog observations to $b \stackrel{\text{def}}{=} b_1 = b_2 = \cdots = b_K$ bits and use the same number of quantization levels $M \stackrel{\text{def}}{=} M_1 = M_2 = \cdots = M_K$. Each sensor uses a scalar deterministic uniform quantizer, whose set of quantization thresholds is known and is the same for all sensors. In other words, $\beta(\ell) \stackrel{\text{def}}{=} \beta_i(\ell)$, $\forall i$ and $\ell = 0, 1, \ldots, M$. In our simulations, the set of quantization thresholds for all sensors is chosen as $\beta(\ell) = \frac{\ell h}{M}$, $\ell = 1, 2, \ldots, M-1$.

The parallel, independent channels between distributed sensors and the FC are assumed to experience Rayleigh fading. The channel fading coefficients are normalized to ensure that $\mathbb{E}\left[|h_i|^2\right] = 1$ (or more generally, $\mathbb{E}\left[|\alpha_i h_i|^2\right] = 1$ when applicable), $i = 1, 2, \ldots, K$. When the sensors are located close to each other and the FC is far away from them, the distance between the sensors and the FC is approximately the same for all sensors, and this assumption is valid. In our simulations, it is assumed that the parallel channels between distributed sensors and the FC have *homogeneous* noises. In other words, it is assumed that the additive white Gaussian noises of the channels between all sensors and the FC are *identically distributed* and their variances are the same, i.e.,

$$\tau^2 \stackrel{\text{def}}{=} \tau_1^2 = \tau_2^2 = \cdots = \tau_K^2.$$

We define the *channel SNR* as

$$\eta \stackrel{\text{def}}{=} \frac{1}{2\tau^2}. \tag{3.36}$$

Again, this definition of η is the SNR at a reference distance from each sensor at which the strength of the faded transmitted signal is unit.

For the analog local processing scheme, the ML estimate of the vector of unknown parameters $\boldsymbol{\theta}$ is found based on the system of non-linear equations given in (3.8). This system of equations is linearized and solved iteratively using Newton's method as briefly described at the end of Section 3.6. For the digital local processing scheme, the linearized EM algorithm given in Equations (3.30)–(3.31) is used as an iterative, efficient approach to numerically find the ML estimate of $\boldsymbol{\theta}$. To solve this system of non-linear equations, Newton's linearization

method is applied, as briefly described at the end of Section 3.6 and summarized in Equation (3.32). In all of our simulations, the *true* values of the parameters to be estimated are $h = 8$, $x_{\mathsf{c}} = 5$, and $y_{\mathsf{c}} = 3$. Moreover, the initial values of the parameters fed to the Newton's linearization algorithm are $h^{\mathsf{Init}} = 7$, $x_{\mathsf{c}}^{\mathsf{Init}} = 4$, and $y_{\mathsf{c}}^{\mathsf{Init}} = 2$.

We have chosen the *integrated mean-squared error* (IMSE) of the estimation, defined as

$$\mathsf{IMSE} \stackrel{\text{def}}{=} \iint_{\mathcal{A}} \mathsf{MSE}\,(x, y)\,\mathrm{d}x\mathrm{d}y, \tag{3.37}$$

as the performance measure to evaluate the ML-estimation techniques developed in this chapter, where $\mathcal{A}$ is the two-dimensional observation environment and $\mathsf{MSE}\,(x, y)$ is the *location-dependent* mean-squared error of the estimation at location $(x, y) \in \mathcal{A}$ defined as

$$\mathsf{MSE}\,(x, y) \stackrel{\text{def}}{=} \mathbb{E}_{\mathbf{h},\mathbf{w},\mathbf{n}}\left[\left(g(x, y) - \widehat{g}(x, y)\right)^2\right], \tag{3.38}$$

where $\mathbb{E}_{\mathbf{h},\mathbf{w},\mathbf{n}}[\cdot]$ denotes the expectation with respect to the observation noise, channel noise, and channel fading coefficients, $g(x, y)$ is the *true* sample of the underlying function at location (x, y), and $\widehat{g}(x, y)$ is its estimated value based on the estimate of the vector of unknown parameters $\boldsymbol{\theta} \stackrel{\text{def}}{=} [h, x_{\mathsf{c}}, y_{\mathsf{c}}]^T$. In our simulations, the values of $\mathsf{MSE}\,(x, y)$ are also averaged with respect to the random location of distributed sensors in the observation environment.

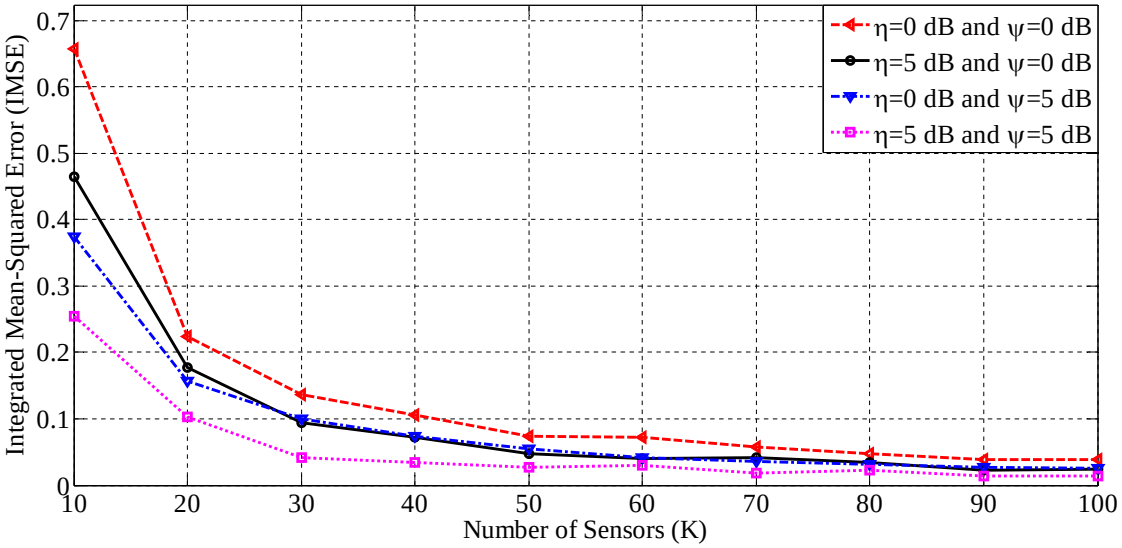
We use a Mante-Carlo method to calculate the IMSE values based on Equation (3.37). In this Monte-Carlo simulation, 100 random placements of K sensors in the observation environment are generated. For each random sensor placement, 10^4 realizations of the observation noise, channel noise, and channel fading coefficients have been generated based on the statistical models described in Section 3.3. The IMSE values shown in the figures depicted in the rest of this section are averaged over all of these realizations. Therefore, our simulation results average the effects of random sensor placement in the observation environment as well as the randomness in the observation noise, channel noise, and channel fading coefficients on the performance of the proposed distributed estimation framework.

3.7.2 Effects of K, and Observation and Channel SNR on the Performance of Distributed Estimation Framework

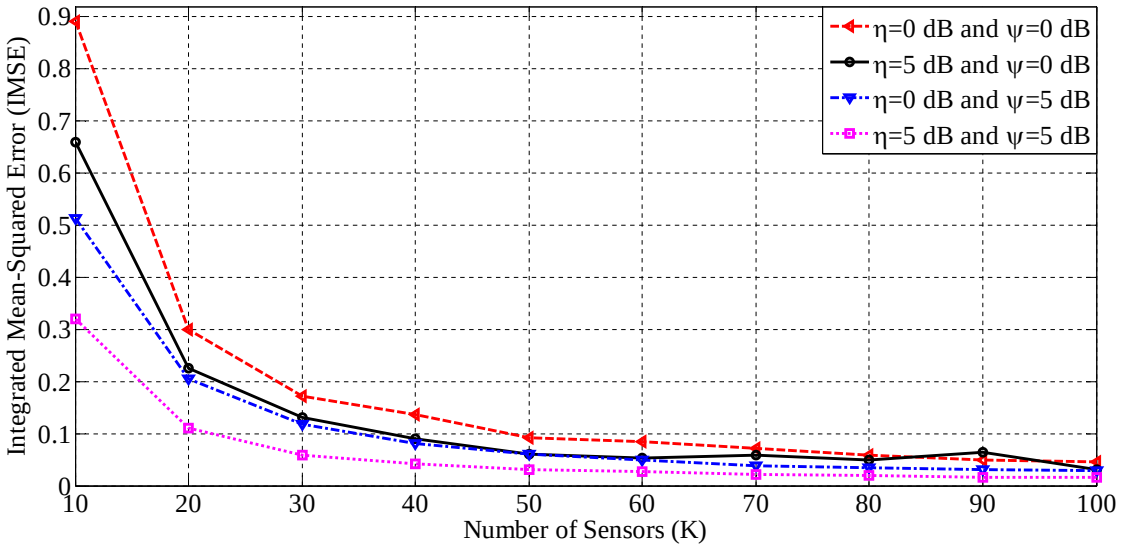
The performance of the proposed distributed estimation framework is, in part, a function of the number of distributed sensors in the observation environment, the observation SNR, and the channel SNR.

In Figure 3.2, the integrated mean-squared error at the FC versus the number of distributed sensors in the observation environment (K) is shown for different values of the observation SNR (ψ) and channel SNR (η). Figure 3.2a shows this performance measure for applying the ML-estimation technique at the FC, when the analog local processing scheme is used. Figure 3.2b shows the IMSE as the estimation performance measure for applying the linearized EM estimation algorithm at the FC, when the digital local processing scheme is used with $M = 8$ quantization levels. As it can be seen in this figure, as the number of sensors in the observation environment increases, the IMSE decreases monotonically. This conclusion is valid for both analog and digital local processing schemes. Furthermore, it can be observed from Figure 3.2 that the performance improvement due to the increase in the density of sensors in the observation environment is more considerable when there are small number of sensors. As the number of sensors increases, the percentage of this performance improvement decreases, and the distributed estimation technique achieves an acceptable performance in terms of the IMSE at a moderate number of sensors in the observation environment.

Figure 3.3 shows the integrated mean-squared error at the FC versus the observation SNR (ψ) for different values of the number of distributed sensors in the observation environment (K) and channel SNR (η). Figure 3.3a shows the IMSE as the performance measure for applying the ML-estimation technique at the FC, when the analog local processing scheme is used. Figure 3.3b shows the IMSE as the estimation performance measure for applying the linearized EM estimation algorithm at the FC, when the digital local processing scheme is used with $M = 8$ quantization levels. Similar to detailed discussions provided in analyzing Figure 3.2, as the observation SNR increases, the IMSE decreases monotonically for both cases of analog and digital local processing schemes. Furthermore, it can be observed from



(a) ML estimation at the FC based on *analog* local processing.



(b) Linearized EM estimation at the FC based on *digital* local processing with $M = 8$ quantization levels.

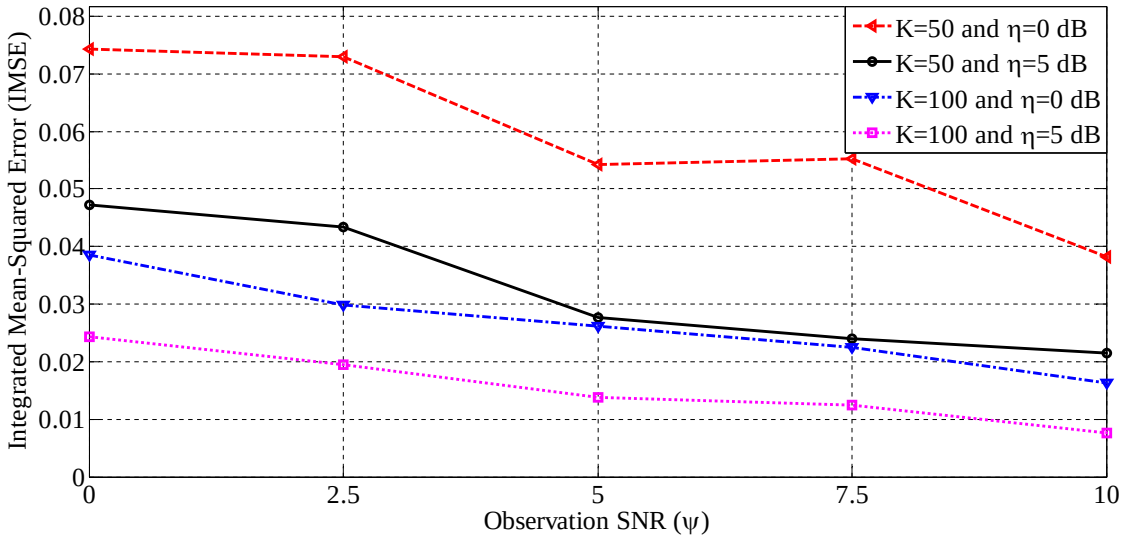
Figure 3.2: IMSE versus the number of distributed sensors in the observation environment (K) for different values of the observation SNR (ψ) and channel SNR (η).

Figure 3.3 that the performance improvement due to the increase in the observation SNR is more considerable at low SNRs. As the observation SNR increases, the percentage of this performance improvement decreases, and the distributed estimation framework achieves an acceptable performance in terms of the IMSE at a moderate observation SNR. Similar results and discussions can be provided for analyzing the effects of channel SNR on the performance of the distributed estimation framework, which are omitted because of the limited space.

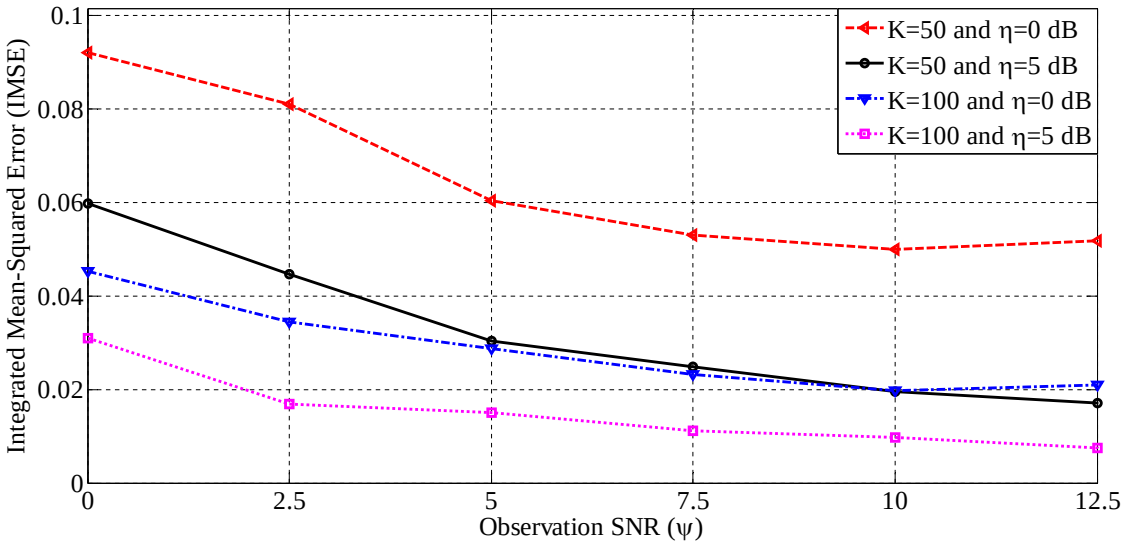
One of the most important points to be noticed in Figures 3.2 and 3.3 is that for the same values of K, observation SNR ψ, and channel SNR η, the ML estimation at the FC based on the analog local processing scheme outperforms the linearized EM estimation at the FC based on the digital local processing scheme with $M = 8$ quantization levels. This observation is expected for two main reasons: First, the linearized EM algorithm is an efficient iterative method for *numeric* calculation of the ML estimate; therefore, it suffers from a degraded performance compared to the exact ML estimate. Second, by performing digital signal processing and quantizing their noisy observations, local sensors are introducing a quantization noise in the processed samples to become available at the FC. Therefore, the results of the estimation at the FC based on the received analog samples are more accurate than those based on the quantized versions of local samples.

3.7.3 Effects of M on the Performance of Distributed Linearized EM Estimation Framework

Besides K, observation SNR ψ, and channel SNR η, one of the major parameters that affects the performance of the linearized EM estimation at the FC, when the digital local processing scheme is used, is the number of quantization levels at local sensors, i.e., M. Figure 3.4 shows the integrated mean-squared error at the FC versus the number of quantization levels at local sensors (M) for different values of the number of distributed sensors in the observation environment (K), observation SNR (ψ), and channel SNR (η), when the digital local processing scheme is used at distributed sensors and the linearized EM estimation algorithm is applied at the FC. As it can be seen in this figure, as the number of quantization levels at local sensors increases, the IMSE decreases monotonically. Again, it can be observed from Figure 3.4



(a) ML estimation at the FC based on *analog* local processing.



(b) Linearized EM estimation at the FC based on *digital* local processing with $M = 8$ quantization levels.

Figure 3.3: IMSE versus the observation SNR (ψ) for different values of the number of distributed sensors in the observation environment (K) and channel SNR (η).

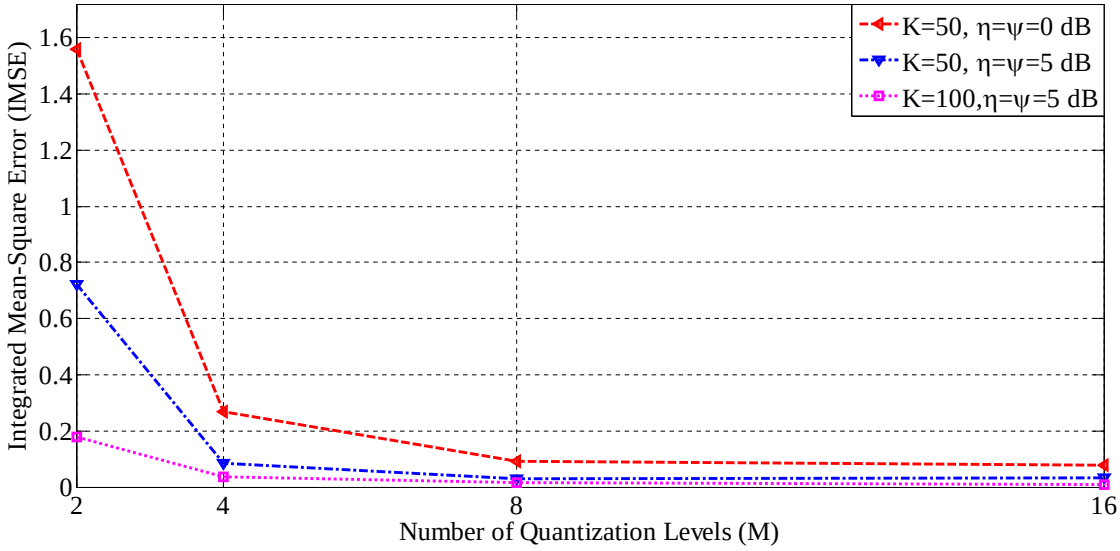


Figure 3.4: IMSE versus the number of quantization levels at local sensors for linearized EM estimation at the FC based on *digital* local processing for different values of the number of distributed sensors in the observation environment (K), observation SNR (ψ), and channel SNR (η).

that the performance improvement due to the increase in the umber of quantization levels at local sensors is more considerable for small values of M. As the umber of quantization levels at local sensors increases, the percentage of this performance improvement decreases, and the distributed estimation technique achieves an acceptable performance in terms of the IMSE at a reasonably low number of quantization levels. It is worth mentioning that Figure 3.4 shows that even $M = 4$ quantization levels at local sensors can achieve a very good performance in terms of the IMSE. In other words, even if the sensors quantize their noisy observations to only two bits, the system shows an acceptable performance. This conclusion emphasizes on the *energy efficiency* of the proposed estimation framework that can lead to a higher lift-time of distributed sensors in the network. In other words, the sensors do not need to waste a lot of energy to send very high-resolution observations quantized to a large number of quantization levels to achieve an acceptable estimation error performance at the FC in terms of the IMSE.

3.8 Conclusions

In this chapter, the problem of distributed estimation of a vector of unknown deterministic parameters associated with a two-dimensional function was considered in the context of WSNs. Each sensor observes a sample of the underlying function at its location, corrupted by an additive white Gaussian observation noise, whose samples are spatially uncorrelated across sensors. After local processing, each sensor transmits its locally processed sample to the FC of a WSN through parallel, coherent fading channels. Two local processing schemes were considered, namely analog and digital. In the analog local processing scheme, each sensor transmits an amplified version of its analog noisy observation to the FC, acting like a relay in a wireless network. In the digital local processing scheme, each sensor quantizes its noisy observations using a deterministic uniform scaler quantizer before transmitting its digitally modulated version to the FC. The ML estimate of the vector of unknown parameters at the FC was derived for both analog and digital local processing schemes. Since the ML estimate for the case of digital local processing scheme was too complicated to be implemented, an efficient iterative EM algorithm was proposed to numerically find the ML estimate in this case. Numerical simulation results were provided to evaluate the performance of the proposed distributed estimation framework in a typical WSN application scenario. As shown in the results of these simulations, the proposed distributed estimation framework achieves a very good performance in terms of the integrated mean-squared error for reasonable values of the parameters of the system, including the number of distributed sensors in the observation environment, the observation SNR, the channel SNR, and the number of quantization levels for digital local processing scheme. In particular, numerical performance analysis showed that even with a low number of quantization levels at distributed sensors, i.e., high energy efficiency, the estimation framework provides a very good performance in terms of the integrated mean-squared error.

Chapter 4

Channel-aware Power Allocation for Distributed BLUE Estimation: Full and Limited Feedback of CSI

4.1 Introduction

Consider a wireless sensor network (WSN) in which the noisy observations of spatially distributed sensors are correlated with an unknown random signal to be estimated. Suppose that the sensors transmit their local analog noisy observations to the fusion center (FC) using an amplify-and-forward strategy as described in Subsection 1.1.2 under Equation (1.2). Analog local processing is considered in this chapter due to its simplicity and practical feasibility. Assume that the FC finds the best linear unbiased estimator (BLUE) of the unknown random signal by combining linearly processed noisy observations of sensors received through orthogonal channels corrupted by fading and additive Gaussian noise. This chapter investigates one of the main challenges of distributed estimation in the case of analog amplify-and-forward local processing, which is finding the optimal local amplification gains [17–20]. The values of these gains set the instantaneous transmission power of sensors; therefore, we refer to their determination as the optimal, adaptive *power allocation* to sensors.

Cui et al. [17] have proposed an optimal power-allocation scheme to minimize the sum of the local transmission powers, given a maximum estimation distortion defined as the

variance of the BLUE estimator of a scalar, random signal at the FC of a WSN. Although optimal with respect to the total transmission power in the network, this strategy can result in assigning very high transmission power to sensors with high quality observations and less noisy channels while assigning zero transmission power to other sensors. The direct consequence of such power allocation is that some sensors will die quickly, which can in turn result in a network partition, while the remaining sensors have either low observation quality or too noisy communication channels. In order to alleviate this drawback, we propose an optimal, adaptive power-allocation strategy that minimizes the L^2-norm of the vector of local transmission powers, given a maximum estimation distortion as defined above. This approach prevents the assignment of high transmission power to sensors by putting a higher penalty on them, which in itself reduces the chances of those sensors dying and the network becoming partitioned. In other words, the proposed scheme results in the increased lifetime of the WSN compared to similar approaches that are based on the minimization of the sum of the local transmission powers. Furthermore, the total transmission power used in the entire network still stays bounded.

As it will be seen in Section 4.3, the limitation of the proposed power-allocation scheme is that the optimal local amplification gains found based on it depend on the instantaneous fading coefficients of the channels between the sensors and FC, as is the case in [17]. Therefore, the FC must feed the exact channel fading coefficients back to sensors through infinite-rate, error-free links. This requirement of the feedback of the instantaneous channel state information (CSI) from the FC to local sensors is not practical in most applications of WSNs, especially when the number of sensors in the network is large. In the remainder of this chapter, we propose a *limited-feedback strategy* to eliminate this requirement. The proposed approach is based on designing an optimal codebook using the generalized Lloyd algorithm with modified distortion metrics, which is used to quantize the space of the optimal power-allocation vectors used by the sensors to set their local amplification gains. Based on this scheme, each sensor amplifies its analog noisy observations using a quantized version of its optimal amplification gain determined by the designed optimal codebook.

In summary, the main contributions of this chapter are as follows:

(1) An optimal, adaptive power-allocation scheme is proposed to minimize the L^2-norm of the vector of local transmission powers, given a maximum estimation distortion (defined as the variance of the BLUE estimator) at the FC. This scheme alleviates the problem of assigning very high transmission power to some sensors while turning off the other ones.

(2) A limited-feedback strategy is proposed to quantize the vector space of the optimal local amplification gains. Appropriate distortion metrics are defined for the application of the generalized Lloyd algorithm in the domain of adaptive power allocation for distributed estimation.

The rest of this chapter is organized as follows: In Section 4.2, the system model of the WSN under study is described. The proposed adaptive power-allocation strategy is derived in Section 4.3. A brief discussion on the motivation for and implementation of the limited feedback for the proposed power-allocation scheme is presented in Section 4.4. Details of the implementation of the proposed limited-feedback scheme are discussed in Section 4.5. Section 4.6 provides the numerical results to show the applicability of the proposed schemes. Finally, the chapter is concluded in Section 4.7.

4.2 System Model

The system model of the WSN considered in this chapter, which is depicted in Figure 4.1, is the same as the one described in Section 1.1. The goal of the WSN is to reliably estimate an unknown random signal θ at its FC using linearly amplified versions of local noisy observations received through parallel (orthogonal) coherent communication channels corrupted by fading and additive Gaussian noise. An example of the unknown random variable θ to be estimated is the intensity of the signal originated from an energy-emitting source and sensed by a set of spatially distributed signal detectors. This estimated variable along with the propagation model of the given signal in the observation environment can then be used to estimate the location of the source. It is assumed that θ has zero mean and variance σ_θ^2, and is otherwise unknown.

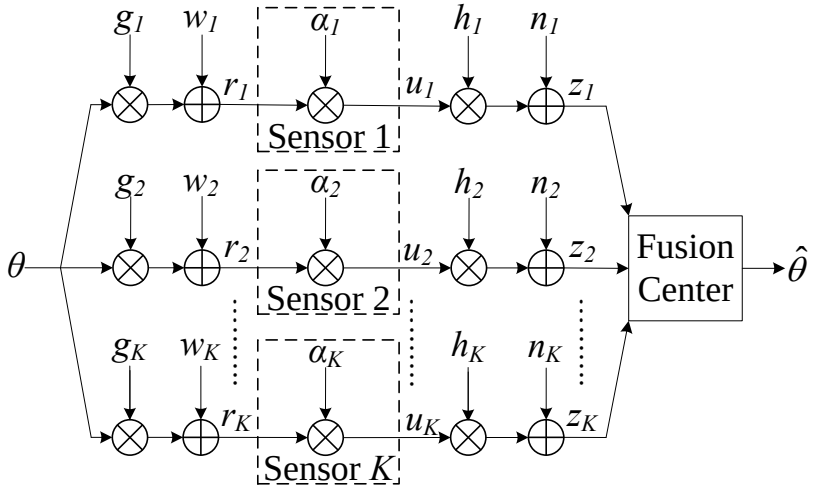


Figure 4.1: System model of a WSN in which the FC finds an estimate of θ.

Assume that the local noisy observation at each sensor is a linear function of the unknown random signal. Therefore, the observation function $\Xi_i(\cdot)$ in Equation (1.1) is defined as

$$\Xi_i(\theta) \stackrel{\text{def}}{=} g_i\theta,$$

where g_i is the fixed local observation gain at sensor i, *completely known* at the sensor and FC, and the local noisy observation at the ith sensor is found as

$$r_i = g_i\theta + w_i, \qquad i = 1, 2, \ldots, K, \tag{4.1}$$

where w_i is the spatially independent additive observation noise with zero mean and known variance σ_i^2, which is independent from the unknown signal θ. Note that unlike the local observation model described in Subsection 1.1.1, no further assumption is made on the type of the distribution of the random signal to be estimated and on that of the observation noise. We define the *observation signal-to-noise ratio* (SNR) at sensor i as

$$\psi_i \stackrel{\text{def}}{=} \frac{|g_i|^2 \sigma_\theta^2}{\sigma_i^2}, \qquad i = 1, 2, \ldots, K,$$

where $|\cdot|$ denotes the absolute-value operation.

Suppose that the local processing rule at each sensor is the same as the one introduced in Subsection 1.1.2 under "analog local processing" and Equation (1.2), where the local amplification gain α_i is to be optimally found in our analyses. Note that based on this local

processing rule, the instantaneous transmission power of sensor i is found as

$$P_i = \alpha_i^2 \left(|g_i|^2 \sigma_\theta^2 + \sigma_i^2 \right) = \alpha_i^2 \sigma_i^2 \left(1 + \psi_i \right). \tag{4.2}$$

As it can be seen in Equation (4.2), the value of the local amplification gain at each sensor determines the instantaneous transmission power allocated to that sensor. Therefore, we will call any strategy that assigns a set of local amplification gains to sensors a *power-allocation scheme*.

The model of the communication channels between the sensors and the FC is the same as the one introduced in Subsection 1.1.3. It is assumed that the channel fading coefficients h_i are uncorrelated and can reliably be estimated by the FC. We define the *channel SNR* of the signal received from sensor i at the FC as[1]

$$\eta_i \stackrel{\text{def}}{=} \frac{|h_i|^2}{\tau_i^2}, \qquad\qquad i = 1, 2, \ldots, K.$$

4.3 Optimal Power Allocation with Minimal L^2-Norm of Transmission-Power Vector

Given a power-allocation scheme and a realization of the fading coefficients of the communication channels, the FC combines the set of received signals from different sensors to find the *best linear unbiased estimator* for the unknown signal θ as [84, Chapter 6]

$$\widehat{\theta} = \left(\sum_{i=1}^{K} \frac{|g_i|^2 \alpha_i^2 |h_i|^2}{\alpha_i^2 |h_i|^2 \sigma_i^2 + \tau_i^2} \right)^{-1} \sum_{i=1}^{K} \frac{g_i \alpha_i h_i z_i}{\alpha_i^2 |h_i|^2 \sigma_i^2 + \tau_i^2}, \tag{4.3}$$

where the corresponding estimator variance is found as

$$\begin{aligned} \mathsf{Var}\left(\widehat{\theta} \middle| \boldsymbol{\alpha}, \mathbf{h} \right) &= \left(\sum_{i=1}^{K} \frac{|g_i|^2 \alpha_i^2 |h_i|^2}{\alpha_i^2 |h_i|^2 \sigma_i^2 + \tau_i^2} \right)^{-1} \\ &= \sigma_\theta^2 \left(\sum_{i=1}^{K} \frac{\psi_i \eta_i \alpha_i^2 \sigma_i^2}{1 + \eta_i \alpha_i^2 \sigma_i^2} \right)^{-1}, \end{aligned} \tag{4.4}$$

[1]Note that the defined parameter η_i is not an exact definition of the channel SNR. We have loosely defined this parameter to quantify the quality of the communication channel between each sensor and the FC and to simplify our derivations.

in which $\boldsymbol{\alpha} \stackrel{\text{def}}{=} [\alpha_1, \alpha_2, \ldots, \alpha_K]^T$ and $\mathbf{h} \stackrel{\text{def}}{=} [h_1, h_2, \ldots, h_K]^T$ are column vectors containing the set of local amplification gains α_i and fading coefficients of the channels between local sensors and the FC (i.e., h_i), respectively.

One of the goals of this chapter is to find the optimal local amplification gains or equivalently, the optimal power-allocation scheme that minimizes the L^2-norm of the vector of local transmission powers defined as $\mathbf{P} \stackrel{\text{def}}{=} [P_1, P_2, \ldots, P_K]^T$, given a constraint on the variance of the BLUE estimator at the FC as defined in Equation (4.4). This objective can be formulated as the following *convex* optimization problem:

$$\begin{aligned} \underset{\{P_i\}_{i=1}^K}{\text{minimize}} \quad & \left(\sum_{i=1}^{K} P_i^2\right)^{\frac{1}{2}} \\ \text{subject to} \quad & \mathsf{Var}\left(\widehat{\theta}\middle|\boldsymbol{\alpha}, \mathbf{h}\right) \leq D_0 \end{aligned} \tag{4.5}$$

By replacing P_i and $\mathsf{Var}\left(\widehat{\theta}\middle|\boldsymbol{\alpha}, \mathbf{h}\right)$ from Equations (4.2) and (4.4), respectively, Equation (4.5) is converted to the following convex form (in terms of ${\alpha_i}^2$), whose optimization variables are the (squared of the) local amplification gains:

$$\begin{aligned} \underset{\{{\alpha_i}^2\}_{i=1}^K}{\text{minimize}} \quad & \sum_{i=1}^{K} \left[{\alpha_i}^2 \sigma_i^2 \left(1 + \psi_i\right)\right]^2 \\ \text{subject to} \quad & \sum_{i=1}^{K} \frac{\psi_i \eta_i \alpha_i^2 \sigma_i^2}{1 + \eta_i \alpha_i^2 \sigma_i^2} \geq \frac{\sigma_\theta^2}{D_0} \quad \text{and} \quad {\alpha_i}^2 \geq 0 \end{aligned} \tag{4.6}$$

Let b_i be defined as $b_i \stackrel{\text{def}}{=} \frac{\psi_i \eta_i \alpha_i^2 \sigma_i^2}{1 + \eta_i \alpha_i^2 \sigma_i^2}$. The above constrained optimization problem can be rewritten as

$$\begin{aligned} \underset{\{b_i\}_{i=1}^K}{\text{minimize}} \quad & \sum_{i=1}^{K} \left(\frac{\left(1 + \psi_i\right) b_i}{\left(\psi_i - b_i\right) \eta_i}\right)^2 \\ \text{subject to} \quad & \sum_{i=1}^{K} b_i \geq \frac{\sigma_\theta^2}{D_0} \quad \text{and} \quad 0 \leq b_i < \psi_i \end{aligned} \tag{4.7}$$

which is a *convex* optimization problem in terms of b_i.[2] The Lagrangian function for this constrained optimization problem is

$$L\left(\mathbf{b}, \lambda, \boldsymbol{\mu}\right) \stackrel{\text{def}}{=} \sum_{i=1}^{K} \left(\frac{\left(1 + \psi_i\right) b_i}{\left(\psi_i - b_i\right) \eta_i}\right)^2 + \lambda \left(\frac{\sigma_\theta^2}{D_0} - \sum_{i=1}^{K} b_i\right) - \sum_{i=1}^{K} \mu_i b_i, \tag{4.8}$$

[2]Note that b_i can be rewritten as $b_i \stackrel{\text{def}}{=} \frac{\psi_i \eta_i \sigma_i^2}{\eta_i \sigma_i^2 + \frac{1}{\alpha_i^2}}$. Since $\alpha_i^2 \geq 0$, the range over which b_i can change is found as $0 \leq b_i < \psi_i$.

where $\mathbf{b} \stackrel{\text{def}}{=} [b_1, b_2, \ldots, b_K]^T$ is the column vector of target optimized variables, and $\boldsymbol{\mu} \stackrel{\text{def}}{=} [\mu_1, \mu_2, \ldots, \mu_K]^T$ is the Lagrangian multiplier vector. The Karush-Kuhn-Tucker (KKT) conditions [85, Theorem 18.6] for this optimization problem can be written as

$$\frac{\partial L\left(\mathbf{b}, \lambda, \boldsymbol{\mu}\right)}{\partial b_i} = \frac{2\psi_i \left(1+\psi_i\right)^2 b_i}{\left(\psi_i - b_i\right)^3 \eta_i^2} - \lambda - \mu_i = 0, \tag{4.9a}$$

$$\sum_{i=1}^{K} b_i = \frac{\sigma_\theta^2}{D_0}, \tag{4.9b}$$

$$\mu_i b_i = 0, \qquad i = 1, 2, \ldots, K, \tag{4.9c}$$

$$\mu_i \geq 0 \quad \text{and} \quad b_i \geq 0, \qquad i = 1, 2, \ldots, K. \tag{4.9d}$$

Note that since the objective function of the constrained optimization problem in Equation (4.7) is an increasing function with respect to b_i, its minimum is attained at the smallest possible value of b_i. Therefore, we have used the fact that the inequality constraint $\sum_{i=1}^{K} b_i \geq \frac{\sigma_\theta^2}{D_0}$ is converted into an equality constraint as in Equation (4.9b) at the optimal point of the optimization problem.

The first KKT condition in Equation (4.9a) can be simplified as

$$b_i^3 - 3\psi_i b_i^2 + \psi_i \left(3\psi_i + \frac{2\left(1+\psi_i\right)^2}{\left(\lambda+\mu_i\right)\eta_i^2}\right) b_i - \psi_i^3 = 0. \tag{4.10}$$

It can be shown that the cubic equation defined in Equation (4.10) *only* has a *unique*, *real* root as follows, provided that $\lambda + \mu_i > 0$:

$$b_i = \psi_i \underbrace{\left[1 - \sqrt[3]{\frac{\psi_i \, \delta_i^2 \, T_i'}{\lambda+\mu_i}} \left(1 - \frac{2}{3}\sqrt[3]{\frac{\psi_i \, \delta_i^2}{\left(\lambda+\mu_i\right) T_i'^2}}\right)\right]}_{\stackrel{\text{def}}{=} \Gamma_i}, \tag{4.11}$$

where $\delta_i \stackrel{\text{def}}{=} \frac{1+\psi_i}{\psi_i \eta_i}$, $i = 1, 2, \ldots, K$, and T_i' is defined as

$$T_i' \stackrel{\text{def}}{=} 1 + \sqrt{1 + \frac{8\psi_i \, \delta_i^2}{27\left(\lambda+\mu_i\right)}}\,.$$

Based on the complementary slackness requirement of Equation (4.9c) and the constraint of Equation (4.9d), $\mu_i = 0$ when $b_i > 0$, and $b_i = 0$ when $\mu_i > 0$. Therefore, Equation (4.11)

can be simplified to

$$b_i = \psi_i \left[1 - \sqrt[3]{\frac{\psi_i \, \delta_i^2 \, T_i}{\lambda}} \left(1 - \frac{2}{3} \sqrt[3]{\frac{\psi_i \, \delta_i^2}{\lambda T_i^2}} \right) \right]^+ , \tag{4.12}$$

where T_i is defined as

$$T_i \stackrel{\text{def}}{=} 1 + \sqrt{1 + \frac{8\psi_i \, \delta_i^2}{27\lambda}} \, ,$$

and the operator $[\cdot]^+$ is defined as

$$[x]^+ \stackrel{\text{def}}{=} \begin{cases} 0, & \text{if } \; x \leq 0. \\ x, & \text{if } \; x > 0. \end{cases}$$

It should be noted that the value of b_i from Equation (4.12) is in the interval $0 \leq b_i < \psi_i$. Furthermore, when Γ_i in Equation (4.11) is negative, eliminating $\mu_i > 0$ from it increases the value of T_i', which in turn decreases the value of Γ_i and b_i to a smaller negative quantity. Therefore, variable μ_i has been removed in Equation (4.12).

As the observation SNR ψ_i or channel SNR η_i decreases, the value of δ_i increases, which in turn increases the value of T_i and decreases the value of b_i. Therefore, if the sensors are sorted so that $\delta_1 \leq \delta_2 \leq \cdots \leq \delta_K$, *only* the first K_1 sensors with the least values of δ_i will have a positive value for b_i, and $b_i = 0$ for all $i > K_1$. The values of the number of active sensors K_1 for which $b_i > 0$, and the equality-constraint Lagrangian multiplier $\lambda > 0$[3] are unique and can be found by replacing b_i from Equation (4.12) into Equation (4.9b) to derive the following relationship between them:

$$\sum_{i=1}^{K_1} \psi_i \sqrt[3]{\frac{\psi_i \, \delta_i^2 \, T_i}{\lambda}} \left(1 - \frac{2}{3} \sqrt[3]{\frac{\psi_i \, \delta_i^2}{\lambda T_i^2}} \right) = \sum_{i=1}^{K_1} \psi_i - \frac{\sigma_\theta^2}{D_0}. \tag{4.13}$$

The values of K_1 and λ can be found through the water-filling-based iterative process summarized in Algorithm I. It can be shown that the solution of the above iterative algorithm in terms of K_1 and λ always exists and is unique.

Having found b_i through the above process, the local amplification gain α_i is found as

[3]Note that $\lambda + \mu_i > 0$ is the condition under which the cubic equation defined in Equation (4.10) has a unique real root. As $\mu_i = 0$ when $b_i > 0$, this condition is reduced to $\lambda > 0$.

ALGORITHM I: The water-filling-based iterative process to find the unique values for the number of active sensors K_1 and the equality-constraint Lagrangian constant λ.

Require: K, σ_θ^2, $\{\psi_i\}_{i=1}^K$, $\{\eta_i\}_{i=1}^K$, and D_0.

1. **Initialization**
2. **for** $i = 1, 2, \ldots, K$ **do**
3. $\quad \delta_i \longleftarrow \frac{1+\psi_i}{\psi_i \eta_i}$
4. **end for**
5. Sort the sensors based on the ascending values of δ_i so that $\delta_1 \leq \delta_2 \leq \cdots \leq \delta_K$.
6. $K_1 \longleftarrow K$
7. **EndInitialization**
8. **repeat**
9. Using the given value for K_1, find the value of λ by solving Equation (4.13).
10. Replace the value of λ in Equation (4.12) and find the new values of b_i, $i = 1, 2, \ldots, K$.
11. $K_1 \longleftarrow K_1 - 1$
12. **until** The values of b_i do not change from the previous iteration. In particular, $b_i > 0$ for all $i \leq K_1$, and $b_i = 0$ for all $i > K_1$.
13. **return** K_1 and λ.

follows:

$$\alpha_i^2 = \frac{b_i}{(\psi_i - b_i)\,\eta_i \sigma_i^2} = \begin{cases} \frac{1}{\eta_i \sigma_i^2} \left(\frac{\sqrt[3]{\frac{\lambda}{\psi_i \delta_i^2 T_i}}}{1 - \frac{2}{3}\sqrt[3]{\frac{\psi_i \delta_i^2}{\lambda T_i^2}}} - 1 \right), & i \leq K_1. \\ 0, & i > K_1. \end{cases} \tag{4.14}$$

The above power-allocation strategy assigns a zero amplification gain or equivalently, zero transmission power to the sensors for which δ_i is large because either the sensor's observation SNR and/or its channel SNR is too low. The assigned instantaneous transmission power to other sensors is non-zero and based on the value of δ_i for each sensor. Note that based on the above power-allocation scheme, there is a unique one-to-one mapping between $\mathbf{h}$ and $\boldsymbol{\alpha}$ that can be denoted as $\boldsymbol{\alpha} = f(\mathbf{h})$.

4.4 Limited Feedback for Adaptive Power Allocation

The optimal power-allocation scheme proposed in the previous section is based on the assumption that the complete forward CSI is available at the sensors. In other words, Equation (4.14) shows that the optimal value of the local amplification gain at sensor i is a function of its channel SNR η_i, which in itself is a function of the instantaneous fading coefficient of the channel between sensor i and the FC. Therefore, in order to achieve the minimum L^2-norm of the vector of local transmission powers, the FC must feed the instantaneous amplification gain α_i back to each sensor through an infinite-rate, error-free link.[4] This requirement is not practical in most applications, especially in large-scale WSNs, since the feedback information is typically transmitted through finite-rate, *digital* feedback links.

In the rest of this chapter, we propose a limited-feedback strategy to eliminate the above-mentioned requirement for infinite-rate digital feedback links from the FC to local sensors. For each channel realization, the FC first finds the optimal power-allocation scheme using the approach proposed in the previous section. Note that the FC has access to the perfect *backward* CSI; i.e., the instantaneous fading coefficient of the channel between each sensor and itself. Therefore, it can find the *exact* power-allocation strategy of the entire network based on Equation (4.14), given any channel realization. In the next step, the FC broadcasts the *index* of a quantized version of the optimized power-allocation vector to all sensors.

In the limited-feedback strategy summarized above, the FC and sensors must agree on a *codebook* of the local amplification gains or equivalently, a codebook of possible power-allocation schemes. The optimal codebook can be designed offline by quantizing the space of the optimized power-allocation vectors using the *generalized Lloyd algorithm* [86] with modified distortion metrics. Let L be the number of feedback bits that the FC uses to quantize the space of the optimal local power-allocation vectors into 2^L disjoint regions.

[4]Note that instead of feeding α_i back to each sensor, the FC can send back the fading coefficient of the channel between each sensor and itself. However, the knowledge of h_i alone is *not* enough for sensor i to compute the optimal value of its local amplification gain α_i. The sensor must also know whether it should transmit or stay silent. There are two ways that the extra information can be fed back to the sensors: (a) This data can be encoded in an extra one-bit command instructing each sensor to transmit or to stay silent, or (b) Each sensor can listen for the entire vector $\mathbf{h}$ sent by the FC over a broadcast channel and determine whether or not it should transmit based on the details of the power-allocation scheme discussed in Section 4.3. Sending back each value of α_i from the FC avoids this extra communication.

Note that L is the total number of feedback bits broadcast by the FC, and *not* the number of bits fed back to each sensor. A codeword is chosen in each quantization region. The length of each codeword is K, and its ith entry is a *real-valued* number representing a quantized version of the optimal local amplification gain for sensor i. The proposed quantization scheme can then be thought of as a mapping from the space of channel state information to a discrete set of 2^L length-K, real-valued power-allocation vectors. Details of this quantization method are described in the next section.

4.5 Codebook Design Using Lloyd Algorithm

Let $\mathbf{C} = [\boldsymbol{\alpha}_1 \ \boldsymbol{\alpha}_2 \ \cdots \ \boldsymbol{\alpha}_{2^L}]^T$ be a 2^L-by-K codebook matrix of the optimal local amplification gains, where $[\mathbf{C}]_{\ell,i}$ denotes its element in row ℓ and column i as the optimal gain of sensor i in codeword ℓ. Note that each $\boldsymbol{\alpha}_\ell$, $\ell = 1, 2, \ldots, 2^L$ is associated with a realization of the fading coefficients of the channels between sensors and the FC. We apply the generalized Lloyd algorithm with modified distortion metrics to solve the problem of vector quantization in the space of the optimal local amplification gains. This algorithm designs the optimal codebook $\mathbf{C}$ in an iterative process, as explained in the following discussions.

In order to implement the generalized Lloyd algorithm, one distortion metric must be defined for the codebook and one for each codeword. Let $D_{\mathsf{B}}(\mathbf{C})$ denote the average distortion for codebook $\mathbf{C}$ defined as

$$D_{\mathsf{B}}(\mathbf{C}) \stackrel{\text{def}}{=} \mathbb{E}_{\boldsymbol{\alpha}}\left[\min_{\ell\in\{1,2,\ldots,2^L\}} D_{\mathsf{W}}(\boldsymbol{\alpha}_\ell, \boldsymbol{\alpha})\right], \tag{4.15}$$

where $\mathbb{E}_{\boldsymbol{\alpha}}[\cdot]$ denotes the expectation operation with respect to the optimal vector of local amplification gains $\boldsymbol{\alpha}$, and $D_{\mathsf{W}}(\boldsymbol{\alpha}_\ell, \boldsymbol{\alpha})$ represents the distance between codeword $\boldsymbol{\alpha}_\ell$ and an optimal power-allocation vector $\boldsymbol{\alpha}$, defined as

$$D_{\mathsf{W}}(\boldsymbol{\alpha}_\ell, \boldsymbol{\alpha}) \stackrel{\text{def}}{=} |J(\boldsymbol{\alpha}_\ell) - J(\boldsymbol{\alpha})|, \tag{4.16}$$

where $J(\cdot)$ is the optimized cost of a power-allocation vector. Let $\mathbf{P}_\ell$ and $\mathbf{P}$ be the vectors of local transmission powers, when the vector of local amplification gains is $\boldsymbol{\alpha}_\ell$ and $\boldsymbol{\alpha}$, respectively. The cost function $J(\boldsymbol{\alpha})$ is defined as the L^2-norm of the corresponding vector

of transmission powers $\mathbf{P}$, i.e.,

$$J(\boldsymbol{\alpha}) \stackrel{\text{def}}{=} \left(\sum_{i=1}^{K} P_i^2\right)^{\frac{1}{2}} = \left(\sum_{i=1}^{K} \left[\alpha_i^2 \sigma_i^2 (1+\psi_i)\right]^2\right)^{\frac{1}{2}}. \tag{4.17}$$

Let $\mathcal{A} \subseteq \mathbb{R}^{K+}$ be the K-dimensional vector space of the optimal local amplification gains, whose entries are chosen from the set of real-valued, non-negative numbers. Given the distortion metric for codebook $\mathbf{C}$ and that for each one of its codewords defined in Equations (4.15) and (4.16), respectively, the two main conditions of the generalized Lloyd algorithm can be reformulated for our vector-quantization problem as follows [86, Chapter 11]:

Nearest–Neighbor Condition: This condition finds the optimal Voronoi cells of the vector space to be quantized, given a fixed codebook. Based on this condition, given a codebook $\mathbf{C}$, the space $\mathcal{A}$ of optimized power-allocation vectors is divided into 2^L disjoint quantization regions (or Voronoi cells) with the ℓth region represented by codeword $\boldsymbol{\alpha}_\ell \in \mathbf{C}$ and defined as

$$\mathcal{A}_\ell \stackrel{\text{def}}{=} \{\boldsymbol{\alpha} \in \mathcal{A} : D_{\mathsf{W}}(\boldsymbol{\alpha}_\ell, \boldsymbol{\alpha}) \leq D_{\mathsf{W}}(\boldsymbol{\alpha}_k, \boldsymbol{\alpha}), \forall k \neq \ell\}. \tag{4.18}$$

Centroid Condition: This condition finds the optimal codebook, given a specific partitioning of the vector space to be quantized. Based on this condition, given a specific partitioning of the space of the optimized power-allocation vectors $\{\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_{2^L}\}$, the optimal codeword associated with each Voronoi cell $\mathcal{A}_\ell \subseteq \mathcal{A}$ is the *centroid* of that cell with respect to the distance function defined in Equation (4.16) as

$$\boldsymbol{\alpha}_\ell^\star \stackrel{\text{def}}{=} \arg\min_{\boldsymbol{\alpha}_\ell \in \mathcal{A}_\ell} \mathbb{E}_{\boldsymbol{\alpha} \in \mathcal{A}_\ell}\left[D_{\mathsf{W}}(\boldsymbol{\alpha}_\ell, \boldsymbol{\alpha})\right], \tag{4.19}$$

where the expectation operation is performed over the set of members of partition $\mathcal{A}_\ell$.

The optimal codebook is designed offline by the FC using the above two conditions. It can be shown that the average codebook distortion defined in Equation (4.15) will monotonically decrease through the iterative usage of the Centroid Condition and the Nearest-Neighbor Condition [86, Chapter 11]. Details of the codebook-design process are summarized in Algorithm II. The optimal codebook is stored in the FC and is shared with all sensors.

ALGORITHM II: Iterative process of designing an optimal codebook of power-allocation vectors based on the generalized Lloyd algorithm with modified distortion metrics.

Require: K, σ_θ^2, $\{\sigma_i^2\}_{i=1}^K$, $\{\psi_i\}_{i=1}^K$, $\{\tau_i^2\}_{i=1}^K$, D_0, and L.
▷ L is the number of feedback bits.

Require: Fading model of the channel between local sensors and the FC.

Require: M. ▷ M is the number of *training vectors* in space $\mathcal{A}$.

Require: ϵ. ▷ ϵ is the distortion threshold to stop the iterative process.

1. **Initialization**
2. $\mathcal{H}_s \longleftarrow$ A set of M length-K vectors of channel-fading realizations based on the given fading model of the channels between local sensors and the FC.
▷ $M \gg 2^L$.
3. $\mathcal{A}_s \longleftarrow$ The set of optimal local power-allocation vectors associated with the channel fading vectors in $\mathcal{H}_s$, found using Equation (4.14).
▷ $\mathcal{A}_s$ is the set of training vectors, and $\mathcal{A}_s \subseteq \mathcal{A}$.
4. $\{\boldsymbol{\alpha}_\ell^0\}_{\ell=1}^{2^L} \longleftarrow$ *Randomly* select 2^L optimal power-allocation vectors from the set $\mathcal{A}_s$ as the initial set of codewords.
5. $\mathbf{C}^0 \longleftarrow \left[\boldsymbol{\alpha}_1^0 \;\; \boldsymbol{\alpha}_2^0 \;\; \cdots \;\; \boldsymbol{\alpha}_{2^L}^0\right]^T$ ▷ $\mathbf{C}^0$ is the initial codebook.
6. $\mathsf{NewCost} \longleftarrow D_{\mathsf{B}}(\mathbf{C}^0)$ and $j \longleftarrow 0$.
▷ The average distortion of codebook is found using Equation (4.15).
7. **EndInitialization**
8. **repeat**
9. $j \longleftarrow j+1$ and $\mathsf{OldCost} \longleftarrow \mathsf{NewCost}$.
10. Given codebook $\mathbf{C}^{j-1}$, optimally partition the set $\mathcal{A}_s$ into 2^L disjoint subsets based on the *Nearest-Neighbor Condition* using Equation (4.18). Denote the resulted optimal partitions by $\mathcal{A}_\ell^{j-1}$, $\ell = 1, 2, \ldots, 2^L$.
11. **for all** $\mathcal{A}_\ell^{j-1}$, $\ell = 1, 2, \ldots, 2^L$ **do**
12. $\boldsymbol{\alpha}_\ell^j \longleftarrow$ Optimal codeword associated with partition $\mathcal{A}_\ell^{j-1}$ found based on the *Centroid Condition* using Equation (4.19).
13. **end for**
14. $\mathbf{C}^j \longleftarrow \left[\boldsymbol{\alpha}_1^j \;\; \boldsymbol{\alpha}_2^j \;\; \cdots \;\; \boldsymbol{\alpha}_{2^L}^j\right]^T$ ▷ $\mathbf{C}^j$ is the new codebook.
15. $\mathsf{NewCost} \longleftarrow D_{\mathsf{B}}(\mathbf{C}^j)$
16. **until** $\mathsf{OldCost} - \mathsf{NewCost} \leq \epsilon$
17. **return** $\mathbf{C}^{\mathsf{OPT}} \longleftarrow \mathbf{C}^j$.

Upon observing a realization of the channel fading vector $\mathbf{h}$, the FC finds its associated optimal power-allocation vector $\boldsymbol{\alpha}^{\mathsf{OPT}}$, using Equation (4.14) to calculate each one of its elements. It then identifies the closest codeword in the optimal codebook $\mathbf{C}$ to $\boldsymbol{\alpha}^{\mathsf{OPT}}$ with respect to the distance metric defined in Equation (4.16). Finally, the FC broadcasts the L-bit *index* of that codeword over an error-free, digital feedback channel to all sensors as

$$\ell \stackrel{\text{def}}{=} \underset{k\in\{1,2,\ldots,2^L\},\boldsymbol{\alpha}_k\in\mathbf{C}}{\arg\min} D_{\mathsf{W}}\left(\boldsymbol{\alpha}_k, \boldsymbol{\alpha}^{\mathsf{OPT}}\right). \tag{4.20}$$

Upon the reception of the index ℓ, each sensor i knows its quantized local amplification gain or equivalently, its power-allocation weight as $[\mathbf{C}]_{\ell,i}$, where ℓ and i are the row and column indexes of the codebook $\mathbf{C}$, respectively.

4.6 Numerical Analysis

In this section, numerical results are provided to assess the performance of the optimal power-allocation scheme proposed in Section 4.3 and to verify the effectiveness of the limited-feedback strategy proposed in Section 4.5 in achieving the *energy efficiency* close to that of a WSN with full CSI feedback. In our analyses, the energy efficiency of a power-allocation scheme is defined as the L^2-norm of the vector of local transmission powers formulated in Equation (4.17).

In our simulations, we have set $\sigma_\theta^2 = 1$ and the local observation gains g_i are randomly chosen from a Gaussian distribution with unit mean and variance 0.09. In all simulations, the average power of g_i across all sensors is set to be 1.2, i.e., $\mathbb{E}\left[g_i^2\right] = 1.2$. The observation noise variances σ_i^2 are uniformly selected from the interval $(0.05, 0.15)$ such that the average power of the noise variances across all sensors is kept at 0.01, i.e., $\mathbb{E}\left[\sigma_i^2\right] = 0.01$. The channel noise variance for all sensors is set to $\tau_i^2 = -90$ dBm, $i = 1, 2, \ldots, K$. The following fading model is considered for the channels between the sensors and the FC:

$$h_i = \zeta_0 \left(\frac{d_i}{d_0}\right)^{-\frac{\alpha}{2}} f_i, \qquad i = 1, 2, \ldots, K,$$

where $\zeta_0 = -30$ dB is the nominal path loss at the reference distance set to be $d_0 = 1$ meter, d_i is the distance between sensor i and the FC (in meters), $\alpha = 2$ is the path-loss exponent,

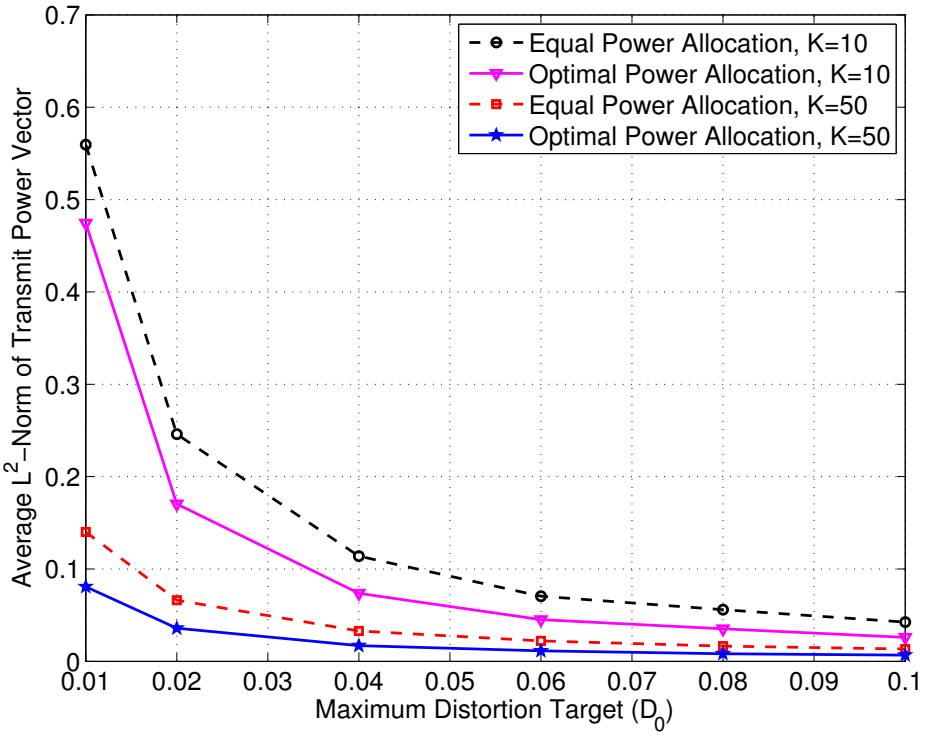


Figure 4.2: Average energy efficiency versus the target estimation distortion D_0 for the proposed adaptive power-allocation scheme and the equal power-allocation strategy.

and f_i is the independent and identically distributed (i.i.d.) Rayleigh-fading random variable with unit variance. The distance between sensors and the FC is uniformly distributed between 50 and 150 meters. The size of the training set in the optimal codebook-design process described in Algorithm II is set to $M = 5,000$, and the codebook-distortion threshold for stopping the iterative algorithm is assumed to be $\epsilon = 10^{-4}$. The results are obtained by averaging over 10,000 Monte-Carlo simulations.

Figure 4.2 illustrates the energy efficiency of the adaptive power-allocation scheme proposed in Section 4.3. The figure depicts the average L^2-norm of the vector of local transmission powers versus the maximum estimation distortion D_0 at the FC for different values of the number of sensors in the network K. The energy efficiency for the case of equal power allocation, i.e., the minimum transmission power required to achieve the given target estimation distortion at the FC, is also shown with dotted lines as a benchmark. As it can be seen in this figure, the energy efficiency of the network improves as the number of sensors increases. This is due to the fact that when there are fewer sensors in the network, each

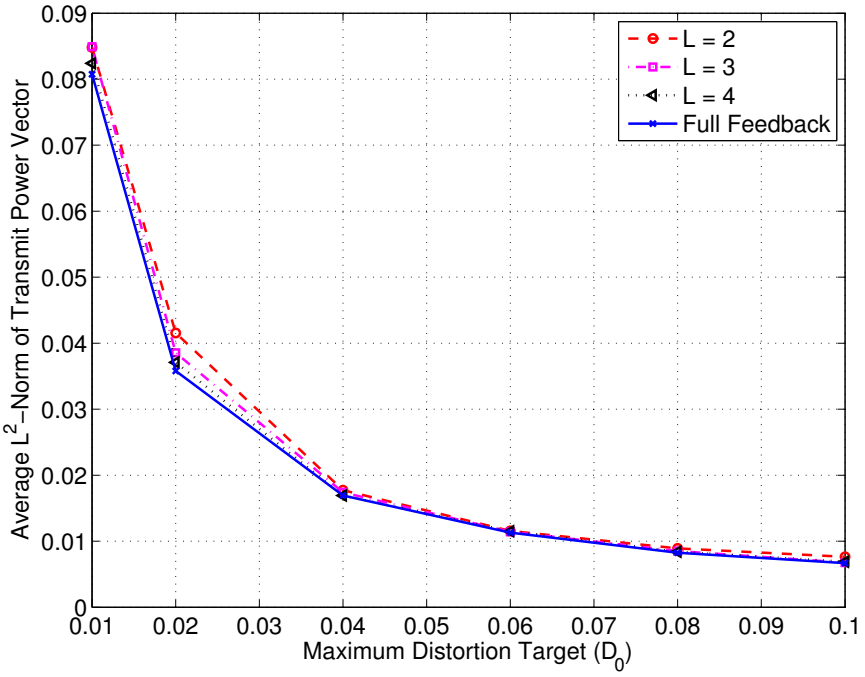


Figure 4.3: Average energy efficiency of the proposed power-allocation scheme versus the target estimation distortion D_0 for different values of the number of feedback bits L, when there are $K = 50$ sensors in the network.

one of them must transmit with a higher power in order for the FC to achieve the same estimation distortion. Note that in our analyses, there is no constraint on the total transmission power consumed in the entire network. Another observation from Figure 4.2 is that the proposed adaptive power allocation scheme achieves a higher energy efficiency than the equal power-allocation strategy. As the maximum estimation-distortion constraint at the FC is relaxed, i.e., the value of D_0 is increased, the gain in the energy efficiency decreases slightly.

Figure 4.3 illustrates the effect of L as the number of feedback bits from the FC to local sensors on the energy efficiency of the proposed power-allocation scheme. It should be emphasized that L is the total number of feedback bits broadcast by the FC, and not the number of bits fed back to each sensor. This figure depicts the average L^2-norm of the vector of local transmission powers versus the maximum estimation distortion D_0 at the FC for different values of the number of feedback bits L, when there are $K = 50$ sensors in the network. As it can be seen in this figure, the energy efficiency of the proposed adaptive

power allocation with limited feedback is very close to that with full feedback and gets closer to it as the number of feedback bits is increased.

4.7 Conclusions

In this chapter, an adaptive power-allocation scheme was proposed that minimizes the L^2-norm of the vector of local transmission powers in a WSN, given a maximum variance for the BLUE estimator of a scalar, random signal at the FC. This approach results in an increase in the lifetime of the network at the expense of a potential slight increase in the sum total transmission power of all sensors. The next contribution of this chapter was to propose a limited-feedback strategy to eliminate the requirement of infinite-rate feedback of the instantaneous forward CSI from the FC to local sensors. This scheme designs an optimal codebook by quantizing the vector space of the optimal local amplification gains using the generalized Lloyd algorithm with modified distortion metrics. Numerical results showed that the proposed adaptive power-allocation scheme achieves a high energy efficiency, and that even with a limited number of feedback bits (small codebook), its average energy efficiency based on the proposed limited-feedback strategy is close to that of a WSN with full CSI feedback.

Chapter 5

Linear Spatial Collaboration for Distributed BLUE Estimation

5.1 Introduction

In most studies in the literature, it is assumed that the spatially distributed sensors forming a wireless sensor network (WSN) do not communicate and/or collaborate with each other, and that the local processing is performed *only* on each sensor's noisy observations [16–26, 32, 33, 87–89]. In this chapter, we investigate the problem of distributed estimation under the assumption that local sensors collaborate with each other by sharing their local noisy observations. Consequently, the processing at each sensor connected to a fusion center (FC) will be performed on the combination of the sensor's own observations and those of the other sensors to which it has access. We study the problem of linear spatial collaboration for distributed estimation in the context in which each sensor can share its local noisy (and potentially spatially correlated) observations with a subset of sensors through error-free, low-cost links. An adjacency matrix defines the connectivity of the network and the pattern by which local sensors share their noisy observations with each other. The signals observed by different sensors are spatially correlated, and the goal of the WSN is for a FC to estimate the vector of unknown signals observed by individual sensors. Each one of the sensors that is connected to the FC forms a linear combination of the noisy observations to which it has access and sends the result of this analog local processing to the FC through an orthogonal

communication channel corrupted by fading and additive Gaussian noise. The FC combines the received data from spatially distributed sensors to find the best linear unbiased estimator (BLUE) of the vector of unknown signals observed by individual sensors. The main novelty of this chapter is the derivation of an optimal power-allocation scheme for a network with linear spatial collaboration in which the set of coefficients or weights used to form linear combinations of shared noisy observations at the sensors connected to the FC is optimized. Through this optimization, the total estimation distortion at the FC (defined as the sum of the estimation variances of the BLUE estimators for different signals observed by individual sensors) is minimized, given a constraint on the maximum average cumulative transmission power in the entire network. Numerical results show that even with a moderate connectivity across the network, spatial collaboration among sensors significantly reduces the estimation distortion at the FC.

The rest of this chapter is organized as follows: In Section 5.2, a summary of the related works to the ideas presented in this chapter will be provided, and the relationship between this work and similar studies in the literature will be explained. Section 5.3 introduces the system model of the WSN analyzed in this chapter. There are several differences between this system model and that introduced in Section 1.1, which will be explained in detail. In Section 5.4, an optimal power-allocation scheme will be derived that optimizes the mixing coefficients of local sensors with an objective to minimize the estimation distortion at the FC, given a constraint on the maximum average transmission power in the entire network. Numerical results in Section 5.5 show the effectiveness of the spatial collaboration among local sensors in improving the performance of the estimator at the FC. Finally, the chapter will be concluded in Section 5.6.

5.2 Related Works

Bahçeci and Khandani [87] have studied a WSN in which spatially distributed sensors make noisy observations of correlated Gaussian signals. They have assumed that each sensor amplifies its own local noisy observation before sending it to the FC through orthogonal channels corrupted by Rayleigh flat-fading and additive white Gaussian noise. The FC

then combines the received data from local sensors to estimate the set of correlated signals observed by the sensors, either using the BLUE or the minimum mean-squared error (MMSE) estimator. They have derived the optimal power-allocation scheme that minimizes the total cumulative transmission power in the entire network, given a constraint on the maximum estimation distortion at the FC, measured either as the estimation variance of each individual signal observed by one of the sensors, or as the average estimation variance of all signals of interest. It is crucial to emphasize that [87] assumes that there is no communication and/or collaboration among sensors.

Kar and Varshney [90] have studied the optimal power allocation for a WSN in which sensors collaborate with each other by sharing their local noisy observations. To the best of our knowledge, this is the first work that has considered sensor collaboration in the context of distributed estimation. In the system model studied in [90], each sensor connected to the FC, which in general can be in a subset of all sensors, forms a linear combination of its own noisy observation and the observations of other sensors to which it has access. This operation is known as the *linear spatial collaboration*. The sensor then sends the resulting linearly processed data to the FC through a coherent multiple access channel (MAC). The FC finds the linear minimum mean-squared error (LMMSE) estimator of a *scalar*, random signal observed by spatially distributed sensors. The gains used to form the linear combinations at local sensors are optimized to minimize the LMMSE estimator distortion at the FC, given a constraint on the maximum per-sensor or cumulative transmission power in the network. The results of their investigations show that even a moderate connectivity in the WSN drastically reduces the estimation distortion at the FC.

As the system model of the WSN described in Section 5.3 shows, our goal in this chapter is to generalize the network model studied in [87] by assuming that (a) the observation noises and channel noises are spatially correlated, (b) a subset of sensors is *not* directly connected to the FC, and more importantly, (c) the sensors collaborate with each other by sharing their local noisy observations through error-free, low-cost links. The most important aspect of this network model is the linear spatial collaboration among the sensors. Furthermore, we will study a generalized version of the problem investigated in [90]. In contrast with [90], the FC in our system model estimates the individual signals observed by distributed sensors and

not just an underlying scalar signal that is collaboratively observed by the entire network. Moreover, we consider the communication channels between the connected sensors and the FC to be orthogonal rather than a coherent MAC. Another contribution of our work is that the FC finds the BLUE estimator of the vector of unknown signals observed by local sensors rather than the LMMSE estimator. Note that unlike the LMMSE estimator, which depends on the statistics of the signals being observed and estimated, the BLUE estimator is independent of the source statistics and is useful when the information about the signals to be estimated is limited.

5.3 System Model

The system model of the WSN considered in this chapter, which is depicted in Figure 5.1, is different from the one described in Section 1.1 in several aspects. These differences will be clarified in this section.

Assume that $M \leq K$ sensors are connected to a FC, where K is the total number of the spatially distributed sensors forming a WSN, each one of which observes a noisy version of a local signal of interest. Using the received faded and noisy versions of locally processed sensor observations from a subset of sensors that are connected to it, the FC tries to find the BLUE estimator of the vector of signals observed by individual sensors.[1] Note that one of the major differences between the system considered here and most of the studies in the literature is that in our model (similar to [87]), the FC estimates the individual signals observed by local sensors rather than combining the observations to estimate a set of signals that are correlated with the collection of local observations and that can directly be observed by all sensors in the network. In other words, it is assumed that each signal to be estimated is *only* observed by one sensor although signals observed by different sensors may be correlated with each other.

Assume that the ith sensor observes a noisy version of a local signal of interest, denoted

[1]It is assumed that the observations of each sensor are communicated to the FC by itself if it is directly connected to the FC, by a subset of connected sensors to the FC with which it shares its observations if it is *not directly* connected to the FC, or by both.

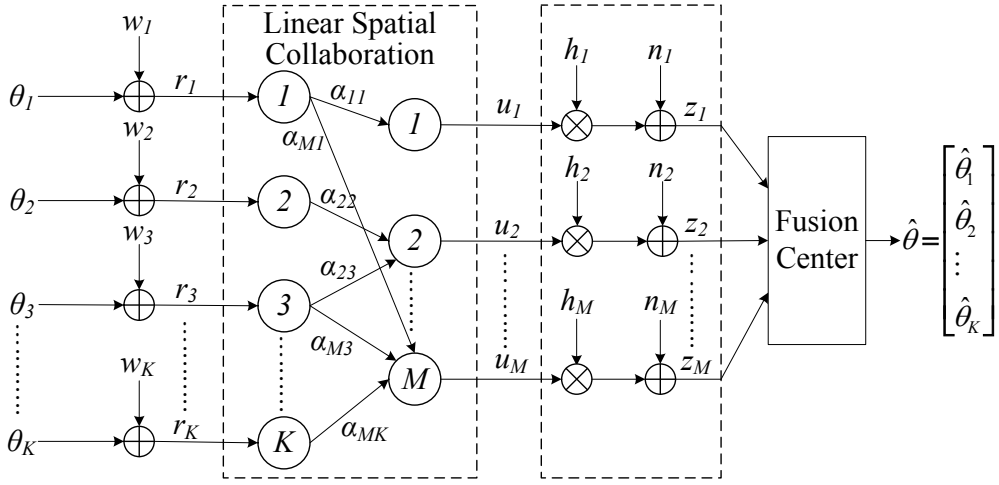


Figure 5.1: System model of a WSN with error-free inter-sensor collaboration in which the FC finds an estimate of $\boldsymbol{\theta} \stackrel{\text{def}}{=} [\theta_1, \theta_2, \ldots, \theta_K]^T$.

by θ_i. Therefore, the observation function $\Xi_i(\cdot)$ in Equation (1.1) is defined as

$$\Xi_i(\boldsymbol{\theta}) \stackrel{\text{def}}{=} \theta_i,$$

and the local noisy observation at the ith sensor is found as

$$r_i = \theta_i + w_i, \qquad i = 1, 2, \ldots, K, \tag{5.1}$$

where θ_i is the local unknown random signal to be estimated at the FC, and w_i is the observation noise. Assume that the random vector of signals observed at different sensors $\boldsymbol{\theta} \stackrel{\text{def}}{=} [\theta_1, \theta_2, \ldots, \theta_K]^T$ has zero mean and is spatially correlated with the known auto-correlation matrix $\mathbf{R}_{\boldsymbol{\theta}} \stackrel{\text{def}}{=} \mathbb{E}\left[\boldsymbol{\theta}\boldsymbol{\theta}^T\right]$, where $(\cdot)^T$ represents the vector/matrix transpose operation and $\mathbb{E}[\cdot]$ denotes the expectation operation. Furthermore, assume that the vector of observation noises $\mathbf{w} \stackrel{\text{def}}{=} [w_1, w_2, \ldots, w_K]^T$ is Gaussian with zero mean and $\mathbf{R}_{\mathbf{w}} \stackrel{\text{def}}{=} \mathbb{E}\left[\mathbf{w}\mathbf{w}^T\right]$ defined as its auto-correlation matrix, i.e., $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_{\mathbf{w}})$. It is assumed that the random vectors $\boldsymbol{\theta}$ and $\mathbf{w}$ are independent.

With the exception of [90], most studies in the literature assume that there is no inter-sensor communication and/or collaboration, and that each sensor processes only its own local noisy observations before transmitting them to the FC. In this chapter, we assume that the sensors share their observations with each other through error-free, low-cost links.[2]

[2]If the distance between sensors is a lot smaller than the distance between sensors and the FC, we can ignore

Suppose that the inter-sensor connectivity is modeled by an M-by-K *adjacency matrix* $\mathbf{A}$, whose elements are either zero or one. If $\mathbf{A}_{j,i} = 1$, then sensor j has access to the local observations of sensor i through a low-cost link. Otherwise, $\mathbf{A}_{j,i} = 0$. Note that in general, the adjacency matrix $\mathbf{A}$ is not necessarily symmetric since a sensor may receive the observations of a subset of other sensors, but it may not share its own observations with them. Moreover, $\mathbf{A}_{j,j} = 1$, $j = 1, 2, \ldots, M$, as each sensor has access to its own local observations.

Suppose that the sensors are sorted so that the first M sensors are connected to the FC. Each connected sensor to the FC uses an amplify-and-forward strategy and forms a linear combination of all local observations to which it has access as

$$u_j = \sum_{\substack{i=1 \\ \mathbf{A}_{j,i}=1}}^{K} \alpha_{j,i} r_i, \qquad j = 1, 2, \ldots, M, \tag{5.2}$$

where u_j is the locally processed and transmitted data by the jth sensor, and $\alpha_{j,i}$ is the weight of the ith observation in the linear combination that sensor j forms to be transmitted to the FC. Note that the above analog local processing can be rewritten in a vector form as

$$\mathbf{u} = \boldsymbol{\alpha}\mathbf{r}, \tag{5.3}$$

where $\mathbf{u} \stackrel{\text{def}}{=} [u_1, u_2, \ldots, u_M]^T$ is the column vector of transmitted signals from the sensors that are connected to the FC, $\mathbf{r} \stackrel{\text{def}}{=} [r_1, r_2, \ldots, r_K]^T$ is the vector of local noisy observations, and $\boldsymbol{\alpha}$ is an M-by-K *mixing matrix*. It can easily be seen that $\boldsymbol{\alpha}_{j,i} = 0$ if $\mathbf{A}_{j,i} = 0$, and $\boldsymbol{\alpha}_{j,i} = \alpha_{j,i}$ if $\mathbf{A}_{j,i} = 1$. Note that the average cumulative transmission power of the entire network is found as

$$\begin{aligned} P_{\mathsf{Total}} &= \mathbb{E}\left[\mathbf{u}^T\mathbf{u}\right] = \mathbb{E}\left[\mathbf{r}^T\boldsymbol{\alpha}^T\boldsymbol{\alpha}\mathbf{r}\right] \\ &= \mathrm{Tr}\left[\mathbb{E}\left[\mathbf{u}\mathbf{u}^T\right]\right] = \mathrm{Tr}\left[\boldsymbol{\alpha}\,\mathbb{E}\left[\mathbf{r}\mathbf{r}^T\right]\boldsymbol{\alpha}^T\right] \\ &= \mathrm{Tr}\left[\boldsymbol{\alpha}\,\mathbb{E}\left[(\boldsymbol{\theta}+\mathbf{w})(\boldsymbol{\theta}+\mathbf{w})^T\right]\boldsymbol{\alpha}^T\right] = \mathrm{Tr}\left[\boldsymbol{\alpha}\left(\mathbf{R}_{\boldsymbol{\theta}}+\mathbf{R}_{\mathbf{w}}\right)\boldsymbol{\alpha}^T\right], \end{aligned} \tag{5.4}$$

where $\mathrm{Tr}[\cdot]$ denotes the trace operation of a square matrix. Therefore, the choice of the mixing matrix $\boldsymbol{\alpha}$ affects the average cumulative transmission power of the network. Hence, determining the mixing matrix $\boldsymbol{\alpha}$ can be considered a power-allocation strategy.

---

the transmission cost of inter-sensor communications as it is a lot smaller than that of the communication between local sensors and the FC.

The model of the communication channels between local sensors and the FC is the same as the one introduced in Subsection 1.1.3. The received signal from sensor j at the FC is modeled as

$$\begin{aligned} z_j = h_j u_j + n_j &= h_j \sum_{\substack{i=1 \\ \mathbf{A}_{j,i}=1}}^{K} \alpha_{j,i} r_i + n_j \\ &= h_j \sum_{\substack{i=1 \\ \mathbf{A}_{j,i}=1}}^{K} \alpha_{j,i} \theta_i + h_j \sum_{\substack{i=1 \\ \mathbf{A}_{j,i}=1}}^{K} \alpha_{j,i} w_i + n_j \end{aligned} \qquad j = 1, 2, \ldots, M. \tag{5.5}$$

Suppose that the vector of channel noises $\mathbf{n} \stackrel{\text{def}}{=} [n_1, n_2, \ldots, n_M]^T$ is Gaussian with zero mean and $\mathbf{R_n} \stackrel{\text{def}}{=} \mathbb{E}\left[\mathbf{n}\mathbf{n}^T\right]$ as its auto-correlation matrix, i.e., $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \mathbf{R_n})$. The above model for the communication channels between local sensors and the FC can be rewritten in a vector form as

$$\mathbf{z} = \mathbf{H}\mathbf{u} + \mathbf{n} = \mathbf{H}\boldsymbol{\alpha}\mathbf{r} + \mathbf{n} = \mathbf{H}\boldsymbol{\alpha}\boldsymbol{\theta} + \mathbf{H}\boldsymbol{\alpha}\mathbf{w} + \mathbf{n}, \tag{5.6}$$

where $\mathbf{z} \stackrel{\text{def}}{=} [z_1, z_2, \ldots, z_M]^T$ is the vector of the received data from local sensors at the FC, and $\mathbf{H} = \text{diag}(h_1, h_2, \ldots, h_M)$ is a diagonal M-by-M matrix, whose mth diagonal element is the fading coefficient of the channel between sensor m and the FC. In the following analyses, we assume that the FC has *perfect* knowledge of the *instantaneous* fading coefficients of the channels between the sensors and itself. This requirement can be satisfied by, for example, using pilot signals.

5.4 Derivation of Optimal Power Allocation

It can be seen from Equation (5.6) that, due to the independence of $\mathbf{w}$ and $\mathbf{n}$, given a realization of the vector of locally observed signals $\boldsymbol{\theta}$ and a realization of the fading coefficients of the channels between local sensors and the FC (i.e., $\mathbf{H}$), the received vector of signals at the FC (i.e., $\mathbf{z}$) is a Gaussian random vector with mean $\boldsymbol{\mu}_{\mathbf{z}|\{\boldsymbol{\theta},\mathbf{H}\}} = \mathbf{H}\boldsymbol{\alpha}\boldsymbol{\theta}$ and covariance matrix $\mathbf{R}_{\mathbf{z}|\{\boldsymbol{\theta},\mathbf{H}\}} = \mathbf{H}\boldsymbol{\alpha}\mathbf{R}_{\mathbf{w}}\boldsymbol{\alpha}^T\mathbf{H}^T + \mathbf{R}_{\mathbf{n}}$. In other words,

$$\mathbf{z} \Big| \{\boldsymbol{\theta}, \mathbf{H}\} \sim \mathcal{N}\left(\mathbf{H}\boldsymbol{\alpha}\boldsymbol{\theta}, \mathbf{H}\boldsymbol{\alpha}\mathbf{R}_{\mathbf{w}}\boldsymbol{\alpha}^T\mathbf{H}^T + \mathbf{R}_{\mathbf{n}}\right).$$

Upon receiving the faded and noisy version of the vector of locally processed observations, the FC finds the BLUE estimator for the vector of observed signals $\boldsymbol{\theta}$ as follows [84, Chapter 6]:

$$\widehat{\boldsymbol{\theta}} = \left(\boldsymbol{\alpha}^T\mathbf{H}^T\left(\mathbf{H}\boldsymbol{\alpha}\mathbf{R_w}\boldsymbol{\alpha}^T\mathbf{H}^T + \mathbf{R_n}\right)^{-1}\mathbf{H}\boldsymbol{\alpha}\right)^{-1}\boldsymbol{\alpha}^T\mathbf{H}^T\left(\mathbf{H}\boldsymbol{\alpha}\mathbf{R_w}\boldsymbol{\alpha}^T\mathbf{H}^T + \mathbf{R_n}\right)^{-1}\mathbf{z}, \tag{5.7}$$

where the corresponding covariance matrix of the BLUE estimator is found as

$$\begin{aligned}\mathbf{R}_{\widehat{\boldsymbol{\theta}}} &= \mathbb{E}\left[\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)^T\right] \\ &= \left(\boldsymbol{\alpha}^T\mathbf{H}^T\left(\mathbf{H}\boldsymbol{\alpha}\mathbf{R_w}\boldsymbol{\alpha}^T\mathbf{H}^T + \mathbf{R_n}\right)^{-1}\mathbf{H}\boldsymbol{\alpha}\right)^{-1}.\end{aligned} \tag{5.8}$$

Note that the calculation of the BLUE estimator is independent of the statistics of the signal to be estimated $\boldsymbol{\theta}$. As it can readily be observed from Equation (5.8), the choice of the mixing matrix $\boldsymbol{\alpha}$ affects the estimation distortion at the FC, which can be defined based on the given covariance matrix of the BLUE estimator.

The goal of this chapter is to derive an optimal mixing matrix $\boldsymbol{\alpha}$ that minimizes the total distortion in the estimation of $\boldsymbol{\theta}$ at the FC, given a constraint on the average cumulative transmission power of the sensors. We define the total estimation distortion at the FC as the "trace of the covariance matrix of the BLUE estimator," which is the sum of the estimation variances for different components of $\boldsymbol{\theta}$. This objective can be formulated as the following optimization problem:

$$\begin{aligned}\underset{\boldsymbol{\alpha}}{\text{minimize}} \quad & \operatorname{Tr}\left[\boldsymbol{\alpha}^T\mathbf{H}^T\left(\mathbf{H}\boldsymbol{\alpha}\mathbf{R_w}\boldsymbol{\alpha}^T\mathbf{H}^T + \mathbf{R_n}\right)^{-1}\mathbf{H}\boldsymbol{\alpha}\right]^{-1} \\ \text{subject to} \quad & \operatorname{Tr}\left[\boldsymbol{\alpha}\left(\mathbf{R}_{\boldsymbol{\theta}} + \mathbf{R_w}\right)\boldsymbol{\alpha}^T\right] \leq P_{\mathsf{T}}\end{aligned} \tag{5.9}$$

where P_{T} is the constraint on the total average transmission power in the entire network. The following lemma will be used to simplify the objective function of the above constrained optimization problem.

Lemma 5.4.1. *A lower bound on* $\operatorname{Tr}\left[\mathbf{R}_{\widehat{\boldsymbol{\theta}}}\right]$ *can be found as*

$$\operatorname{Tr}\left[\mathbf{R}_{\widehat{\boldsymbol{\theta}}}\right] \geq \frac{K^2}{\operatorname{Tr}\left[\boldsymbol{\alpha}^T\mathbf{H}^T\left(\mathbf{H}\boldsymbol{\alpha}\mathbf{R_w}\boldsymbol{\alpha}^T\mathbf{H}^T + \mathbf{R_n}\right)^{-1}\mathbf{H}\boldsymbol{\alpha}\right]}. \tag{5.10}$$

Proof. It is proved in [88, Lemma 1] that for any arbitrary real matrix $\boldsymbol{\Phi}$ and any positive semi-definite real matrix $\boldsymbol{\Lambda}$ of proper sizes, the following inequality holds:

$$\operatorname{Tr}\left[\boldsymbol{\Phi}^T\boldsymbol{\Lambda}^{-1}\boldsymbol{\Phi}\right] \geq \frac{\left(\operatorname{Tr}\left[\boldsymbol{\Phi}^T\boldsymbol{\Phi}\right]\right)^2}{\operatorname{Tr}\left[\boldsymbol{\Phi}^T\boldsymbol{\Lambda}\boldsymbol{\Phi}\right]}.$$

Let $\mathbf{\Phi} \stackrel{\text{def}}{=} \mathbf{I}_K$ and $\mathbf{\Lambda} \stackrel{\text{def}}{=} \mathbf{R}_{\widehat{\boldsymbol{\theta}}}$, where $\mathbf{I}_K$ denotes the K-by-K identity matrix. The lower bound of the lemma is readily derived. □

Using the result of Lemma 5.4.1, the optimization problem of Equation (5.9) can be rewritten as follows:

$$\begin{aligned} \underset{\boldsymbol{\alpha}}{\text{maximize}} \quad & \text{Tr}\left[\boldsymbol{\alpha}^T\mathbf{H}^T\left(\mathbf{H}\boldsymbol{\alpha}\mathbf{R}_{\mathbf{w}}\boldsymbol{\alpha}^T\mathbf{H}^T+\mathbf{R}_{\mathbf{n}}\right)^{-1}\mathbf{H}\boldsymbol{\alpha}\right] \\ \text{subject to} \quad & \text{Tr}\left[\boldsymbol{\alpha}\left(\mathbf{R}_{\boldsymbol{\theta}}+\mathbf{R}_{\mathbf{w}}\right)\boldsymbol{\alpha}^T\right] \leq P_{\mathsf{T}} \end{aligned} \tag{5.11}$$

Lemma 5.4.2. *The optimization problem given in Equation* (5.11) *is equivalent to the following form:*

$$\begin{aligned} \underset{\boldsymbol{\alpha},\gamma,\mathbf{\Gamma}}{\text{minimize}} \quad & \gamma \\ \text{subject to} \quad & \text{Tr}\left[\boldsymbol{\alpha}\left(\mathbf{R}_{\boldsymbol{\theta}}+\mathbf{R}_{\mathbf{w}}\right)\boldsymbol{\alpha}^T\right] \leq P_{\mathsf{T}} \\ & \begin{pmatrix} \mathbf{\Gamma} & \mathbf{R}_{\mathbf{w}}^{-1} \\ \mathbf{R}_{\mathbf{w}}^{-1} & \boldsymbol{\alpha}^T\mathbf{H}^T\mathbf{R}_{\mathbf{n}}^{-1}\mathbf{H}\boldsymbol{\alpha}+\mathbf{R}_{\mathbf{w}}^{-1} \end{pmatrix} \succeq 0 \\ & \text{Tr}\left[\mathbf{\Gamma}\right] \leq \gamma \end{aligned} \tag{5.12}$$

where γ is a real scalar, $\mathbf{\Gamma}$ is a symmetric K-by-K real matrix, and $\mathbf{\Upsilon} \succeq 0$ denotes that the matrix $\mathbf{\Upsilon}$ is positive semi-definite.

Proof. Based on the Woodbury matrix inversion lemma [91, Page 19], for any arbitrary matrix $\mathbf{\Upsilon}$ and any non-singular matrices $\mathbf{\Phi}$ and $\mathbf{\Lambda}$ of proper sizes, if the matrix $\mathbf{\Phi}+\mathbf{\Upsilon}\mathbf{\Lambda}\mathbf{\Upsilon}^T$ is non-singular, then the following identity holds:

$$\left(\mathbf{\Phi}+\mathbf{\Upsilon}\mathbf{\Lambda}\mathbf{\Upsilon}^T\right)^{-1} = \mathbf{\Phi}^{-1} - \mathbf{\Phi}^{-1}\mathbf{\Upsilon}\left(\mathbf{\Lambda}^{-1}+\mathbf{\Upsilon}^T\mathbf{\Phi}^{-1}\mathbf{\Upsilon}\right)^{-1}\mathbf{\Upsilon}^T\mathbf{\Phi}^{-1}.$$

Let $\mathbf{\Phi} \stackrel{\text{def}}{=} \mathbf{R}_{\mathbf{w}}^{-1}$, $\mathbf{\Lambda} \stackrel{\text{def}}{=} \mathbf{R}_{\mathbf{n}}^{-1}$, and $\mathbf{\Upsilon} \stackrel{\text{def}}{=} \boldsymbol{\alpha}^T\mathbf{H}^T$. Using the above matrix identity, the argument of the trace operation in the objective function of Equation (5.11) can be simplified as

$$\boldsymbol{\alpha}^T\mathbf{H}^T\left(\mathbf{H}\boldsymbol{\alpha}\mathbf{R}_{\mathbf{w}}\boldsymbol{\alpha}^T\mathbf{H}^T+\mathbf{R}_{\mathbf{n}}\right)^{-1}\mathbf{H}\boldsymbol{\alpha} = \mathbf{R}_{\mathbf{w}}^{-1} - \mathbf{R}_{\mathbf{w}}^{-1}\left(\boldsymbol{\alpha}^T\mathbf{H}^T\mathbf{R}_{\mathbf{n}}^{-1}\mathbf{H}\boldsymbol{\alpha}+\mathbf{R}_{\mathbf{w}}^{-1}\right)^{-1}\mathbf{R}_{\mathbf{w}}^{-1}.$$

Hence, the optimization problem defined in Equation (5.11) can be rewritten as

$$\begin{aligned} \underset{\boldsymbol{\alpha}}{\text{minimize}} \quad & \text{Tr}\left[\mathbf{R}_{\mathbf{w}}^{-1}\left(\boldsymbol{\alpha}^T\mathbf{H}^T\mathbf{R}_{\mathbf{n}}^{-1}\mathbf{H}\boldsymbol{\alpha}+\mathbf{R}_{\mathbf{w}}^{-1}\right)^{-1}\mathbf{R}_{\mathbf{w}}^{-1}\right] \\ \text{subject to} \quad & \text{Tr}\left[\boldsymbol{\alpha}\left(\mathbf{R}_{\boldsymbol{\theta}}+\mathbf{R}_{\mathbf{w}}\right)\boldsymbol{\alpha}^T\right] \leq P_{\mathsf{T}} \end{aligned} \tag{5.13}$$

Let γ be a real scalar such that for any mixing matrix $\boldsymbol{\alpha}$, the following inequality holds:

$$\mathrm{Tr}\left[\mathbf{R}_{\mathbf{w}}^{-1}\left(\boldsymbol{\alpha}^T\mathbf{H}^T\mathbf{R}_{\mathbf{n}}^{-1}\mathbf{H}\boldsymbol{\alpha}+\mathbf{R}_{\mathbf{w}}^{-1}\right)^{-1}\mathbf{R}_{\mathbf{w}}^{-1}\right]\leq\gamma. \tag{5.14}$$

There exists a symmetric K-by-K real matrix $\boldsymbol{\Gamma}$ such that [89]

$$\mathbf{R}_{\mathbf{w}}^{-1}\left(\boldsymbol{\alpha}^T\mathbf{H}^T\mathbf{R}_{\mathbf{n}}^{-1}\mathbf{H}\boldsymbol{\alpha}+\mathbf{R}_{\mathbf{w}}^{-1}\right)^{-1}\mathbf{R}_{\mathbf{w}}^{-1}\preceq\boldsymbol{\Gamma} \tag{5.15a}$$

$$\text{and}\quad \mathrm{Tr}\left[\boldsymbol{\Gamma}\right]\leq\gamma \tag{5.15b}$$

where $\boldsymbol{\Phi}\preceq\boldsymbol{\Lambda}$ means that the matrix $\boldsymbol{\Lambda}-\boldsymbol{\Phi}$ is positive semi-definite, denoted as $\boldsymbol{\Lambda}-\boldsymbol{\Phi}\succeq 0$. In other words,

$$\boldsymbol{\Gamma}-\mathbf{R}_{\mathbf{w}}^{-1}\left(\boldsymbol{\alpha}^T\mathbf{H}^T\mathbf{R}_{\mathbf{n}}^{-1}\mathbf{H}\boldsymbol{\alpha}+\mathbf{R}_{\mathbf{w}}^{-1}\right)^{-1}\mathbf{R}_{\mathbf{w}}^{-1}\succeq 0. \tag{5.16}$$

Based on the Schur's complement theorem [91, Page 472], for any arbitrary matrix $\boldsymbol{\Upsilon}$ and any symmetric matrices $\boldsymbol{\Phi}$ and $\boldsymbol{\Lambda}$ of proper sizes, if $\boldsymbol{\Lambda}$ is invertible and $\boldsymbol{\Lambda}\succ 0$, then $\boldsymbol{\Phi}-\boldsymbol{\Upsilon}\boldsymbol{\Lambda}^{-1}\boldsymbol{\Upsilon}^T\succeq 0$ if and only if

$$\begin{pmatrix}\boldsymbol{\Phi} & \boldsymbol{\Upsilon}\\ \boldsymbol{\Upsilon}^T & \boldsymbol{\Lambda}\end{pmatrix}\succeq 0,$$

where $\boldsymbol{\Lambda}\succ 0$ means that the matrix $\boldsymbol{\Lambda}$ is positive definite. Let $\boldsymbol{\Phi}\overset{\text{def}}{=}\boldsymbol{\Gamma}$, $\boldsymbol{\Upsilon}\overset{\text{def}}{=}\mathbf{R}_{\mathbf{w}}^{-1}$, and $\boldsymbol{\Lambda}\overset{\text{def}}{=}\boldsymbol{\alpha}^T\mathbf{H}^T\mathbf{R}_{\mathbf{n}}^{-1}\mathbf{H}\boldsymbol{\alpha}+\mathbf{R}_{\mathbf{w}}^{-1}$. Note that any real symmetric matrix $\boldsymbol{\Lambda}$ is positive semidefinite if and only if it can be factored as $\boldsymbol{\Lambda}\overset{\text{def}}{=}\boldsymbol{\Psi}\boldsymbol{\Psi}^T$, where $\boldsymbol{\Psi}$ is an arbitrary matrix [91]. Using the Schur's complement, the condition shown in Equation (5.16) is equivalent to the following matrix being positive semi-definite:

$$\begin{pmatrix}\boldsymbol{\Gamma} & \mathbf{R}_{\mathbf{w}}^{-1}\\ \mathbf{R}_{\mathbf{w}}^{-1} & \boldsymbol{\alpha}^T\mathbf{H}^T\mathbf{R}_{\mathbf{n}}^{-1}\mathbf{H}\boldsymbol{\alpha}+\mathbf{R}_{\mathbf{w}}^{-1}\end{pmatrix}\succeq 0.$$

Based on the above discussions, the optimization problem given in Equation (5.13) is equivalent to the constrained optimization problem defined in Equation (5.12), and the proof of Lemma 5.4.2 is concluded. □

The constrained optimization problem defined in Equation (5.12) is a *linear programming with bi-linear matrix-inequality constraints*. It can efficiently be solved using numerical solvers such as PENBMI [92], which is fully integrated within the MATLAB® environment through version 3.0 of the YALMIP interface library [93].

5.5 Numerical Analysis

In this section, the numerical results are presented to show the effect of spatial collaboration among sensors on the estimation performance at the FC of a WSN. Suppose that $K = 6$ sensors are randomly and uniformly distributed in the two-dimensional rectangle of $[-10, 10] \times [-5, 5]$, where $\times$ denotes the Cartesian product of two sets. It is assumed that all sensors are connected to the FC, i.e., $M = K$. Let the covariance between the signals observed by sensors i and j be defined as

$$\mathbf{R}_{\boldsymbol{\theta}_{i,j}} \stackrel{\text{def}}{=} \mathbb{E}\left[\theta_i \theta_j\right] \stackrel{\text{def}}{=} \sigma_\theta^2 \, \rho_{i,j}, \qquad i, j = 1, 2, \ldots, K, \tag{5.17}$$

where σ_θ^2 is the variance of each component of the vector of signals to be estimated $\boldsymbol{\theta}$, and $\rho_{i,j}$ is the inter-sensor correlation coefficient that monotonically decreases with the increase of the distance between sensors as

$$\rho_{i,j} \stackrel{\text{def}}{=} e^{-\left(\frac{d_{i,j}}{\kappa_1}\right)^{\kappa_2}}, \qquad i, j = 1, 2, \ldots, K, \tag{5.18}$$

where $d_{i,j}$ is the distance between sensors i and j, $\kappa_1 > 0$ is the normalizing factor of the distances, and $0 < \kappa_2 \leq 2$ controls the rate of the decay of the correlation coefficients with distance. Note that $\rho_{i,i} = 1$, $i = 1, 2, \ldots, K$. Assume that the vectors of observation noise $\mathbf{w}$ and channel noise $\mathbf{n}$ are homogeneous and equi-correlated with their covariance matrix defined as

$$\mathbf{R}_{\mathbf{w}} = \sigma_w^2 \left[(1 - \lambda_w)\, \mathbf{I}_K + \lambda_w \mathbf{1}\mathbf{1}^T\right] \tag{5.19a}$$

$$\mathbf{R}_{\mathbf{n}} = \sigma_n^2 \left[(1 - \lambda_n)\, \mathbf{I}_M + \lambda_n \mathbf{1}\mathbf{1}^T\right] \tag{5.19b}$$

where σ_w^2 and σ_n^2 are the variances of each component of the vectors of observation noise $\mathbf{w}$ and channel noise $\mathbf{n}$, respectively, λ_w and λ_n are the constant correlation coefficients between each pair of distinct components of $\mathbf{w}$ and $\mathbf{n}$, respectively, and $\mathbf{1}$ is a column vector of all ones with appropriate length. To generate our numerical results, the coefficients of the communication channels between local sensors and the FC are assumed to be unit, i.e., $h_j = 1$, $j = 1, 2, \ldots, M$. The following values are used for the parameters of the system to generate the numerical results presented in this section: $\sigma_\theta^2 = 1$, $\kappa_1 = 6$, $\kappa_2 = 3$, $\sigma_w^2 = 0.1$, $\sigma_n^2 = 0.01$, and $\lambda_w = \lambda_n = 0.1$.

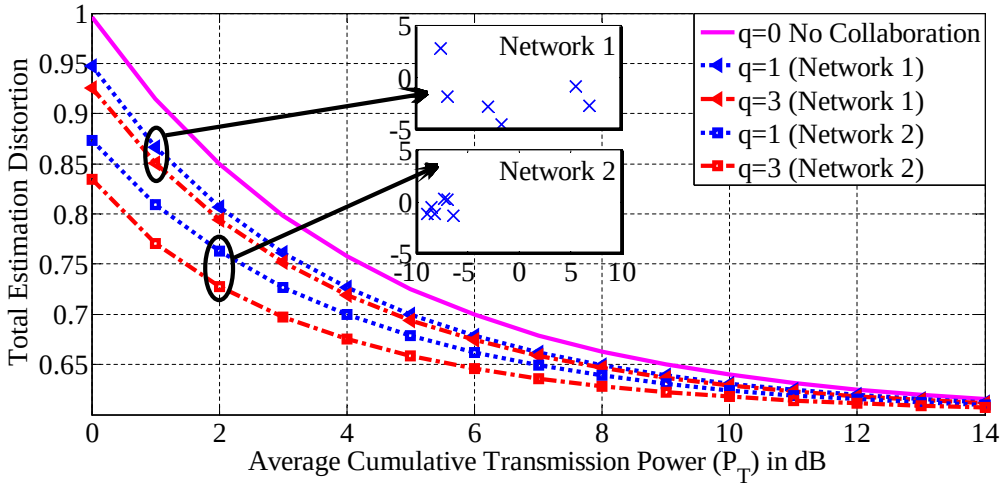


Figure 5.2: Total estimation distortion at the FC versus the average cumulative transmission power for different degrees of spatial collaboration within two random network realizations.

Figure 5.2 shows the total estimation distortion at the FC, as defined by the objective function of the optimization problem in Equation (5.9), versus the total average transmission power in the entire network P_{T} for two network realizations. Each sensor collaborates with its q closest neighbors by sharing its local noisy observations with them through error-free, low-cost links. Note that $q = 0$ represents a network without any spatial collaboration, and $q = K - 1$ corresponds to a network with full spatial collaboration. As evident from this figure, even moderate collaboration among sensors decreases the estimation distortion at the FC. The *collaboration gain* is more significant when the signals to be estimated have a higher correlation, i.e., when the sensors observing them are located more closely, as depicted in Network 2.

5.6 Conclusions

In this chapter, we studied the effect of spatial collaboration on the performance of the BLUE estimator at the FC of a WSN that tries to estimate the vector of spatially correlated signals observed by sensors, rather than the well-studied case of estimating a common signal observed by a collection of sensors. An optimal linear spatial-collaboration scheme was derived that minimizes the sum of the estimation variances of different signals observed by

the network, given a constraint on the average cumulative transmission power in the entire network. The numerical results showed that even a small degree of connectivity and spatial collaboration in the network improves the quality of the estimators at the FC, and that the collaboration gain is more significant when the signals observed by sensors have a higher spatial correlation.

Chapter 6

Effects of Spatial Randomness on Source Localization with Distributed Sensors

6.1 Introduction

The problem of energy-based source localization using a set of spatially distributed, randomly located, limited-power sensors forming a wireless sensor network (WSN) has recently attracted a lot of attention in the research community [28, 29, 81, 94–98]. An effective source localization can be a first step in a broad range of other applications such as navigation, tracking, and geographic routing. In this context, the sensors make noisy observations of the energy transmitted by, for example, an RF or acoustic source at their locations, process their noisy observations locally by, for instance, quantizing them, and send their processed data to a central entity in the network, known as the fusion center (FC), for further processing. The FC will then combine the received signals from the sensors, which are potentially corrupted by the communication channels between the sensors and itself, to estimate the location of the energy-transmitting source. As is common in the literature, it is reasonable to assume that the locations of the sensors are known at the FC, which can be achieved using any form of cooperative localization schemes (e.g., [99–103]).

The analyses and performance assessments in most of the works proposed in the literature

for source localization can easily be generalized to a generic case in which the sensors are *randomly* located within the surveillance region covered by the network. Of course, the realization of the network geometry after its deployment should be known at the FC. However, the results of the performance analysis are usually presented for a fixed network topology such as a regular grid deployment [29] or for an average behavior of a number of random network realizations [96]. To the best of our knowledge, the effect of randomness of the sensor placement on the performance of source-localization schemes has been relatively unexplored, beyond analyzing the network's average behavior [104, 105]. Srinivasa and Haenggi [104] have considered the problem of distributed estimation of the path-loss exponent in an environment in which an RF signal is broadcast, where the sensors are distributed according to a Poisson point process and sensor transmissions can interfere with each other.

The goal of this chapter is to assess the performance of a typical source-localization scheme under different scenarios of random network realizations using numerical simulations. In other words, the question that we are trying to answer is as follows: Given a specific localization scheme, how does a randomly deployed WSN within a fixed surveillance region perform in terms of the localization accuracy? Note that answering this question gives significantly more insight into the design of a network than predicting only the *average* behavior of a randomly deployed system. Therefore, we are not proposing a new localization scheme, but rather we are applying concepts from stochastic geometry and point processes [106–109] to investigate the performance of a refined and special version of a recently proposed source-localization algorithm [29]. A novel performance measure called the *localization outage* will be introduced to assess the performance of a typical localization algorithm. Numerical methods will be used to determine what parameters affect the performance of the given localization scheme when the sensors are placed according to a binomial point process with repulsion, which is also known as a uniform clustering process. The results of this analysis can be used to guide network deployment. If these guidelines are followed, a randomly formed network can be guaranteed (with some confidence) to achieve a minimum performance in terms of the localization accuracy.

The rest of this chapter is organized as follows: Section 6.2 describes the system model considered in our analyses. In Section 6.3, the source-localization scheme proposed in [29] is

summarized and the Cramér-Rao lower bound (CRLB) for the location estimator based on the binary quantized data at the sensors is derived. The effects of the random realization of the network geometry on the aforementioned localization scheme are shown through numerical simulations in Section 6.4. Section 6.5 introduces the concept of localization outage for a random network realization and discusses the effects of exclusion zones around sensors on the performance of an arbitrary random realization of the network geometry. Finally, the chapter is concluded in Section 6.6.

6.2 System Model

Suppose that a WSN is composed of a FC and K sensors *arbitrarily* located in the two-dimensional space $\mathbb{R}^2$ within a circular surveillance region $\mathcal{S} \subseteq \mathbb{R}^2$ with radius R and spatially distributed according to any point process. Assume that a point source located at $(x_{\mathsf{T}}, y_{\mathsf{T}}) \in \mathcal{S}$ emits energy omni-directionally and that its power received by an arbitrary sensor i located at $(x_i, y_i) \in \mathcal{S}$ is

$$P_i = P_0 \left(\frac{d_0}{d_i}\right)^{\alpha}, \qquad d_i \geq d_0 \text{ and } i = 1, 2, \ldots, K, \tag{6.1}$$

where P_0 is the received power from the source at the reference distance d_0, α is the power-decay exponent, and d_i is the distance between the source and sensor i defined as

$$d_i = \sqrt{(x_{\mathsf{T}} - x_i)^2 + (y_{\mathsf{T}} - y_i)^2}, \qquad i = 1, 2, \ldots, K.$$

An example of the random realization of such network topology is shown in Figure 6.1, where $K = 50$ sensors are randomly distributed in a circular region with radius $R = 50$. Other parameters shown in the figure are introduced later in this chapter. It should be mentioned that in addition to RF point sources, one of the most well-studied sources that satisfies the above power-decay model is the acoustic source, whose localization has widely been studied in the literature [95].

Let $\boldsymbol{\theta} \stackrel{\text{def}}{=} [P_0, x_{\mathsf{T}}, y_{\mathsf{T}}]^T$ denote the vector of deterministic parameters associated with the source, where $[\cdot]^T$ represents the transpose operation. The ultimate goal of the WSN is to estimate these parameters. More specifically, the focus of this chapter is on the estimation

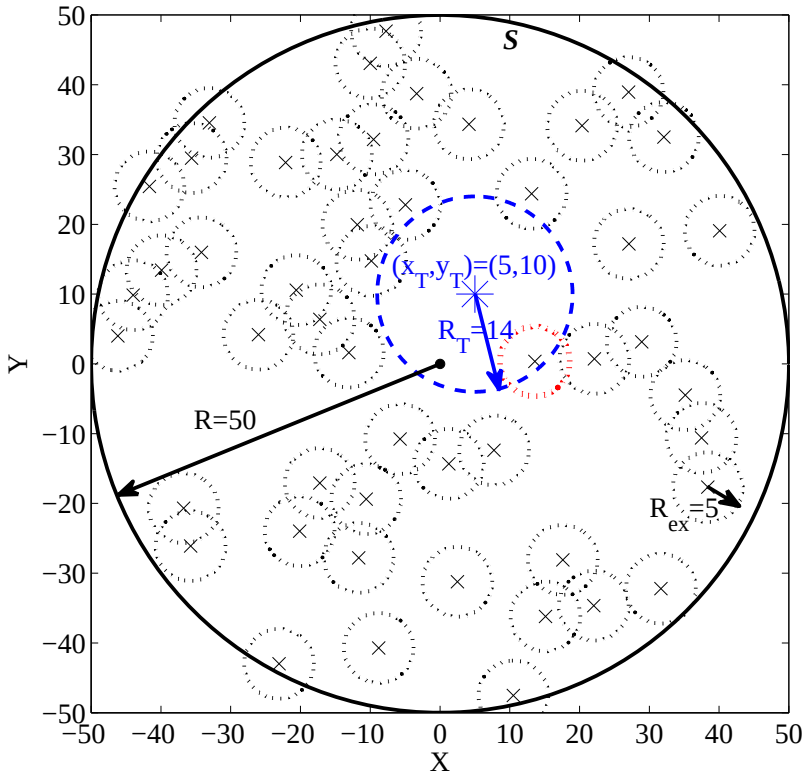


Figure 6.1: The network topology of an example WSN consisting of $K = 50$ sensors denoted by '$\times$', whose objective is to localize a source target denoted by '$*$' and located at $(x_{\mathsf{T}}, y_{\mathsf{T}}) = (5, 10)$. The sensors are randomly placed in the circular surveillance region with radius $R = 50$ and centered at the origin according to a uniform clustering process. Each sensor is surrounded by an exclusion zone with radius $R_{\mathsf{ex}} = 5$, shown by a dotted circle around the sensor. A dashed circle with radius $R_{\mathsf{T}} = 14$ is depicted around the source target, within which there is $K_{\mathsf{T}} = 1$ sensor enclosed.

of the source location. Figure 6.2 shows the functional diagram of the WSN. Assume that the ith sensor observes a noisy version of the received power of the broadcast signal at its location. Therefore, the observation function $\Xi_i(\cdot)$ in Equation (1.1) is defined as

$$\Xi_i(\boldsymbol{\theta}) \stackrel{\text{def}}{=} \sqrt{P_i}.$$

Based on the above power-decay model, the local noisy observation at the ith sensor is found as

$$r_i = \sqrt{P_i} + w_i, \qquad i = 1, 2, \ldots, K. \tag{6.2}$$

Note that each sensor can make N consecutive observations over a specific time interval T

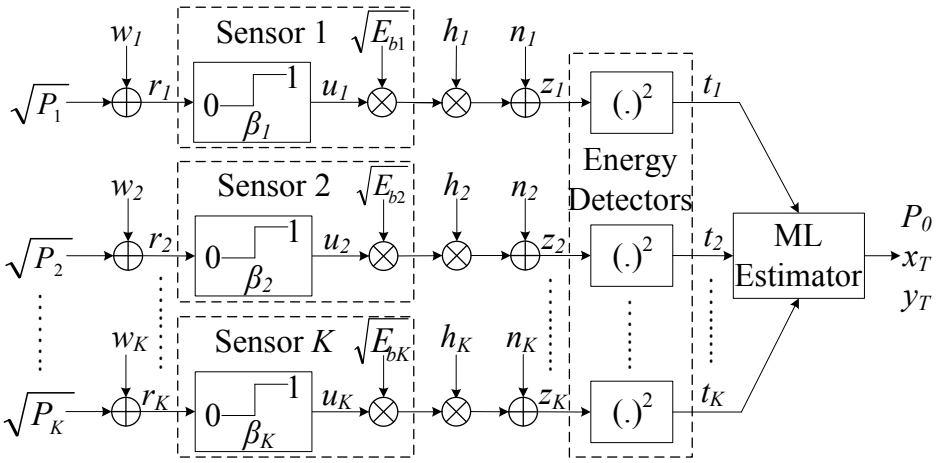


Figure 6.2: Functional system model of a WSN in which the FC localizes a source of energy.

and average them to find the observed signal as shown in Equation (6.2). The time averaging of the local observations results in an observation noise with smaller variance. We define the *observation signal-to-noise ratio* (SNR) at sensor i as $\psi_i \stackrel{\text{def}}{=} \frac{P_0}{\sigma_i^2}$, $i = 1, 2, \ldots, K$. Upon observing the received noisy signal, each sensor uses a binary quantization scheme to quantize its local observation as

$$u_i = \begin{cases} 0, & \text{if } r_i < \beta_i \\ 1, & \text{if } r_i \geq \beta_i \end{cases} \qquad i = 1, 2, \ldots, K, \tag{6.3}$$

where β_i is the binary quantization threshold at sensor i. This is a special case of the digital local processing scheme introduced in Subsection 1.1.2 under Equation (1.3). Note that the sensors can process their noisy observations using various processing schemes. The simple binary quantization method considered as an example does not limit the generality of the following discussions and has only been used to emphasize the main objective of this chapter, which is to study how spatial randomness affects the performance of a typical localization scheme.

The model of the communication channels between the sensors and the FC is the same as the one introduced in Subsection 1.1.3. More specifically, each sensor will use an on-off keying (OOK) scheme to send its quantized data to the FC through orthogonal channels corrupted by fading and additive white Gaussian noise (AWGN). The received signal from

sensor i at the FC is

$$z_i = h_i \sqrt{\mathcal{E}_{b_i}}\, u_i + n_i, \qquad i = 1, 2, \ldots, K, \tag{6.4}$$

where $\sqrt{\mathcal{E}_{b_i}} u_i$ is transmitted by the ith sensor, h_i is the multiplicative fading coefficient of the channel between sensor i and the FC, and n_i is the spatially independent, zero-mean, complex Gaussian random variable with variance τ_i^2, i.e., $n_i \sim \mathcal{CN}(0, \tau_i^2)$. In this chapter, it is assumed that the channels between the sensors and the FC experience Rayleigh fading and therefore, the random variable h_i is assumed to be spatially independent, zero-mean, complex Gaussian with unit power, i.e., $h_i \sim \mathcal{CN}(0, 1)$. It is also assumed that the FC does *not* have access to the instantaneous channel coefficients and that it only knows their distribution. We define the *channel SNR* for sensor i as $\eta_i \stackrel{\text{def}}{=} \frac{\mathcal{E}_{b_i}}{\tau_i^2}$, $i = 1, 2, \ldots, K$.

Upon receiving the signal from sensor i, the FC finds the energy of the received signal as $t_i = |z_i|^2$, $i = 1, 2, \ldots, K$, where $|\cdot|$ denotes the absolute-value operation. Having access to $\mathbf{t} \stackrel{\text{def}}{=} [t_1, t_2, \ldots, t_K]^T$, the FC finds the maximum likelihood (ML) estimate of the vector of unknown parameters $\boldsymbol{\theta}$ as explained in the following section.

6.3 Derivation of ML Estimator and CRLB

As mentioned in the previous section, the sensors are *arbitrarily* located in the surveillance region $\mathcal{S}$. However, it is assumed that the FC knows their exact locations after the deployment of the WSN. This assumption can in practice be satisfied using any localization scheme [99–103]. Note that the method and criteria for the localization of distributed sensors can in general be quite different from those of a single energy-emitting source, as considered in this chapter.

Let Ω be a variable denoting a realization of the network geometry, when the WSN is deployed and the set of $\{(x_i, y_i)\}_{i=1}^K$ (and consequently, the set of $\{d_i\}_{i=1}^K$) is fixed. It is intuitive that the performance of any source-localization scheme, including the one studied in this chapter, depends on the specific realization of the network geometry. The main goal of this chapter is to study the effect of variable Ω as the network geometry on the performance of a source-localization scheme similar to the one proposed in [29]. In the rest of this section,

the ML estimator and its corresponding CRLB proposed in [29] are summarized in order to assess the effect of network geometry on their performance.

6.3.1 Derivation of ML Estimator

Based on the observation model introduced in Equation (6.2) and the binary quantization rule specified in Equation (6.3), the probability density function (pdf) of each sensor's quantized data parameterized by the vector of unknown parameters to be estimated, given a realization of the network geometry Ω, is found as

$$f_{U_i|\Omega}\left(u_i:\boldsymbol{\theta}|\Omega\right)=Q\left(\frac{\sqrt{P_i}-\beta_i}{\sigma_i}\right)\delta\left[u_i\right]+Q\left(\frac{\beta_i-\sqrt{P_i}}{\sigma_i}\right)\delta\left[u_i-1\right], \tag{6.5}$$

where $\delta\left[\cdot\right]$ denotes the discrete Dirac delta function, and $Q(\cdot)$ is the complementary cumulative distribution function (CCDF) of the standard Gaussian random variable defined as

$$Q(x)\stackrel{\text{def}}{=}\frac{1}{\sqrt{2\pi}}\int_x^{\infty}e^{-\frac{t^2}{2}}\,\mathrm{d}t.$$

Based on the channel model introduced in Equation (6.4), given any binary sensor decision u_i, the signal received from sensor i at the FC is a complex Gaussian random variable with zero mean and variance $\mathcal{E}_{b_i}u_i^2+\tau_i^2$, i.e., $z_i|u_i\sim\mathcal{CN}\left(0,\mathcal{E}_{b_i}u_i^2+\tau_i^2\right)$. Note that the channel fading coefficient and the channel AWGN are assumed to be independent. Based on this result, the energy of the received signal from sensor i at the FC, given the sensor's binary decision, is exponentially distributed with parameter $\zeta_i\stackrel{\text{def}}{=}\frac{1}{\mathcal{E}_{b_i}u_i^2+\tau_i^2}$, i.e., $t_i|u_i\sim\mathcal{E}\left(\frac{1}{\mathcal{E}_{b_i}u_i^2+\tau_i^2}\right)$. Therefore, given a realization of the network geometry Ω, the joint pdf of the vector of received energies from different sensors at the FC, parameterized by the vector of unknown parameters to be estimated can be written as

$$f_{\mathbf{T}|\Omega}\left(\mathbf{t}:\boldsymbol{\theta}|\Omega\right)=\prod_{i=1}^{K}f_{T_i|\Omega}\left(t_i:\boldsymbol{\theta}|\Omega\right), \tag{6.6}$$

where

$$\begin{aligned}f_{T_i|\Omega}\left(t_i:\boldsymbol{\theta}|\Omega\right)&=\int f_{T_i|U_i}\left(t_i|u_i\right)f_{U_i|\Omega}\left(u_i:\boldsymbol{\theta}|\Omega\right)\,\mathrm{d}u_i\\&\stackrel{(a)}{=}\frac{1}{\tau_i^2}e^{-\frac{t_i}{\tau_i^2}}Q\left(\frac{\sqrt{P_i}-\beta_i}{\sigma_i}\right)+\frac{1}{\mathcal{E}_{b_i}+\tau_i^2}e^{-\frac{t_i}{\mathcal{E}_{b_i}+\tau_i^2}}Q\left(\frac{\beta_i-\sqrt{P_i}}{\sigma_i}\right)\end{aligned} \tag{6.7}$$

where (a) is based on the sifting property of the Dirac delta function. It is well known that the ML estimate of the vector of unknown parameters at the FC using the vector of the received energies from local sensors is found as [84, Chapter 7]

$$\widehat{\boldsymbol{\theta}}_{\Omega} = \arg\max_{\boldsymbol{\theta}} \ \ln\left(f_{\mathbf{T}|\Omega}\left(\mathbf{t}:\boldsymbol{\theta}|\Omega\right)\right), \tag{6.8}$$

where the subscript Ω for the ML estimator signifies that it depends on the realization of the network geometry.

6.3.2 Derivation of CRLB

The performance of any estimator can be quantified by its variance. The CRLB expresses a lower bound on the variance of any *unbiased* estimator $\widehat{\boldsymbol{\theta}}_{\Omega}$ as [84, Chapter 3]

$$\mathbb{E}\left[\left(\widehat{\boldsymbol{\theta}}_{\Omega}-\boldsymbol{\theta}\right)\left(\widehat{\boldsymbol{\theta}}_{\Omega}-\boldsymbol{\theta}\right)^{T}\right] \succeq \boldsymbol{\mathcal{I}}_{\Omega}^{-1}\left(\boldsymbol{\theta}\right), \tag{6.9}$$

where $\mathbb{E}\left[\cdot\right]$ represents the expectation operation with respect to the joint pdf of the vector of received energies from different sensors at the FC, $\boldsymbol{\Phi} \succeq \boldsymbol{\Lambda}$ means that the matrix $\boldsymbol{\Phi}-\boldsymbol{\Lambda}$ is positive semi-definite, and $\boldsymbol{\mathcal{I}}_{\Omega}\left(\boldsymbol{\theta}\right)$ denotes the Fisher information matrix (FIM) for the given realization of the network geometry Ω, whose element in row m and column n is defined as

$$\left[\boldsymbol{\mathcal{I}}_{\Omega}\left(\boldsymbol{\theta}\right)\right]_{m,n} = -\mathbb{E}\left[\frac{\partial^2 \ln\left(f_{\mathbf{T}|\Omega}\left(\mathbf{t}:\boldsymbol{\theta}|\Omega\right)\right)}{\partial\theta_m \partial\theta_n}\right].$$

Based on the joint pdf of the vector of received energies from different sensors at the FC defined in Equations (6.6)–(6.7), the FIM for the given observation and channel models and given a realization of the network geometry Ω is found as follows [29]:

$$\boldsymbol{\mathcal{I}}_{\Omega}\left(\boldsymbol{\theta}\right) = \sum_{i=1}^{K} \frac{\boldsymbol{\mathcal{G}}_{i,\Omega}\left(\boldsymbol{\theta}\right) P_i}{8\pi\sigma_i^2 P_0} e^{-\frac{\left(\sqrt{P_i}-\beta_i\right)^2}{\sigma_i^2}} \int_0^{\infty} \frac{1}{f_{T_i|\Omega}\left(t_i:\boldsymbol{\theta}|\Omega\right)} \left(\frac{1}{\mathcal{E}_{b_i}+\tau_i^2} e^{-\frac{t_i}{\mathcal{E}_{b_i}+\tau_i^2}} - \frac{1}{\tau_i^2} e^{-\frac{t_i}{\tau_i^2}}\right)^2 \mathrm{d}t_i, \tag{6.10}$$

where $f_{T_i|\Omega}\left(t_i:\boldsymbol{\theta}|\Omega\right)$ is found in Equation (6.7), and $\boldsymbol{\mathcal{G}}_{i,\Omega}\left(\boldsymbol{\theta}\right)$ is a symmetric 3-by-3 matrix defined as

$$\boldsymbol{\mathcal{G}}_{i,\Omega}\left(\boldsymbol{\theta}\right) \stackrel{\text{def}}{=} \begin{pmatrix} \frac{1}{P_0} & \frac{\alpha}{d_i^2}\left(x_i-x_{\mathsf{T}}\right) & \frac{\alpha}{d_i^2}\left(y_i-y_{\mathsf{T}}\right) \\ \frac{\alpha}{d_i^2}\left(x_i-x_{\mathsf{T}}\right) & \frac{P_0\alpha^2}{d_i^4}\left(x_{\mathsf{T}}-x_i\right)^2 & \frac{P_0\alpha^2}{d_i^4}\left(x_{\mathsf{T}}-x_i\right)\left(y_{\mathsf{T}}-y_i\right) \\ \frac{\alpha}{d_i^2}\left(y_i-y_{\mathsf{T}}\right) & \frac{P_0\alpha^2}{d_i^4}\left(x_{\mathsf{T}}-x_i\right)\left(y_{\mathsf{T}}-y_i\right) & \frac{P_0\alpha^2}{d_i^4}\left(y_{\mathsf{T}}-y_i\right)^2 \end{pmatrix}. \tag{6.11}$$

Note that the matrix $\boldsymbol{\mathcal{G}}_{i,\Omega}(\boldsymbol{\theta})$ and consequently, the FIM and CRLB depend on the realization of the network geometry.

6.3.3 Performance-Assessment Metric for Localization Schemes

One of the main measures used to assess the performance of any source-localization scheme is the *geometric location-estimation error* (GLE) defined as [81]

$$\mathsf{GLE}_{\Omega} \stackrel{\text{def}}{=} \sqrt{\left(\widehat{x}_{\mathsf{T},\Omega} - x_{\mathsf{T}}\right)^2 + \left(\widehat{y}_{\mathsf{T},\Omega} - y_{\mathsf{T}}\right)^2}, \tag{6.12}$$

where the subscript Ω signifies that the GLE depends on the realization of the network geometry. Note that given a specific realization of the network geometry Ω, the following lower bound can be established on the mean squared GLE using the CRLB as defined in Equation (6.9):

$$\begin{aligned} \mathsf{MSGLE}_{\Omega} &\stackrel{\text{def}}{=} \mathbb{E}\left[\mathsf{GLE}_{\Omega}^2\right] = \mathbb{E}\left[\left(\widehat{x}_{\mathsf{T},\Omega} - x_{\mathsf{T}}\right)^2 + \left(\widehat{y}_{\mathsf{T},\Omega} - y_{\mathsf{T}}\right)^2\right] \\ &\geq \left[\boldsymbol{\mathcal{I}}_{\Omega}^{-1}(\boldsymbol{\theta})\right]_{2,2} + \left[\boldsymbol{\mathcal{I}}_{\Omega}^{-1}(\boldsymbol{\theta})\right]_{3,3}, \end{aligned} \tag{6.13}$$

where MSGLE_{Ω} denotes the mean squared GLE, given a specific realization of the network geometry Ω, and the expectation operation is calculated with respect to the distributions of the observation noise, channel fading coefficients, and channel noise.

Since there is no closed-form equation for finding MSGLE_{Ω}, we resort to a Monte-Carlo approach for its calculation as follows. For a fixed arbitrary realization of the network geometry Ω, the set of sensors' locations $\{(x_i, y_i)\}_{i=1}^{K}$ and consequently, the sets of their distances to the source target $\{d_i\}_{i=1}^{K}$ and the received power from the source at their locations $\{P_i\}_{i=1}^{K}$ (defined in Equation (6.1)) are fixed. In order to find the empirical mean squared GLE, N_{MC} Monte-Carlo trials are performed for the given network geometry by generating random observation noises, channel fading coefficients, and channel noises based on their respective distributions introduced in Section 6.2. The empirical mean squared GLE for the given network realization can then be found as

$$\begin{aligned} \mathsf{MSGLE}_{\Omega} &= \frac{1}{N_{\mathsf{MC}}} \sum_{m=1}^{N_{\mathsf{MC}}} \mathsf{SGLE}_{\Omega}^{(m)} \\ &= \frac{1}{N_{\mathsf{MC}}} \sum_{m=1}^{N_{\mathsf{MC}}} \left(\widehat{x}_{\mathsf{T},\Omega}^{(m)} - x_{\mathsf{T}}\right)^2 + \left(\widehat{y}_{\mathsf{T},\Omega}^{(m)} - y_{\mathsf{T}}\right)^2, \end{aligned} \tag{6.14}$$

where $\mathsf{SGLE}_\Omega \stackrel{\text{def}}{=} \mathsf{GLE}_\Omega^2$ and GLE_Ω is defined in Equation (6.12), and the superscript m denotes the result obtained in the mth Monte-Carlo trial.

6.3.4 Derivation of Optimal Local Quantization Thresholds

Note that the performance of both empirical mean squared GLE and its corresponding CRLB for any fixed network realization is a function of the sensors' binary quantization thresholds. In this chapter, the optimal set of local quantization thresholds are found based on the approach proposed in [28]. According to this method, since the main focus of this chapter is the accurate localization of the source target and *not* so much the accurate estimation of P_0 as the received power from the source at the reference distance d_0, the binary quantization thresholds are found such that the CRLB on the mean squared GLE as defined in Equation (6.13) is minimized. In other words, the optimal set of binary quantization thresholds for the optimal source-localization scheme can be found as [28]

$$\left\{\beta_i^{\mathsf{OPT}}\right\}_{i=1}^{K} = \underset{\{\beta_i\}_{i=1}^{K}}{\arg\min} \left(\left[\boldsymbol{\mathcal{I}}_\Omega^{-1}(\boldsymbol{\theta})\right]_{2,2} + \left[\boldsymbol{\mathcal{I}}_\Omega^{-1}(\boldsymbol{\theta})\right]_{3,3} \right), \tag{6.15}$$

where the conditional FIM $\boldsymbol{\mathcal{I}}_\Omega(\boldsymbol{\theta})$, given the current network realization Ω is found using Equations (6.10)–(6.11).

6.4 Numerical Performance Assessment

Ozdemir et al. [29] have reported the performance of their proposed source-localization scheme summarized in Subsections 6.3.1 and 6.3.2 for a WSN deployed in a regular grid configuration. As mentioned previously, the performance of the source-localization method is heavily affected by the realization of the network geometry. In order to observe this dependence, suppose that a WSN consisting of $K = 50$ sensors is randomly deployed to estimate the location and parameter P_0 of a source target located at $(x_\mathsf{T}, y_\mathsf{T}) = (5, 10)$, for which $P_0 = 10,000$, $d_0 = 1$, and the power-decay exponent is $\alpha = 2$. The sensors are randomly placed in the circular surveillance region with radius $R = 50$ and centered at the origin according to a *uniform clustering process*. The local observation noises are assumed to be identically distributed with the same variance $\sigma^2 \equiv \sigma_i^2 = \frac{P_0}{\psi}$, where $\psi \equiv \psi_i$ is the common

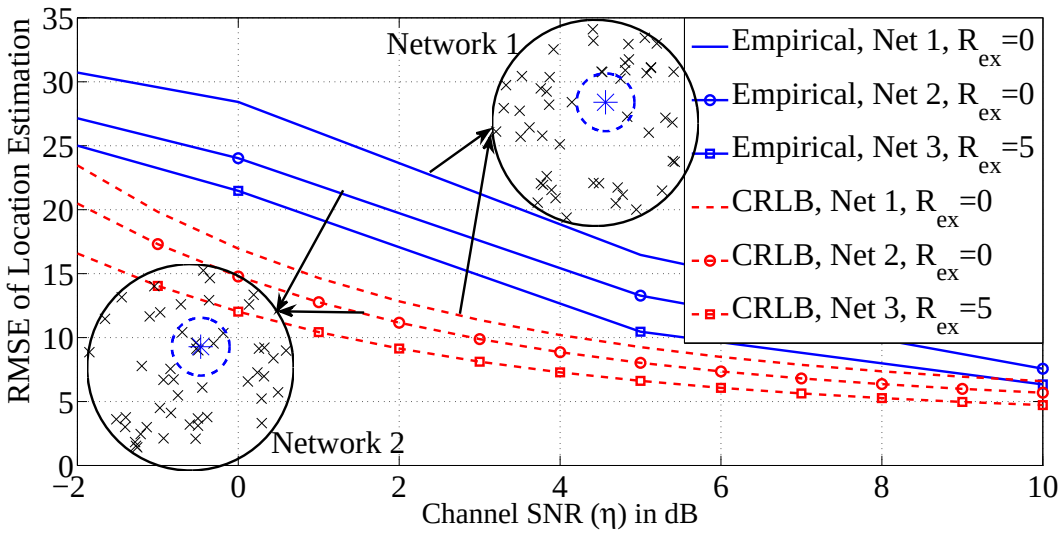


Figure 6.3: Empirical RMSE of the source-location estimation, shown by solid lines, and its corresponding CRLB, shown by dashed lines, vs channel SNR in dB for three different random realizations of the network geometry, when the observation SNR is $\psi = 40$ dB.

observation SNR. Similarly, the local channel noises are assumed to be identically distributed with the same variance $\tau^2 \equiv \tau_i^2 = \frac{\mathcal{E}_b}{\eta}$, where $\mathcal{E}_b \equiv \mathcal{E}_{b_i} = 1$ dB is the common transmission energy when $u_i = 1$ is sent, and $\eta \equiv \eta_i$ is the common channel SNR. Due to the homogeneous nature of the network, all of the binary quantization thresholds are assumed to be identical to $\beta \equiv \beta_i$. The results have been found by averaging over $N_{\mathsf{MC}} = 10,000$ Monte-Carlo trials as explained in Subsection 6.3.3.

Figure 6.3 shows the empirical root mean-squared error (RMSE) for the source-location estimation, plotted by solid lines, and its corresponding CRLB, plotted by dashed lines, as functions of the channel SNR η for three different random network realizations, when the observation SNR is fixed at $\psi = 40$ dB. Details of generating each random network geometry are explained in the next section. The first and second network realizations corresponding to the curves without marker and with the circle marker, respectively, are shown in the corners of Figure 6.3, where sensors are denoted by '$\times$'. In these two network realizations, there is no exclusion zone considered around the sensors (i.e., $R_{\mathsf{ex}} = 0$) and therefore, they may be placed very close to each other. The network realization corresponding to the curves shown by the square marker '$\square$' is depicted in Figure 6.1. In this case, each sensor is surrounded

by an exclusion zone with radius $R_{\mathsf{ex}} = 5$ and therefore, all of the sensors will be apart from each other by at least 5 units of length. It can be seen in this figure that the performance of the source-localization scheme highly depends on the realization of the network geometry. It also shows that as the channel SNR increases, the error in the localization decreases and gets closer to its CRLB, as expected. Similar results can be found by considering the localization performance as a function of the observation SNR.

A close look at all network configurations shown in Figure 6.1 and the corners of Figure 6.3 reveals that a circle with radius $R_{\mathsf{T}} = 14$ is centered at the source target and shown by a dashed line. In Network 1 shown at the top of Figure 6.3, there are only two sensors located within this region surrounding the target, whereas in Network 2 shown at the bottom of this figure, there are six such sensors in the same vicinity of the target. This difference partially explains why the performance of the localization scheme using the two different network realizations is completely different. When there are more sensors within the immediate vicinity of the target, the localization error will be lower since the observations are of higher quality. This point will further be discussed with more details in Subsection 6.5.2.

6.5 Spatial Dependence of Source Localization

In this section, we study the effects of spatial randomness, i.e., random realization of the network geometry, on the performance of the source-localization scheme proposed in [29] and summarized in Section 6.3, through a numerical Monte-Carlo approach. Note that the performance evaluations presented here can easily be extended to any other source-localization method.

Let the *localization outage event* in the space of random realizations of network geometry be defined as

$$\mathcal{A}(\gamma) \stackrel{\text{def}}{=} \left\{\Omega : \mathsf{MSGLE}_{\Omega} > \gamma^2\right\}.$$

Based on the above definition, a *localization outage* occurs when the root mean squared distance between the estimated location of the source $(\widehat{x}_{\mathsf{T}}, \widehat{y}_{\mathsf{T}})$ and its true value $(x_{\mathsf{T}}, y_{\mathsf{T}})$ exceeds a prespecified threshold γ. In other words, a realization of the network geometry Ω is said to be in outage if *on average*, the source-location estimation using that network

deployment produces an error beyond an acceptable threshold γ.

It can be observed that the localization outage is a random variable depending on the distribution of the network geometry. In order to assess the dependence of the localization outage on the realization of the network geometry, we will use the CCDF of the random variable MSGLE_Ω defined as

$$\overline{F}_{\mathsf{MSGLE}}(\gamma) \stackrel{\text{def}}{=} \mathbb{P}\left[\mathsf{MSGLE}_\Omega > \gamma^2\right], \tag{6.16}$$

where the right-hand side of the equation is the probability that an arbitrary network geometry is in outage, as defined above.

In the following discussions, the performance assessments will be based on a Monte-Carlo approach with 500 simulation trials ran as follows. In each simulation trial, a realization of the network geometry is obtained by randomly placing K sensors within the circular region of radius R. There can be an exclusion zone around each sensor within which no other sensors can be placed. The sensors are located within the surveillance region successively according to a *uniform clustering process* as follows. A pair of independent random variables (x_i, y_i), $i = 1, 2, \ldots, K$, is selected from a uniform distribution over $[-R, R]$. If the sensor falls outside of the circular disk of radius R, i.e., $x_i^2 + y_i^2 > R^2$, this process is repeated until the sensor falls inside the surveillance region. If an exclusion zone of radius R_{ex} is considered around each sensor, the distances between the ith sensor and all $i-1$ other sensors are found, and the above process of assigning new random location to the ith sensor is repeated as many times as necessary until the sensor is located outside of the exclusion zones of all other previously located sensors in the network.

In the next step, the empirical mean squared GLE (i.e., MSGLE_Ω) is found for a fixed random realization of the network geometry Ω using the Monte-Carlo approach described in Subsection 6.3.3 with $N_{\mathsf{MC}} = 1,000$ trials per network realization. The CRLB on the mean squared GLE is found *only once* for each network realization as defined in Equation (6.13). The optimal local binary quantization threshold is fixed and found only once for any network realization, using the approach discussed in Subsection 6.3.4. All of the simulation parameters are exactly the same as those summarized in Section 6.4. The observation SNR and channel SNR are fixed at $\psi = 40$ dB and $\eta = 0$ dB, respectively.

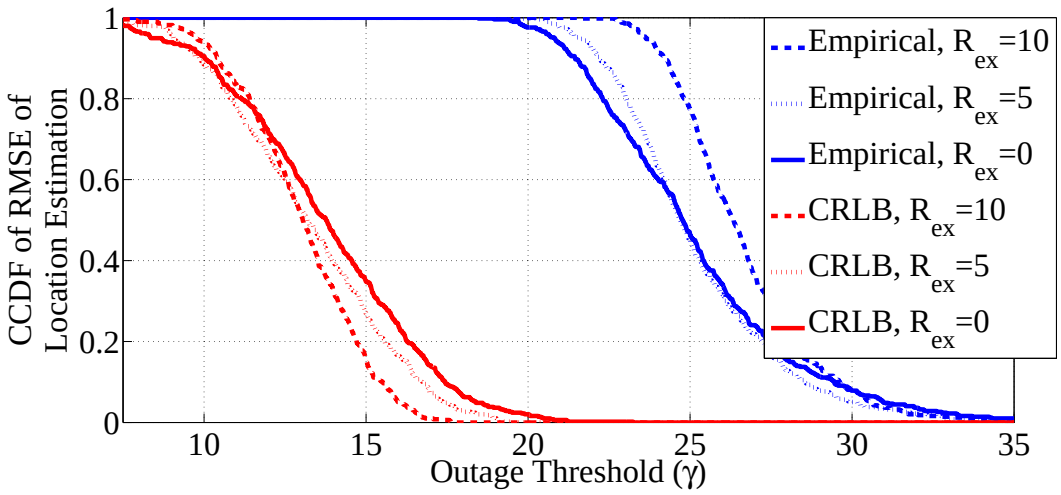


Figure 6.4: CCDF of the empirical RMSE of the source-location estimation and its corresponding CRLB vs the outage threshold γ for different settings of the network geometry. The observation SNR and channel SNR are fixed at $\psi = 40$ dB and $\eta = 0$ dB, respectively.

6.5.1 Effect of Sensor Exclusion Zones on Source Localization

One of the parameters affecting the performance of any source-localization scheme, which is based on the assumption of random sensor placement, is the *exclusion zone* around each sensor. The exclusion zone is a circular disk around each sensor within which no other sensors can be placed. It can be the result of a physical limitation that does not allow such proximity of two sensors or it can be controlled by the network administrator during the network deployment in order to guarantee a proper coverage of the surveillance region. Figure 6.4 depicts the CCDF of the empirical RMSE of the source-location estimation, as defined by Equation (6.16), and its corresponding CRLB as functions of the outage threshold γ for different values of the radius of sensor exclusion zones R_{ex}. The results were obtained using 500 Monte-Carlo trials for generating random network realizations as described at the beginning of this section.

As it can be seen from Figure 6.4, the probability of an empirical localization outage increases as the radius of the sensor exclusion zones increases. In other words, as the exclusion zone around each sensor expands, the probability that the average GLE of an arbitrary random network deployment exceeds a prescribed threshold increases, due to the fact that the expansion of the exclusion zones around sensors results in them being located farther

apart. Therefore, the number of sensors that can be located close to a target decreases on average. This will result in a lower number of strong local measurements, which in turn decreases the quality of the data available at the FC as more sensors are likely to have sent zeros. It should be mentioned that for lower probabilities of localization outage, i.e., higher values of outage threshold γ, the exclusion zones around sensors do not have much effect as almost any network realization can on average satisfy the required accuracy of location estimation. The same argument applies to the CCDF for the CRLB values on the root mean squared location estimation error.

6.5.2 Effect of the Closest Sensors to Source on Localization

It is intuitive that the performance of any source-localization scheme depends mainly on the observation and channel qualities of the closest sensors to the target. In order to investigate this effect, consider a scenario in which there is no exclusion zone around the sensors, i.e., $R_{\mathsf{ex}} = 0$. Note that similar results and discussions can be found for the network realizations with an arbitrary exclusion zone around sensors. Let R_{T} denote the radius of a circular region around the target within which we assume the most important sensors to the performance of the source-localization scheme are located. Let K_{T} denote the number of sensors located within this region. In the network realization depicted in Figure 6.1, the region around the target is shown by a dashed line as a circle with radius $R_{\mathsf{T}} = 14$, and the number of sensors within this region is $K_{\mathsf{T}} = 1$. Note that in general, $0 \leq R_{\mathsf{T}} \leq 2R$ and $0 \leq K_{\mathsf{T}} \leq K$, where R is the radius of the surveillance region. Figure 6.5 depicts the CCDF of the empirical RMSE of the source-location estimation, as defined by Equation (6.16), as a function of the outage threshold γ for different values of R_{T} and K_{T}, when there is no exclusion zone around the sensors, i.e., $R_{\mathsf{ex}} = 0$. The results were obtained in a similar way to the procedure explained at the beginning of this section.

As it can be seen in Figure 6.5, for a given R_{T}, the probability of localization outage decreases as K_{T} increases. In other words, if in the random realizations of the network geometry, the number of sensors located within a fixed radius around the target increases, the probability that an arbitrary network deployment is in outage drastically decreases. In

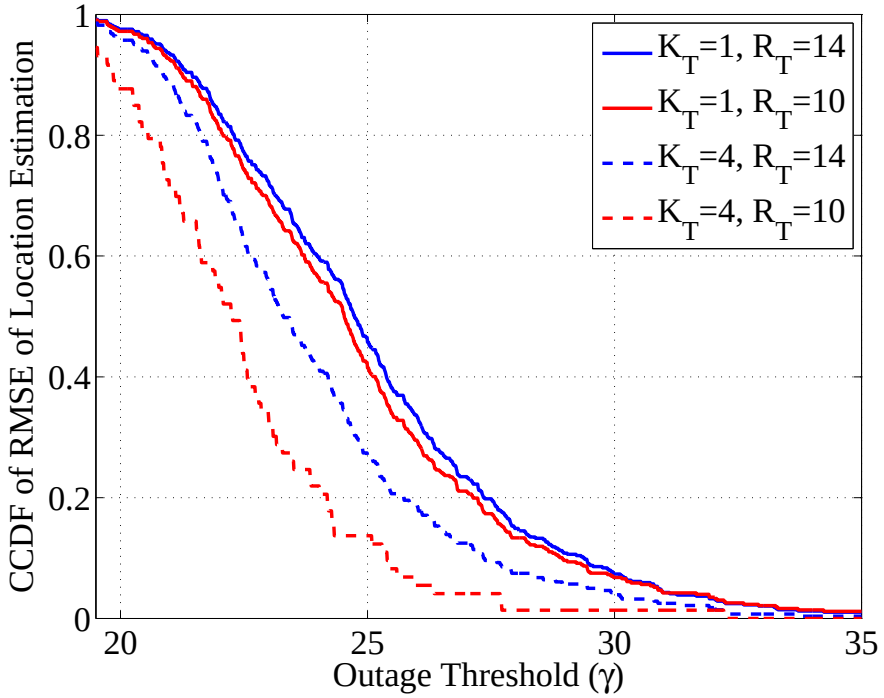


Figure 6.5: CCDF of the empirical RMSE of the source-location estimation vs the outage threshold γ for different values of R_{T} and K_{T}, when $R_{\mathsf{ex}} = 0$. The network geometries are generated without considering any sensor exclusion zones. The observation SNR and channel SNR are $\psi = 40$ dB and $\eta = 0$ dB, respectively.

a similar discussion, for a given K_{T}, the probability of outage decreases as the radius R_{T} decreases. In other words, if we need to expand the region around the target to have a specific, fixed number of sensors located close to it, the probability of outage increases as the region expands. Figure 6.5 shows that the effect of increasing K_{T} for a fixed R_{T} is always noticeable, whereas the effect of decreasing R_{T} for a fixed K_{T} is more noticeable when the number of sensors considered within the neighborhood of the target is larger. The important implication of this discussion in practical network design is that the density of the randomly deployed network should be above a threshold to guarantee that the sensors are so closely located that if the target location is anywhere within the surveillance region, there are enough number of sensors in its proximity.

6.6 Conclusions

The main focus of this chapter was to quantify the effects of spatial randomness on the performance of source-localization schemes. To this end and for demonstration purposes, a recently proposed approach was studied in which the FC of a WSN finds the maximum-likelihood estimate of the location of a point source (in addition to another parameter related to the power-decay model of its power radiations). The FC uses the energies received from a set of spatially distributed sensors that make binary quantization of their local noisy power observations. The random realization of the network geometry was assumed to be according to a uniform clustering process. The concept of localization outage was defined to be a realization of the network geometry that on average fails to satisfy a required threshold on the localization accuracy. The numerical results verified that the source-localization performance is heavily affected by the realization of sensor deployment and that it highly depends on the number of sensors that are within a close proximity of the source. This conclusion suggests a guideline that the sensor density in the network should appropriately be chosen such that enough number of sensors will be close to a target arbitrarily located within a random realization of the network geometry. As the network density increases, resulting in a higher number of sensors in a fixed disk around the source, the performance of the localization scheme improves drastically. The effect of exclusion zones around sensors was also studied based on which increasing the minimum sensor separation increases the localization-outage probability, i.e., if the sensors are forced to be farther separated, it is more likely that a random network realization will be in outage.

Chapter 7

Summary and Future Work

In this dissertation, we investigated the problem of distributed detection, classification, and estimation in WSNs, which enables a wide range of applications such as event detection, classification, localization, system identification, and target tracking. In this chapter, a summary of the problems considered in the previous chapters and the main results of each proposed solution are presented in Section 7.1. Section 7.2 presents several avenues for further exploration that can be considered a starting point for continued research in the domain of distributed estimation in WSNs.

7.1 Summary

In Chapter 2, we presented an extensive literature review on the distributed detection and classification in WSNs and summarized the results of major research accomplishments in this area up to date. Furthermore, we proposed an approach for distributed multi-hypothesis classification of an underlying hypothesis at the FC of a WSN using local *binary* decisions based on the known influence fields characterizing different hypotheses. The main contribution of this chapter is the formulation of local and fusion decision rules that maximize the probability of correct global classification at the FC, along with an algorithm for channel-aware global optimization of the decision thresholds at local sensors and the FC. The results of numerical analyses showed that the proposed approach simplifies decision making at the sensors while achieving an acceptable performance in terms of the global average probability of correct

classification at the FC, for a wide range of the parameters of the system, including the number of distributed sensors, the observation SNR, and the channel SNR. Furthermore, it was shown that a global optimization of the local decision thresholds improves the probability of correct classification at the FC compared to the case in which local thresholds are only locally optimized.

In Chapter 3, the problem of distributed estimation of a vector of unknown parameters associated with a deterministic, two-dimensional function using its spatially distributed, noisy samples was considered in the context of WSNs. The ML estimate of the vector of unknown parameters at the FC was derived for both analog and digital local processing schemes. Since the ML estimate for the case of digital local processing scheme was too complicated to be implemented, an efficient iterative EM algorithm was proposed to numerically find the ML estimate in this case. Numerical simulation results proved that the proposed distributed estimation framework achieves a very good performance in terms of the integrated mean-squared error for reasonable values of the parameters of the system, including the number of distributed sensors in the observation environment, the observation SNR, the channel SNR, and the number of quantization levels for digital local processing scheme. In particular, numerical performance analysis showed that even with a low number of quantization levels at distributed sensors, i.e., high energy efficiency, the estimation framework provides a very good performance in terms of the integrated mean-squared error.

In Chapter 4, an adaptive power-allocation scheme was proposed to find the optimal local amplification gains in a distributed estimation framework. The proposed power allocation minimizes the L^2-norm of the vector of local transmission powers in a WSN, given a maximum variance for the BLUE estimator of a scalar, random signal at the FC. This approach results in an increase in the lifetime of the network at the expense of a potential slight increase in the sum total transmission power of all sensors. This is because the proposed scheme prevents the assignment of high transmission power to sensors by putting a higher penalty on them, compared to similar approaches that are based on the minimization of the sum of the local transmission powers. The limitation of the proposed power-allocation scheme is that the optimal local amplification gains found based on it depend on the instantaneous fading coefficients of the channels between the sensors and FC. The next contribution of this

chapter was to propose a limited-feedback strategy to eliminate the requirement of infinite-rate, error-free feedback of the instantaneous forward CSI from the FC to local sensors. This scheme designs an optimal codebook by quantizing the vector space of the optimal local amplification gains using the generalized Lloyd algorithm with modified distortion metrics. Numerical results showed that the proposed adaptive power-allocation scheme achieves a high energy efficiency, and that even with a limited number of feedback bits (small codebook), its average energy efficiency based on the proposed limited-feedback strategy is close to that of a WSN with full CSI feedback.

In Chapter 5, we investigated the effect of linear spatial collaboration on the performance of the BLUE estimator at the FC of a WSN that estimates the vector of spatially correlated signals observed by sensors. In this context, each sensor can collaborate with a subset of other sensors by sharing its local observations with them through error-free, low-cost links. An optimal linear spatial-collaboration scheme was derived in which the set of coefficients or weights used to form linear combinations of shared noisy observations at the sensors connected to the FC is derived. The objective of this power-allocation scheme is to minimize the sum of the estimation variances of different signals observed by the network, given a constraint on the average cumulative transmission power in the entire network. The numerical results showed that even a small degree of connectivity and spatial collaboration in the network improves the quality of the estimators at the FC, and that the collaboration gain is more significant when the signals observed by sensors have a higher spatial correlation.

In Chapter 6, the effects of spatial randomness on the performance of source-localization schemes were demonstrated using a recently proposed approach in which the FC of a WSN finds the ML estimate of the location of a point source based on the energies received by a set of spatially distributed sensors that make binary quantization of their local noisy power observations. The random realization of the network geometry was assumed to be according to a uniform clustering process. The concept of localization outage was defined to be a realization of the network geometry that on average fails to satisfy a required threshold on the localization accuracy. The numerical results verified that the source-localization performance is heavily affected by the realization of sensor deployment, and that it highly depends on the number of sensors that are within a close proximity of the source. This

conclusion suggests a guideline that the sensor density in the network should appropriately be chosen such that enough number of sensors will be close to a target arbitrarily located within a random realization of the network geometry. As the network density increases, resulting in a higher number of sensors in a fixed disk around the source, the performance of the localization scheme improves drastically. The effect of exclusion zones around sensors was also studied based on which increasing the minimum sensor separation increases the localization-outage probability, i.e., if the sensors are forced to be farther separated, it is more likely that a random network realization will be in outage.

7.2 Future Work

In this section, several research ideas are summarized that can be considered as a continuation of the works presented in the previous chapters.

7.2.1 Extension of Chapter 4

The analyses and results presented in this chapter can be extended in the following aspects:

1. Cui et al. [17] have introduced the concepts of *estimation outage* and *estimation diversity*. The estimation outage probability is defined as

$$\mathcal{P}_{\mathsf{out}}(\gamma) \stackrel{\text{def}}{=} \mathbb{P}\left[\mathsf{Var}\left(\widehat{\theta}\right) > \gamma\right],$$

   where γ is a predefined threshold. The estimation diversity is defined as the exponential rate at which the asymptotic estimation outage probability for large WSNs (i.e., large number of distributed sensors K) decreases as the total transmission power in the network increases. It is an interesting problem to derive the estimation outage and estimation diversity for the proposed power-allocation scheme and to compare them with those for the cases of equal power allocation and optimal power allocation based on the minimization of the sum of the local transmission powers.

2. It is also interesting to quantify and/or analyze through numerical simulations the effects of the proposed limited-feedback scheme on the estimation outage probability

and estimation diversity of the proposed power-allocation approach. In other words, one can attempt to find the potential increase in the estimation outage probability and the potential loss in the estimation diversity due to the limited feedback.

7.2.2 Extension of Chapter 5

The analyses and results presented in this chapter can be extended in the following aspects:

1. The proposed optimal power allocation in this chapter minimizes the total estimation distortion at the FC subject to a constraint on the average transmission power in the entire network. A more meaningful constraint in this optimization problem can be put on the average transmission power of each sensor. One can attempt to solve this problem with the new constraint and discuss the differences between the two solutions.

2. In our analyses in this chapter, we assumed that the sensors collaborate with each other through error-free, low-cost links. These assumptions may not be valid in practical WSNs. One can attempt to analyze our power-allocation scheme under the assumption of costly collaboration. In this case, the available power budget should appropriately be divided between collaboration and transmission. Furthermore, the problem can be extended to a case in which a collaboration noise is added to local observations. In this case, the optimal mixing matrix is expected to also depend on the level of unreliability in the collaboration channel, i.e., how noisy the channel between two collaborating sensors is.

7.2.3 Extension of Chapter 6

In Chapter 6, we analyzed the effects of spatial randomness of sensor locations on the performance of an energy-based source-localization algorithm through extensive *numerical simulations* using the measure of localization outage. More *theoretical* analyses can be performed based on the rich results of the theory of stochastic geometry. In other words, the following question can be considered as the backbone of a future research study:

> "Through theoretical analysis based on the theory of stochastic geometry, what is the probability that a randomly placed network of wireless sensors *on average* can

achieve a minimum localization accuracy at the FC for a given source-localization scheme?"

References

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Communications Magazine*, vol. 40, no. 8, pp. 102–114, August 2002. 1

[2] J. Yick, B. Mukherjee, and D. Ghosal, "Wireless sensor network survey," *Elsevier Computer Networks*, vol. 52, no. 12, pp. 2292–2330, August 2008. 1

[3] T. Bokareva, W. Hu, S. Kanhere, B. Ristic, T. Bessell, M. Rutten, and S. Jha, "Wireless sensor networks for battlefield surveillance," in *Land Warfare Conference*, Brisbane, Queensland, Australia, October 2006. 1

[4] K. Lu, Y. Qian, D. Rodriguez, W. Rivera, and M. Rodriguez, "Wireless sensor networks for environmental monitoring applications: A design framework," in *IEEE Global Telecommunications Conference (GLOBECOM)*, Washington, DC, USA, November 2007, pp. 1108–1112. 1

[5] P. Corke, T. Wark, R. Jurdak, W. Hu, P. Valencia, and D. Moore, "Environmental wireless sensor networks," *Proceedings of the IEEE*, vol. 98, no. 11, pp. 1903–1917, November 2010. 1

[6] X. Yu, P. Wu, W. Han, and Z. Zhang, "A survey on wireless sensor network infrastructure for agriculture," *Elsevier Computer Standards & Interfaces*, vol. 35, no. 1, pp. 59–64, January 2013. 1

[7] A. ur Rehman, A. Z. Abbasi, N. Islam, and Z. A. Shaikh, "A review of wireless sensors and networks' applications in agriculture," *Elsevier Computer Standards & Interfaces*, vol. 36, no. 2, pp. 263–270, February 2014. 1

[8] H. Yan, Y. Xu, and M. Gidlund, "Experimental e-health applications in wireless sensor networks," in *WRI International Conference on Communications and Mobile Computing (CMC)*, vol. 1, Kunming, Yunnan, China, January 2009, pp. 563–567. 1

[9] J. Ko, C. Lu, M. Srivastava, J. Stankovic, A. Terzis, and M. Welsh, "Wireless sensor networks for healthcare," *Proceedings of the IEEE*, vol. 98, no. 11, pp. 1947–1960, November 2010. 1

[10] H. Alemdar and C. Ersoy, "Wireless sensor networks for healthcare: A survey," *Elsevier Computer Networks*, vol. 54, no. 15, pp. 2688–2710, October 2010. 1

[11] Z. Wei, N. Shangpin, and L. Lu, "Wireless sensor networks for in-home healthcare: Issues, trend and prospect," in *International Conference on Computer Science and Network Technology (ICCSNT)*, vol. 2, Harbin, China, December 2011, pp. 970–973. 1

[12] C.-Y. Chong and S. Kumar, "Sensor networks: Evolution, opportunities, and challenges," *Proceedings of the IEEE*, vol. 91, no. 8, pp. 1247–1256, August 2003. 1

[13] I. F. Akyildiz, D. Pompili, and T. Melodia, "Underwater acoustic sensor networks: Research challenges," *Ad hoc networks*, vol. 3, no. 3, pp. 257–279, 2005. 1

[14] D. Li, K. D. Wong, Y. H. Hu, and A. M. Sayeed, "Detection, classification, and tracking of targets in distributed sensor networks," *IEEE Signal Processing Magazine*, vol. 19, no. 2, pp. 17–29, March 2002. 1, 12

[15] J.-F. Chamberland and V. Veeravalli, "Wireless sensors in distributed detection applications," *IEEE Signal Processing Magazine*, vol. 24, no. 3, pp. 16–25, May 2007. 1, 14, 15, 16, 17, 18, 19

[16] M. H. Chaudhary and L. Vandendorpe, "Performance of power-constrained estimation in hierarchical wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 61, no. 3, pp. 724–739, February 2013. 1, 87

[17] S. Cui, J.-J. Xiao, A. J. Goldsmith, Z.-Q. Luo, and H. V. Poor, "Estimation diversity and energy efficiency in distributed sensing," *IEEE Transactions on Signal Processing*, vol. 55, no. 9, pp. 4683–4695, September 2007. 1, 2, 70, 71, 87, 120

[18] J.-J. Xiao, S. Cui, Z.-Q. Luo, and A. J. Goldsmith, "Linear coherent decentralized estimation," *IEEE Transactions on Signal Processing*, vol. 56, no. 2, pp. 757–770, February 2008. 1, 2, 46, 47, 70, 87

[19] M. K. Banavar, C. Tepedelenlioğlu, and A. Spanias, "Estimation over fading channels with limited feedback using distributed sensing," *IEEE Transactions on Signal Processing*, vol. 58, no. 1, pp. 414–425, January 2010. 1, 2, 51, 70, 87

[20] M. Fanaei, M. C. Valenti, N. A. Schmid, and M. M. Alkhweldi, "Distributed parameter estimation in wireless sensor networks using fused local observations," in *SPIE Wireless Sensing, Localization, and Processing VII*, vol. 8404, Baltimore, MD, USA, May 2012. 1, 2, 8, 70, 87

[21] J.-J. Xiao and Z.-Q. Luo, "Decentralized estimation in an inhomogeneous sensing environment," *IEEE Transactions on Information Theory*, vol. 51, no. 10, pp. 3564–3575, October 2005. 1, 2, 87

[22] J.-J. Xiao, S. Cui, Z.-Q. Luo, and A. J. Goldsmith, "Power scheduling of universal decentralized estimation in sensor networks," *IEEE Transactions on Signal Processing*, vol. 54, no. 2, pp. 413–422, February 2006. 1, 2, 87

[23] Z.-Q. Luo, "An isotropic universal decentralized estimation scheme for a bandwidth constrained ad hoc sensor network," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 4, pp. 735–744, April 2005. 1, 2, 87

[24] A. Ribeiro and G. B. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks–Part I: Gaussian case," *IEEE Transactions on Signal Processing*, vol. 54, no. 3, pp. 1131–1143, March 2006. 1, 2, 4, 47, 87

[25] ——, "Bandwidth-constrained distributed estimation for wireless sensor networks–Part II: Unknown probability density function," *IEEE Transactions on Signal Processing*, vol. 54, no. 7, pp. 2784–2796, July 2006. 1, 2, 48, 87

[26] P. Ishwar, R. Puri, K. Ramchandran, and S. Pradhan, "On rate-constrained distributed estimation in unreliable sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 4, pp. 765–775, April 2005. 1, 2, 47, 87

[27] A. Arora, P. Dutta, S. Bapat, V. Kulathumani, H. Zhang, V. Naik, V. Mittal, H. Cao, M. Demirbas, M. Gouda, Y. Choi, T. Herman, S. Kulkarni, U. Arumugam, M. Nesterenko, A. Vora, and M. Miyashita, "A line in the sand: A wireless sensor network for target detection, classification, and tracking," *Computer Networks*, vol. 46, no. 5, pp. 605–634, December 2004. 2, 7, 13, 30

[28] R. Niu and P. K. Varshney, "Target location estimation in sensor networks with quantized data," *IEEE Transactions on Signal Processing*, vol. 54, no. 12, pp. 4519–4528, December 2006. 4, 48, 100, 109

[29] O. Ozdemir, R. Niu, and P. K. Varshney, "Channel aware target localization with quantized data in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 57, no. 3, pp. 1190–1202, March 2009. 4, 48, 100, 101, 105, 106, 107, 109, 111

[30] M. Chiani, A. Conti, and R. Verdone, "Partial compensation signal-level-based up-link power control to extend terminal battery duration," *IEEE Transactions on Vehicular Technology*, vol. 50, no. 4, pp. 1125–1131, July 2001. 6

[31] M. Fanaei, M. C. Valenti, N. A. Schmid, and V. Kulathumani, "Channel-aware distributed classification in wireless sensor networks using binary local decisions," in *SPIE Signal Processing, Sensor Fusion, and Target Recognition XX*, I. Kadar, Ed., vol. 8050, Orlando, FL, USA, May 2011, pp. 841–844. 7

[32] M. Fanaei, M. C. Valenti, and N. A. Schmid, "Limited-feedback-based channel-aware power allocation for linear distributed estimation," in *Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, USA, November 2013. 9, 87

[33] ——, "Power allocation for distributed BLUE estimation with full and limited feedback of CSI," in *IEEE Military Communications Conference (MILCOM)*, San Diego, CA, USA, November 2013. 9, 87

[34] M. Fanaei, M. Valenti, A. Jamalipour, and N. Schmid, "Optimal power allocation for distributed BLUE estimation with linear spatial collaboration," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Florence, Italy, May 2014, pp. 5452–5456. 10

[35] M. Fanaei, M. Valenti, and N. Schmid, "Effects of spatial randomness on locating a point source with distributed sensors," in *IEEE International Conference on Communications (ICC) Workshop on Advances in Network Localization and Navigation (ANLN)*, Sydney, Australia, June 2014, pp. 186–192. 10

[36] B. Chen, L. Tong, and P. Varshney, "Channel-aware distributed detection in wireless sensor networks," *IEEE Signal Processing Magazine*, vol. 23, no. 4, pp. 16–26, July 2006. 14

[37] R. R. Tenny and N. R. Sandell, "Detection with distributed sensors," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 17, no. 4, pp. 501–510, July 1981. 14

[38] J. N. Tsitsiklis, "Decentralized detection," *Advances in Statistical Signal Processing*, vol. 2, pp. 297–344, 1993. 15, 16, 17, 18

[39] R. Viswanathan and P. K. Varshney, "Distributed detection with multiple sensors–Part I: Fundamentals," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 54–63, January 1997. 15, 16

[40] R. S. Blum, S. A. Kassam, and H. V. Poor, "Distributed detection with multiple sensors–Part II: Advanced topics," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 64–79, January 1997. 15

[41] H. L. V. Trees, *Detection, Estimation, and Modulation Theory, Part I: Detection, Estimation, and Linear Modulation Theory*, reprint first ed. New York: John Wiley and Sons, 2001. 15

[42] Z. Chair and P. K. Varshney, "Distributed Bayesian hypothesis testing with distributed data fusion," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 18, no. 5, pp. 695–699, September/October 1988. 16, 23

[43] J. N. Tsitsiklis and M. Athans, "On the complexity of decentralized decision making and detection problems," *IEEE Transactions on Automatic Control*, vol. 30, no. 5, pp. 440–446, May 1985. 16

[44] S. A. Aldosari and J. M. F. Moura, "Detection in sensor networks: The saddle-point approximation," *IEEE Transactions on Signal Processing*, vol. 55, no. 1, pp. 327–340, January 2007. 17

[45] X. Nguyen, M. J. Wainwright, and M. I. Jordan, "Nonparametric decentralized detection using kernel methods," *IEEE Transactions on Signal Processing*, vol. 53, no. 11, pp. 4053–4066, November 2005. 17

[46] R. S. Blum and S. A. Kassam, "Optimum distributed detection of weak signals in dependent sensors," *IEEE Transactions on Information Theory*, vol. 38, no. 3, pp. 1066–1079, May 1992. 17, 24

[47] M. Cherikh and P. B. Kantor, "Counterexamples in distributed detection," *IEEE Transactions on Information Theory*, vol. 38, no. 1, pp. 162–165, January 1992. 17

[48] J. N. Tsitsiklis, "Decentralized detection by a large number of sensors," *Mathematics of Control, Signals, and Systems*, vol. 1, no. 2, pp. 167–182, 1988. 17

[49] P. Swaszek and P. Willett, "Parley as an approach to distributed detection," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 31, no. 1, pp. 447–457, January 1995. 19

[50] A. D'Costa, V. Ramachandran, and A. M. Sayeed, "Distributed classification of Gaussian space-time sources in wireless sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 6, pp. 1026–1036, August 2004. 19

[51] J.-F. Chamberland and V. V. Veeravalli, "Asymptotic results for decentralized detection in power constrained wireless sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 6, pp. 1007–1015, August 2004. 20, 21, 22, 23, 24

[52] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. John Wiley and Sons, 2006. 20, 21

[53] S. A. Aldosari and J. M. F. Moura, "Detection in decentralized sensor networks," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 2, Montreal, Quebec, Canada, May 2004, pp. 277–280. 22

[54] M. A. Lexa, C. J. Rozell, S. Sinanović, and D. H. Johnson, "To cooperate or not to cooperate: Detection strategies in sensor networks," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 3, Montreal, Quebec, Canada, May 2004, pp. 841–844. 22

[55] B. Chen, R. Jiang, T. Kasetkasem, and P. K. Varshney, "Channel aware decision fusion in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 52, no. 12, pp. 3454–3458, December 2004. 22, 23

[56] J. J. Xiao and Z. Q. Luo, "Universal decentralized detection in a bandwidth-constrained sensor network," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 2617–2624, August 2005. 22

[57] R. Niu, B. Chen, and P. K. Varshney, "Fusion of decisions transmitted over Rayleigh fading channels in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 54, no. 3, pp. 1018–1027, March 2006. 23

[58] V. Aalo and R. Viswanathan, "On distributed detection with correlated sensors: Two examples," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 25, no. 3, pp. 414–421, May 1989. 24

[59] E. Drakopoulos and C. C. Lee, "Optimum multisensor fusion of correlated local decisions," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 27, no. 4, pp. 593–606, July 1991. 24

[60] P. Willett, P. F. Swaszek, and R. S. Blum, "The good, bad and ugly: Distributed detection of a known signal in dependent Gaussian noise," *IEEE Transactions on Signal Processing*, vol. 48, no. 12, pp. 3266–3279, December 2000. 24

[61] M. Kam, Q. Zhu, and W. S. Gray, "Optimal data fusion of correlated local decisions in multiple sensor detection systems," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 28, no. 3, pp. 916–120, January 1992. 24

[62] R. S. Blum and S. A. Kassam, "Distributed cell-averaging CFAR detection in dependent sensors," *IEEE Transactions on Information Theory*, vol. 41, no. 2, pp. 513–518, March 1995. 25

[63] J. G. Chen and N. Ansari, "Adaptive fusion of correlated local decisions," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 28, no. 2, pp. 276–281, May 1998. 25

[64] R. S. Blum, "Locally optimum distributed detection of correlated random signals based on ranks," *IEEE Transactions on Information Theory*, vol. 42, no. 3, pp. 931–942, May 1996. 25

[65] J.-F. Chamberland and V. V. Veeravalli, "How dense should a sensor network be for detection with correlated observations?" *IEEE Transactions on Information Theory*, vol. 52, no. 11, pp. 5099–5106, November 2006. 25

[66] M. H. Shalaby and A. Papamarcou, "Error exponents for distributed detection of Markov sources," *IEEE Transactions on Information Theory*, vol. 40, no. 2, pp. 397–408, March 1994. 25

[67] F. A. Sadjadi, "Hypotheses testing in a distributed environment," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 22, no. 2, pp. 134–137, March 1986. 25

[68] T.-Y. Wang, Y. Han, P. Varshney, and P.-N. Chen, "Distributed fault-tolerant classification in wireless sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 4, pp. 724–734, April 2005. 26

[69] T.-Y. Wang, Y. Han, B. Chen, and P. Varshney, "A combined decision fusion and channel coding scheme for distributed fault-tolerant classification in wireless sensor networks," *IEEE Transactions on Wireless Communications*, vol. 5, no. 7, pp. 1695–1705, July 2006. 26, 27

[70] T.-Y. Wang, Y. Han, and P. Varshney, "Fault-tolerant distributed classification based on non-binary codes in wireless sensor networks," *IEEE Communications Letters*, vol. 9, no. 9, pp. 808–810, September 2005. 27

[71] H.-T. Pai, Y. S. Han, and J.-T. Sung, "Two-dimensional coded classification schemes in wireless sensor networks," *IEEE Transactions on Wireless Communications*, vol. 7, no. 5, pp. 1450–1455, May 2008. 27

[72] X. Zhu, Y. Yuan, C. Rorres, and M. Kam, "Distributed M-ary hypothesis testing with binary local decisions," *Elsevier Information Fusion*, vol. 5, no. 3, pp. 157–167, September 2004. 27

[73] Q. Zhang and P. K. Varshney, "Towards the fusion of distributed binary decision tree classifiers," in *2nd International Conference on Information Fusion*, Sunnyvale, CA, USA, July 1999. 27, 28

[74] K. Nguyen, T. Alpcan, and T. Basar, "Distributed hypothesis testing with a fusion center: The conditionally dependent case," in *47th IEEE Conference on Decision and Control (CDC)*, Cancún, Mexico, December 2008, pp. 4164–4169. 28

[75] Z.-B. Tang, K. Pattipati, and D. Kleinman, "A distributed M-ary hypothesis testing problem with correlated observations," *IEEE Transactions on Automatic Control*, vol. 37, no. 7, pp. 1042–1046, July 1992. 28

[76] S. Bapat, V. Kulathumani, and A. Arora, "Reliable estimation of influence fields for classification and tracking in an unreliable sensor network," in *24th IEEE Symposium on Reliable Distributed Systems (SRDS)*, Orlando, FL, USA, October 2005, pp. 60–69. 29

[77] J. Li and G. AlRegib, "Distributed estimation in energy-constrained wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 57, no. 10, pp. 3746–3758, October 2009. 30

[78] M. Gastpar, "Uncoded transmission is exactly optimal for a simple Gaussian sensor network," *IEEE Transactions on Information Theory*, vol. 54, no. 11, pp. 5247–5251, November 2008. 47

[79] R. Nowak, U. Mitra, and R. Willett, "Estimating inhomogeneous fields using wireless sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 6, pp. 999–1006, August 2004. 47

[80] E. Maşazade, R. Niu, P. K. Varshney, and M. Keskinoz, "Energy aware iterative source localization for wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 58, no. 9, pp. 4824–4835, September 2010. 48

[81] X. Li, "RSS-based location estimation with unknown path-loss model," *IEEE Transactions on Wireless Communications*, vol. 5, no. 12, pp. 3626–3633, December 2006. 49, 100, 108

[82] N. Salman, M. Ghogho, and A. H. Kemp, "On the joint estimation of the RSS-based location and path-loss exponent," *IEEE Wireless Communications Letters*, vol. 1, no. 1, pp. 34–37, February 2012. 49

[83] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society, Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, December 1977. 55

[84] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, 1st ed. NJ: Prentice Hall, 1993. 74, 94, 107

[85] T. K. Moon and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing*. New Jersey: Prentice Hall, 2000. 76

[86] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Boston: Kluwer Academic Publishers, 1992. 79, 81

[87] I. Bahçeci and A. Khandani, "Linear estimation of correlated data in wireless sensor networks with optimum power allocation and analog modulation," *IEEE Transactions on Communications*, vol. 56, no. 7, pp. 1146–1156, July 2008. 87, 88, 89, 90

[88] J. Fang and H. Li, "Power constrained distributed estimation with correlated sensor data," *IEEE Transactions on Signal Processing*, vol. 57, no. 8, pp. 3292–3297, August 2009. 87, 94

[89] U. Rashid, H. D. Tuan, P. Apkarian, and H. H. Kha, "Globally optimized power allocation in multiple sensor fusion for linear and nonlinear networks," *IEEE Transactions on Signal Processing*, vol. 60, no. 2, pp. 903–915, February 2012. 87, 96

[90] S. Kar and P. Varshney, "Linear coherent estimation with spatial collaboration," *IEEE Transactions on Information Theory*, vol. 59, no. 6, pp. 3532–3553, June 2013. 89, 91

[91] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1991. 95, 96

[92] M. Kočvara and M. Stingl, "PENBMI," Version 2.1, 2004, See www.penopt.com for a free developer version. 96

[93] J. Löfberg, "YALMIP: A toolbox for modeling and optimization in MATLAB®," in *IEEE International Symposium on Computer Aided Control Systems Design (CACSD)*, Taipei, Taiwan, September 2004. 96

[94] A. Sundaresan and P. K. Varshney, "Location estimation of a random signal source based on correlated sensor observations," *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 787–799, February 2011. 100

[95] D. Li and Y. H. Hu, "Energy-based collaborative source localization using acoustic microsensor array," *EURASIP Journal on Advances in Signal Processing*, vol. 2003, no. 4, pp. 321–337, 2003. 100, 102

[96] C. Meesookho, U. Mitra, and S. Narayanan, "On energy-based acoustic source localization for sensor networks," *IEEE Transactions on Signal Processing*, vol. 56, no. 1, pp. 365–377, January 2008. 100, 101

[97] Y. H. Kim and A. Ortega, "Quantizer design for energy-based source localization in sensor networks," *IEEE Transactions on Signal Processing*, vol. 59, no. 11, pp. 5577–5588, November 2011. 100

[98] D. Dardari, A. Conti, C. Buratti, and R. Verdone, "Mathematical evaluation of environmental monitoring estimation error through energy-efficient wireless sensor networks," *IEEE Transactions on Mobile Computing*, vol. 6, no. 7, pp. 790–802, July 2007. 100

[99] M. Win, A. Conti, S. Mazuelas, Y. Shen, W. Gifford, D. Dardari, and M. Chiani, "Network localization and navigation via cooperation," *IEEE Communications Magazine*, vol. 49, no. 5, pp. 56–62, March 2011. 100, 105

[100] N. Patwari, J. Ash, S. Kyperountas, A. Hero, R. Moses, and N. Correal, "Locating the nodes: Cooperative localization in wireless sensor networks," *IEEE Signal Processing Magazine*, vol. 22, no. 4, pp. 54–69, July 2005. 100, 105

[101] Y. Shen and M. Win, "Fundamental limits of wideband localization–Part I: A general framework," *IEEE Transactions on Information Theory*, vol. 56, no. 10, pp. 4956–4980, October 2010. 100, 105

[102] Y. Shen, H. Wymeersch, and M. Win, "Fundamental limits of wideband localization–Part II: Cooperative networks," *IEEE Transactions on Information Theory*, vol. 56, no. 10, pp. 4981–5000, October 2010. 100, 105

[103] H. Wymeersch, J. Lien, and M. Win, "Cooperative localization in wireless networks," *Proceedings of the IEEE*, vol. 97, no. 2, pp. 427–450, February 2009. 100, 105

[104] S. Srinivasa and M. Haenggi, "Path loss exponent estimation in large wireless networks," in *Information Theory and Applications Workshop (ITA)*, San Diego, CA, USA, February 2009, pp. 124–129. 101

[105] F. Zabini and A. Conti, "Process estimation from randomly deployed wireless sensors with position uncertainty," in *IEEE Global Telecommunications Conference (GLOBECOM)*, Houston, TX, USA, December 2011. 101

[106] D. Stoyan, W. S. Kendall, and J. Mecke, *Stochastic geometry and its applications*, 2nd ed. John Wiley and Sons, 1996. 101

[107] M. Haenggi, *Stochastic geometry for wireless networks*, 1st ed. Cambridge Unversity Press, 2012. 101

[108] M. Win, P. Pinto, and L. Shepp, "A mathematical theory of network interference and its applications," *Proceedings of the IEEE*, vol. 97, no. 2, pp. 205–230, February 2009. 101

[109] J. Andrews, R. Ganti, M. Haenggi, N. Jindal, and S. Weber, "A primer on spatial modeling and analysis in wireless networks," *IEEE Communications Magazine*, vol. 48, no. 11, pp. 156–163, November 2010. 101