

**RESOURCE AWARE DISTRIBUTED DETECTION  
AND ESTIMATION OF RANDOM EVENTS  
IN WIRELESS SENSOR NETWORKS**

by  
**ENGİN MAŞAZADE**

Submitted to the Graduate School of Engineering  
and Natural Sciences in partial fulfillment of the requirements  
for the degree of Doctor of Philosophy

Sabancı University

June 2010

**RESOURCE AWARE DISTRIBUTED DETECTION  
AND ESTIMATION OF RANDOM EVENTS  
IN WIRELESS SENSOR NETWORKS**

by Engin Maşazade

APPROVED BY

Assoc. Prof. Dr. Mehmet Keskinöz .....  
(Thesis Advisor)

Prof. Dr. Pramod K. Varshney .....  
(Thesis Co - Advisor)

Assist. Prof. Dr. Ruixin Niu .....

Assist. Prof. Dr. Güllü Kızıldağ Şendur .....

Assist. Prof. Dr. Müjdat Çetin .....

DATE OF APPROVAL: June 25, 2010

*Annem Zehra ve babam Rifat'a. Sevginiz, sabrınız ve desteęiniz için çok teřekkürler.*

©Engin Maşazade, 2010

All Rights Reserved

**RESOURCE AWARE DISTRIBUTED DETECTION  
AND ESTIMATION OF RANDOM EVENTS  
IN WIRELESS SENSOR NETWORKS**

Engin Maşazade

PhD Thesis, 2010

Thesis Advisor: Assoc. Prof. Dr. Mehmet Keskinöz

Thesis Co - advisor: Prof. Dr. Pramod K. Varshney

**Keywords:** *Distributed detection, distributed estimation, multi-objective optimization, source localization, sensor selection, wireless sensor networks, wireless communication*

In this dissertation, we develop several resource aware approaches for detection and estimation in wireless sensor networks (WSNs). Tolerating an acceptable degradation from the best achievable performance, we seek more resource efficient solutions than the state-of-the-art methods. We first define a multi-objective optimization problem and find the trade-off solutions between two conflicting objectives for the distributed detection problem in WSNs: minimizing the probability of error and minimizing the total energy consumption. Simulation results show that Pareto-optimal solutions can provide significant energy savings at the cost a slight increase in the probability of error

from its minimum achievable value.

Having detected the presence of the source, accurate source localization is another important task to be performed by a WSN. The state-of-the-art one-shot location estimation scheme requires simultaneous transmission of all sensor data to the fusion center. We propose an iterative source localization algorithm where a small set of anchor sensors first detect the presence of the source and arrive at a coarse location estimate. Then a number of non-anchor sensors are selected in an iterative manner to refine the location estimate. The iterative localization scheme reduces the communication requirements as compared to the one-shot location estimation while introducing some estimation latency. For sensor selection at each iteration, two metrics are proposed which are derived based on the mutual information (MI) and the posterior Cramér-Rao lower bound (PCRLB) of the location estimate. In terms of computational complexity, the PCRLB-based sensor selection metric is more efficient as compared to the MI-based sensor selection metric, and under the assumption of perfect communication channels between sensors and the fusion center, both sensor selection schemes achieve the similar estimation performance that is the mean squared error of the source location gets very close to the PCRLB of one-shot location estimator within a few iterations. The proposed iterative method is further extended to the case which considers fading on the channels between sensors and the fusion center. Simulation results are presented for the cases when partial or complete channel knowledge are available at the fusion center.

We finally consider a heterogenous sensing field and define a distributed parameter estimation problem where the quantization data rate of a sensor is determined as a function of its observation SNR. The inverse of the average Fisher information is then defined as a lower bound on the average PCRLB which is hard to compute. The inverse of the average Fisher information is minimized subject to the total bandwidth and bandwidth utilization constraints and we find the optimal transmission probability of each possible quantization rate. Under stringent bandwidth availability, the proposed scheme outperforms the scheme where the total bandwidth is equally distributed among sensors.

TELSİZ DUYARGA AĞLARI İCİN  
RASTLANTISAL OLAYLARIN KAYNAK DUYARLI  
DAĞINIK TESBİT VE KESTİRİMİ

Engin Maşazade

Doktora Tezi, 2010

Tez Danışmanı: Doç. Dr. Mehmet Keskinöz

Tez Ek Danışmanı: Prof. Dr. Pramod K. Varshney

**Anahtar Kelimeler:** *Telsiz duyurga ağları, dağınık tesbit ve kestirim problemleri, çok amaçlı eniyileme, duyurga seçimi, sönmülemeli kanallar, telsiz haberleşme*

Bu tezde, telsiz duyurga ağları için dağınık tesbit ve kestirim problemleri kaynak duyarlılığı altında incelenmiştir. Duyurgalar ufak, pille beslenen cihazlar olduğundan, kaynakların (enerji, bandgenişliği) tasarruflu kullanımı önemlidir. Bu doğrultuda ulaşılabilen en iyi başarımdan fazla ödün vermeden kaynaklardan önemli ölçüde tasarruf eden yöntemler sunulmaktadır. İlk olarak dağınık tesbit sorunu ele alınmıştır. Amaçların tümleştirme merkezi karar hata olasılığının ve ağın toplam enerji sarfiyatının en aza indirgenmesi olduğu bir çok amaçlı en iyileme problemi tanımlanmıştır. Bu problemlerden elde edilen sonuçlar, ulaşılabilecek en düşük hata olasılığına yakın ama önemli

ölçüde enerji tasarrufu sağlayan karar eşiklerinin de olduğunu göstermiştir.

Olayın varlığının tesbiti kadar olay yerinin hassas kestirimi türlü uygulamalar açısından önemlidir. Tüm duyarga verisini tek bir defada göndermek yerine, tekrarlı bir kestirim yöntemi sunulmaktadır. Az sayıda duyarganın verisi kullanılarak olay yeri önce kabaca kestirilir. Yöntemin bir diğer tekrarında, verisi istenecek duyargalar müşterek bilgi veya sonsal Cramér-Rao alt sınırı esaslı metrikler yardımıyla seçilmektedir. Müşterek bilgi veya sonsal Cramér-Rao alt sınırı temelli duyarga seçim metrikleri kusursuz iletim kanalları varsayımı altında benzer kestirim başarımları gösterebilir de, hesaplarımız sonsal Cramér-Rao alt sınırı temelli duyarga seçim metriğinin karmaşıklığının müşterek bilgi temelli duyarga seçim metriğine göre daha az olduğunu göstermiştir. Benzetim sonuçlarımız, tekrarlı olay yeri kestirimi yönteminin, ufak bir gecikme pahasına, kestirim için ağdaki haberleşme gereksinimini tüm duyarga verisini istemeye göre önemli ölçüde azalttığını göstermiştir. Önerilen tekrarlı olay yeri kestirim yöntemi sönümlemeli kanallar için de geliştirilmiştir. Öncelikle, kestirimde temel başarımları kısıtlı olan sonsal Cramér-Rao alt sınırı tüm duyarga verisi için türetilmiş, saydığımız iki duyarga seçim metriği tüm ve kısmi kanal bilgisi varsayımları altında yinelenmiştir.

Son olarak, ölçüm gürültüsünün her bir duyarga için farklı olduğu ayrışık durum incelenmiştir. Her bir duyarganın ölçümünü temsilde kullandığı kuantalama hızı, ölçümünün işaret-gürültü oranlarına bağlı olarak belirlenmiştir. Verilen bandgenişliği altında, genel bir dağılım kestirim problemi incelenmiştir. Toplam bandgenişliğini aşmamak için belirli bir kuantalama hızında temsil edilen ölçüm, tümleştirme merkezine yine belirli bir gönderim olasılığı ile iletilmektedir. Her bir kuantalama hızının, en uygun gönderim olasılığını bulmak için toplam bandgenişliği ve band kullanımı kısıtları altında ortalama Fisher bilgisinin tersi en aza indirgenmiştir. Benzetim sonuçları, kısıtlı bandgenişliğinde, önerdiğimiz yöntemin kestirim hatasını bandgenişliğini duyargalar arasında eşit olarak bölüştüren yöntemle göre oldukça düşürdüğünü göstermektedir.



# Acknowledgments

I would like to thank my advisor Prof. Mehmet Keskinöz for his invaluable guidance and support throughout my graduate study at Sabancı University. I am heartily thankful to my thesis co-advisor, Pramod K. Varshney, whose inspiration, encouragement, guidance and support enabled me not only to develop the understanding of the research topic but also to improve my personal skills. I would like to express the deepest appreciation to Prof. Ruixin Niu who always made time for every question I had, and I really enjoyed discussing my research with him. This dissertation would not have been possible without their guidance.

Also, I would like to thank Prof. Güllü Kızıldaş Şendur for being a defense committee member and providing an excellent research direction on Multiobjective optimization methods. I would like to thank Prof. Müjdat Çetin also for being a committee member and I appreciate him for the insights he has shared. I would like to thank Dr. Ramesh Rajagopalan for introducing me the multiobjective optimization methods, Dr. Hao Chen and Prof. Chilukuri K. Mohan for their invaluable comments and suggestions.

I would like to thank all my friends from Sabancı and Syracuse Universities. I would like to thank my friends at Sensor Fusion Lab. Priyadip, Arun, Onur, İlker, Thakshila, Swarnendu, Satish, Ashok, Renbin and Long for their wonderful friendship and the research environment. Also, I would like to thank my friends from Communications and Networking Lab., Kayhan, Ali, Mehmet and Yunus for being the best office mates. Last but not least, I would like to thank my dear friends Ertuğrul, Osman Gökhan, Can, Sarper, Selçuk, Erdem, Eray, Hadi, Gökhan, Özlem, Aslı, Hülya, Hatice

and Öznur for their invaluable friendship and support during the hard years of study. I wish you all the success, fortune and happiness in the world.

Finally, and most importantly, I would like to thank my parents Zehra and Rifat Maşazade and my sister Deniz. I would not be here without them and without their unconditional love, patience and support. I send my love to Furkan, Hilal, and Fatih whom I have missed watching them grow up.

This work is partially supported by The Scientific and Technological Research Council of Turkey (TUBITAK) under the Research Abroad Support Scheme, TUBITAK research grant under 105E161, ARO Grant W911NF-09-1-0244, and AFOSR under grant FA-9550-06-1-0277.

# TABLE OF CONTENTS

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Preliminaries . . . . .	6
1.1.1	Bayesian Detection Theory . . . . .	6
1.1.2	Estimation Theory . . . . .	8
1.1.3	Information Measures . . . . .	9
1.1.4	Monte Carlo Methods . . . . .	11
1.1.5	Multiobjective Optimization . . . . .	13
1.2	Research Motivation and Approach . . . . .	13
1.3	Major Contributions and Dissertation Organization . . . . .	17
1.4	Notes . . . . .	20
<b>2</b>	<b>A Multi-objective Optimization Approach to Obtain Decision Thresholds for Distributed Detection</b>	<b>22</b>
2.1	Problem Definition . . . . .	26
2.1.1	Wireless Sensor Network Model and Statement of the MOP . . . . .	26
2.1.2	Parallel Decision Fusion . . . . .	29
2.1.3	Serial Decision Fusion . . . . .	33
2.2	Multiobjective Optimization . . . . .	38
2.2.1	Normal Boundary Intersection (NBI) . . . . .	38
2.2.2	Non-Dominating Sorting Genetic Algorithm - II (NSGA-II) . . . . .	42
2.2.3	Performance Metrics . . . . .	45
2.3	Simulation Results . . . . .	46

2.3.1	Simulation Settings . . . . .	46
2.3.2	Performance Comparison of NBI and NSGA-II . . . . .	47
2.3.3	Optimal Pareto Fronts . . . . .	48
2.3.4	The Performance of WSN . . . . .	50
2.3.5	Identical Decision Thresholds . . . . .	53
2.4	Discussion . . . . .	56
<b>3</b>	<b>Energy Aware Iterative Source Localization</b>	<b>58</b>
3.1	System Model . . . . .	62
3.1.1	Source detection using multi-bit sensor data . . . . .	64
3.1.2	Source location estimation using multi-bit sensor data . . . . .	65
3.2	Iterative Source Location Estimation Method . . . . .	66
3.2.1	Source Location Estimation Based on Monte Carlo Methods . . . . .	69
3.2.2	Sensor Selection Methods . . . . .	70
3.3	Sensor Data Compression . . . . .	80
3.4	Simulation Results . . . . .	82
3.4.1	Source Detection Performance . . . . .	83
3.4.2	Computational Cost . . . . .	85
3.4.3	Iterative Location Estimation Performance . . . . .	86
3.4.4	The trade-off between estimation performance and communication cost . . . . .	91
3.5	Discussion . . . . .	93
<b>4</b>	<b>Channel Aware Iterative Source Localization</b>	<b>102</b>
4.1	System Model . . . . .	103
4.1.1	WSN assumptions . . . . .	103
4.1.2	CRLB of the source location estimate . . . . .	106
4.2	Iterative Source Location Estimation under Channel Fading . . . . .	108
4.2.1	PCRLB of the source location estimate . . . . .	108
4.2.2	Source localization using a Sequential Monte-Carlo method . . . . .	109
4.2.3	Mutual Information Based Sensor Selection under Channel Fading . . . . .	110

4.2.4	PCRLB Based Sensor Selection under Channel Fading . . . . .	113
4.3	Simulation Results . . . . .	116
4.3.1	Performance of the one-shot location estimator . . . . .	117
4.3.2	Performance of the iterative location estimation . . . . .	118
4.4	Discussion . . . . .	120
<b>5</b>	<b>A Probabilistic Rate Transmission Scheme for Distributed Estimation</b>	<b>126</b>
5.1	Problem Formulation . . . . .	128
5.1.1	Data rate decision of each sensor . . . . .	129
5.1.2	Determination of the quantization thresholds . . . . .	132
5.2	Probabilistic Rate Transmission . . . . .	133
5.2.1	The average FI of the estimate . . . . .	133
5.2.2	Total Rate Constraint . . . . .	136
5.2.3	Bandwidth utilization constraint . . . . .	136
5.3	Parameter Estimation based on Received Sensor Data . . . . .	139
5.3.1	The Likelihood under probabilistic rate transmission . . . . .	140
5.3.2	The Likelihood under equal rate transmission . . . . .	140
5.4	Simulation Results . . . . .	141
5.5	Discussion . . . . .	142
<b>6</b>	<b>Concluding Remarks and Suggestions for Future Work</b>	<b>144</b>
<b>A</b>	<b>Proof of Lemmas</b>	<b>147</b>
A.1	Proof of Lemma 4.1 . . . . .	147
A.2	Proof of Lemma 5.1 . . . . .	151

# LIST OF TABLES

2.1	Generational Distance between NBI and NSGA-II, Spread Metric and Mean Execution Times (E.T.) for NSGA-II and NBI. . . . .	47
2.2	Domination Metric between NBI and NSGA-II. . . . .	48
3.1	The sensor layouts to evaluate the detection performance. . . . .	83
3.2	Mean CPU times of MI and PCRLB . . . . .	86
3.3	Final MSE at the end of the 9 <sup>th</sup> iteration. $A = 1$ , PCRLB based sensor selection. . . . .	90
3.4	$A = 1$ , Average number of bits used to represent the $M = 3$ , $M = 4$ , $M = 5$ and $M = 6$ bit sensor data. . . . .	90
4.1	Estimation Improvement by increasing $M$ , $\epsilon_b = 1$ . . . . .	117
4.2	Estimation Improvement by increasing $M$ , $\epsilon_b = 5$ . . . . .	118
5.1	Optimal transmission probabilities of each possible transmission rate $R_k = j$ bits. (Conservative case: $\epsilon = 0.01$ , $\delta = 0.5$ ) . . . . .	142
5.2	Optimal transmission probabilities of each possible transmission rate $R_k = j$ bits. (Liberal case: $\epsilon = 0.1$ , $\delta = 0.7$ ) . . . . .	142

# LIST OF FIGURES

1.1	An example wireless sensor network . . . . .	3
1.2	System model for sensor and resource management based on feedback from recursive estimator. . . . .	6
2.1	Wireless Sensor Network Model with Parallel Decision Fusion . . . . .	27
2.2	Wireless Sensor Network Model with Serial Decision Fusion . . . . .	34
2.3	The Pareto optimal front found by minimizing the weighted sum of the objective functions, $N = 5$ . . . . .	39
2.4	The point P is the solution of the single-objective constrained NBI sub- problem outlined with the dashed line $v$ . . . . .	41
2.5	Contour lines of the objective functions with $N = 2$ Sensors, parallel configuration (a) Probability of Error (b) Energy Consumption . . . . .	42
2.6	Contour lines of the objective functions with $N = 2$ Sensors, serial con- figuration (a) Probability of Error (b) Energy Consumption . . . . .	43
2.7	Pareto Optimal Solutions generated via NBI and NSGA-II methods for parallel fusion and non-identical decision thresholds at each sensor . . .	49
2.8	Pareto Optimal Solutions obtained by NBI and NSGA-II methods for serial fusion and non-identical decision thresholds at each sensor. . . . .	50
2.9	Parallel Decision Fusion - Local sensor error probability as a function of its mean distance to the event location . . . . .	52
2.10	Parallel Decision Fusion - Local sensor energy consumption as a function of its distance to the fusion center . . . . .	53

2.11	Serial Decision Fusion - Global error probability as a function of the hop count on the routing path . . . . .	54
2.12	Serial Decision Fusion - Global energy consumption as a function of the hop count on the routing path . . . . .	55
2.13	Parallel Decision Fusion, Pareto Optimal Solutions for identical and non-identical sensor thresholds . . . . .	56
2.14	Serial Decision Fusion, Pareto Optimal Solutions for identical and non-identical sensor thresholds . . . . .	57
3.1	The signal intensity contours of a source located in a sensor field. . . . .	62
3.2	Wireless Sensor Network Model. Black Points: Sensor Locations; Blue Squares: Anchor Sensors used for initial iteration; Green Circles: Activated Sensors after 10 iterations for the example considered in Section V; Red Star: Source. $A = 1$ sensor is activated / iteration . . . . .	67
3.3	The flow chart of the algorithm. The dashed blocks represent the state-of-the-art Mutual Information based sensor selection method. The entire set of solid blocks represent the PCRLB based algorithm. . . . .	68
3.4	Conditional Entropy of non-anchor sensors in the field given the multi-bit decisions of the anchor sensors at the beginning of the first iteration. (a) $M = 5$ bit, (b) $M = 6$ bit . . . . .	81
3.5	Grid Sensor Layout specified by inter-sensor distance (ISD) and distance of the nearest sensor to the mean source location (D-NS-MSL). (Black points: Sensors, Red Star: Source) . . . . .	84
3.6	Detection performance of anchor sensor layouts . . . . .	85
3.7	Mean computation times of objective functions . . . . .	86
3.8	$\Gamma(1, 1)$ and $\Gamma(2, 2)$ at the first iteration as a function of $\delta$ ( $\sqrt{G} = 100$ , $M = 5$ ). . . . .	88
3.9	The MSE performance of the PCRLB based sensor selection. Comparison of Numerical computation and Gaussian approximation for the FIM of the prior. (a) $M = 3$ , (b) $M = 4$ , (c) $M = 5$ . . . . .	96



3.10	MSE at each iteration sensor selection is based on MI, PCRLB and nearest sensor to the estimated source location. (a) $M = 3$ , (b) $M = 4$ , (c) $M = 5$ , (d) $M = 6$ bits quantization . . . . .	98
3.11	(a) MSE performance of MI and PCRLB based sensor selection schemes. $N = 361$ , $K = 16$ , $M = 5$ , $A = 2$ ; (b) $M = 5$ bit quantization of each sensor measurement, MSE performance of source localization with PCRLB based sensor selection and data compression. $A = 1$ , $A = 2$ and $A = 3$ sensor activations / iteration. . . . .	99
3.12	Stopping metric vs. the number of sensors to be selected. The black line with triangle markers indicates the accuracy threshold. ( $i = 1, \epsilon = 5$ ) . . . . .	100
3.13	(a) Average number of iterations until the termination of the algorithm. (b) Average number of bits transmitted to the fusion center until the termination of the algorithm. ( $M = 5, M = 6, \epsilon = 5, 100$ different trials.)	101
4.1	Trace of the MSE matrix of $N$ sensor data, (a) $\epsilon_b = 1$ , (b) $\epsilon_b = 5$ ( $M = 3$ )	121
4.2	MSE performance of the iterative scheme using MI-based sensor selection (a) $e_c = 1$ (b) $e_c = 5$ , (The dashed line is the trace of the MSE matrix of $N = 49$ sensor data.) . . . . .	122
4.3	Mean channel gain $ h_k $ ; mean distance between the source location and the selected sensor $ d_k $ at each iteration (MI-based sensor selection). . . . .	123
4.4	Evaluation of the prior FIM as a function of $\delta$ , ( $M = 3, \text{ROI} = 20 \times 20$ )	123
4.5	MSE performance of the iterative scheme using MI and PCRLB-based sensor selection (a) $\epsilon_b = 1$ (b) $\epsilon_b = 5$ , (The dashed line is the trace of the MSE matrix of $N = 49$ sensor data.) . . . . .	124
4.6	(a) Mean channel gain and (b) Mean distance to the source location of the sensor selected at each iteration (MI and PCRLB based sensor selection). . . . .	125
5.1	Fisher information of a single sensor at various $SNR_k$ . Dashed lines represent the FI for the case where a sensor transmits analog data and $M = 6$ . . . . .	131

5.2	Optimal data rate which maximizes the FI per bit vs. $\tau_k$ . $M = 6$ , $\mu_\theta = 0$ and $\sigma_\theta^2 = 10$ . . . . .	132
5.3	MSE of probabilistic vs equal rate transmission schemes, $\mu_\theta = 0$ and $\sigma_\theta^2 = 10$ . . . . .	143

# Chapter 1

## Introduction

A wireless sensor network (WSN) consists of a large number of spatially distributed sensors that have signal processing abilities. Sensors have finite battery lifetime and thus limited computing and communication capabilities. When properly programmed and networked, sensors in a WSN cooperate to perform different tasks that are useful in a wide range of applications such as battlefield surveillance, environment and health monitoring, and disaster relief operations. Therefore, WSNs have recently been considered as an attractive low-cost technology for a wide range of surveillance and monitoring applications [1].

A WSN may be employed to monitor the occurrence of random events in a variety of applications. A random event may occur at an unknown time, at an unknown location in a region of interest (ROI) or one of its attributes (such as energy or frequency) can be random and may vary in time. However, we may have a statistical model for the event or may be able to learn it. As an example, many random arrival or counting processes follow the Poisson distribution. The time interval of interest can be of any length like seconds, minutes, or years. Examples of temporal Poisson counting processes might include the number of illegal border crossing per day or the number of earthquakes per year. Examples of spatial Poisson counting processes might include the number of people per square mile/kilometer or the number of border crossing attempts per mile/kilometer [2]. Failures represent another class of interesting events. Component failures may lead to system failures if they are not detected and corrected. For example,

a truss on a bridge may buckle or a solenoid in a printer may burn out. The Weibull distribution can be used to model the time to failure of a component, measured from some specified time until the component fails [2]. If the task of the WSN is to monitor room temperature, slight deviations from the desired temperature of the room can be modeled using a Normal (Gaussian) distribution. In a battlefield scenario, the location of a source transmitting energy by a friend or foe unit can be assumed to be uniformly distributed over the entire region of interest (ROI).

In this dissertation, we focus on distributed detection and estimation of random events in WSNs. In distributed detection, multiple sensors work collaboratively to distinguish between two hypotheses such as the absence or presence of an event. In distributed estimation, an underlying event or a specific attribute of the event is estimated based on the sensor observations. Since sensors are tiny battery powered devices with limited signal processing capabilities, prolonging the lifetime of a WSN is important for both commercial and tactical applications. With non-rechargeable batteries, this requirement places stringent energy constraints on the design of all WSN operations. Energy limitation is one of the major differences between a WSN and other wireless networks such as wireless local area networks. Also, WSNs are often self-configured networks with little or no pre-established infrastructure as well as a topology that can change dynamically. Moreover, there may be channel impairments such as fading and path loss in the network environment that can considerably degrade the quality of wireless links among sensors. Such challenges should be taken into account while designing the communication and local signal processing algorithms for the WSN. For instance, to maximize battery lifetime and reduce communication bandwidth, it is essential for each sensor to locally compress its observed data so that only low rate inter-sensor or sensor to fusion center communication is required.

The design of distributed detection and estimation algorithms depends on the underlying WSN topology. In the literature, several popular WSN deployments characterized by the presence or absence of a fusion center have been considered. In a parallel fusion topology [3], [4], [5], [6] there is no inter-sensor communication, that is the communication is only between sensors and the fusion center. The fusion center collects

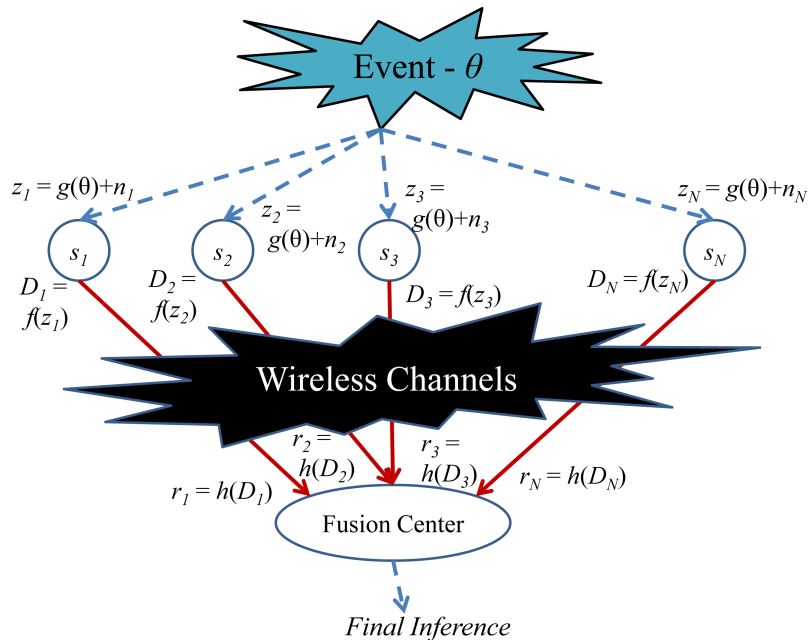


Figure 1.1: An example wireless sensor network

locally processed data and produces a final inference. In an ad hoc WSN [7], [8], there is no fusion center. The network itself is responsible for processing the collected information, and sensors communicate with each other through the shared wireless medium and arrive at a consensus. Furthermore, hybrid schemes are also possible in which the WSN is partitioned into clusters with a hierarchical structure [9], [10]. Each cluster has a local fusion center generating intermediate estimates, which are combined to obtain a final result at the global fusion center.

In this dissertation, we consider the case where sensors communicate directly with the fusion center. In Fig. 1.1, we show an example WSN with  $N$  distributed sensors where the sensor measurements or their compressed versions are directly transmitted to a central fusion center. Let  $s_k$  represent a sensor where  $k \in \{1, 2, \dots, N\}$ . A sensor receives a noisy measurement  $z_k$  from the event of interest  $\theta$  which has the form,

$$z_k = g(\theta) + n_k \quad (1.1)$$

If a sensor observes only the noisy version of the event  $\theta$  itself, then  $g(\theta) = \theta$ . If the

received signal at each sensor is subject to path loss, the observation function  $g(\cdot)$  should be defined using a suitable path loss model which we define later in the dissertation. We assume that the observation noise  $n_k$  is generated from the Gaussian distribution and is independent across sensors.

Since the sensors suffer from severe energy, computation and storage limitations, the transmission of raw measurements to the fusion center is not desirable as it incurs excessive energy and bandwidth consumption. In distributed detection or estimation, the sensors first locally process their measurements and quantized versions of the decision statistics are sent to the fusion center for making the final inference. The quantized measurement of each sensor  $D_k$  is obtained from the raw measurement  $z_k$  according to,

$$D_k = \begin{cases} 0 & -\infty < z_k < \eta_{k,1} \\ 1 & \eta_{k,1} < z_k < \eta_{k,2} \\ \vdots & \\ L-2 & \eta_{k,L-2} < z_k < \eta_{k,L-1} \\ L-1 & \eta_{k,L-1} < z_k < \infty \end{cases} \quad (1.2)$$

where  $\boldsymbol{\eta}_k = [\eta_{k,0}, \eta_{k,1}, \dots, \eta_{k,L}]^T$  is the set of quantization thresholds for  $s_k$  with  $\eta_{k,0} = -\infty$  and  $\eta_{k,L} = \infty$ .

Let  $r_k$  be the received data of each sensor at the fusion center. The quantized data of sensors are either assumed to be sent to the fusion center over perfect communication links ( $r_k = D_k$ ) [11] or over imperfect channels ( $r_k = h(D_k)$ ) where the channel impairments  $h(\cdot)$  are modeled by suitable channel fading and noise models [4], [12].

The task of the fusion center is to make an inference about the event of interest. If the fusion center is responsible to detect either the absence or the presence of the event, the problem is called the detection problem. Let  $D_0$  be the binary decision of the fusion center, which is defined as follows,

$$D_0 = \begin{cases} 0 & \text{Fusion center decides on } H_0 \\ 1 & \text{Fusion center decides on } H_1 \end{cases} \quad (1.3)$$

where Hypothesis 0 ( $H_0$ ) and Hypothesis 1 ( $H_1$ ) denote the absence and the presence of the event.

Moreover, if the fusion center is responsible for estimating an attribute of the event of interest ( $\theta$ ), the problem is called an estimation problem. Let  $\mathcal{T}$  be an estimator which is a function of the received sensor data  $\mathbf{R} = [r_1 \ r_2 \ \dots \ r_N]$ , then the estimate of the parameter is represented as  $\hat{\theta}$  and obtained as,

$$\hat{\theta} = \mathcal{T}(\mathbf{R}) \tag{1.4}$$

If the event of interest  $\theta$  is modeled as an unknown constant, under certain regularity conditions, the optimal estimator is the maximum-likelihood (ML) estimator [13]. If the unknown parameter  $\theta$  has a prior distribution, then the Bayesian estimator minimizes the Bayes risk [13]. We leave the details of the estimators for later in the dissertation.

Since a wireless sensor network consists of densely deployed tiny, battery-powered sensors, they have limited on-board energy. Therefore, if a sensor remains continuously active, its energy will be depleted quickly. In order to prolong the network lifetime, sensors should alternate between being active and idle. Note that dense deployment of sensors brings redundancy in coverage. Therefore, selecting a subset of sensors, can still provide information with the desired quality. The sensor management policies define the selection of active sensors to meet the application requirements while minimizing the use of resources [14], [15], [16], [17], [18], [19]. In other words, the problem of adaptive sensor management and resource allocation in sensor networks is to determine the optimal way to manage system resources and task a group of sensors to collect measurements for statistical inference. As shown in Fig. 1.2, the distributed sensors send their observations through band-limited channels to a fusion center, where the data are fused by an estimator to update the object state estimate. The updated state estimate based on the past data is used to guide the sensor management and resource allocation procedure adaptively. Sensor management is carried out in a way such that at the next time step, the best estimation accuracy is achieved under pre-specified resource utilization constraints.

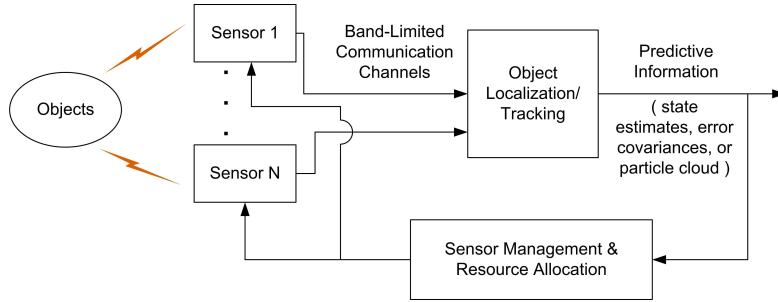


Figure 1.2: System model for sensor and resource management based on feedback from recursive estimator.

## 1.1 Preliminaries

In this section, we briefly summarize the necessary background for several topics that will be considered in the dissertation. Since we deal with detection and estimation in wireless sensor networks, we first present the fundamentals of Bayesian detection and estimation theories. Then we review information measures such as entropy and mutual information. We then discuss Monte-Carlo based methods and finally present the problem formulation of Multi-objective optimization.

### 1.1.1 Bayesian Detection Theory

Let  $H_0$  and  $H_1$  denote the two hypotheses for the binary hypothesis testing problem. Let the observation be denoted as  $z$  so that the conditional densities under the two hypotheses are  $p(z|H_0)$  and  $p(z|H_1)$  respectively. The observations are generated with these conditional densities which are assumed known. The *a priori* probabilities of the two hypotheses are denoted by  $P(H_0)$  and  $P(H_1)$  respectively. In the binary hypothesis testing problem, four possible actions can occur. Let  $C_{i,j}$ ,  $i \in \{0, 1\}$ ,  $j \in \{0, 1\}$  represent the cost of declaring  $H_i$  true when  $H_j$  is present. The Bayes risk function is given by,

$$\begin{aligned}
 \mathcal{R} &= \sum_{i=0}^1 \sum_{j=0}^1 C_{i,j} P(H_j) P(\text{Decide } H_i | H_j \text{ is present}) \\
 &= \sum_{i=0}^1 \sum_{j=0}^1 C_{i,j} P(H_j) \int_{\mathcal{Z}_i} p(z|H_j) dz
 \end{aligned} \tag{1.5}$$



where  $\mathcal{Z}_i$  is the decision region corresponding to hypothesis  $H_i$  which is declared true for any observation falling in the region  $\mathcal{Z}_i$ . Let  $\mathcal{Z}$  be the entire observation space so that  $\mathcal{Z} = \mathcal{Z}_0 \cup \mathcal{Z}_1$  and  $\mathcal{Z}_0 \cap \mathcal{Z}_1 = \emptyset$ .

Collecting the terms in (1.5) yields,

$$\begin{aligned} \mathcal{R} = & P(H_0)C_{1,0} + P(H_1)C_{1,1} + \\ & \int_{\mathcal{Z}_0} \{ [P(H_1)(C_{0,1} - C_{1,1})p(z|H_1)] - [P(H_0)(C_{1,0} - C_{0,0})p(z|H_0)] \} \end{aligned} \quad (1.6)$$

The risk is minimized by assigning those points of  $\mathcal{Z}$  to  $\mathcal{Z}_0$  that make the integrand of (1.6) negative. Then, the minimization results in the likelihood ratio test (LRT)

$$\frac{p(z|H_1)}{p(z|H_0)} \underset{<_{H_0}}{\overset{>_{H_1}}{\geq}} \frac{P(H_0)(C_{1,0} - C_{0,0})}{P(H_1)(C_{0,1} - C_{1,1})} \quad (1.7)$$

The quantity on the left hand side is known as the likelihood ratio and the quantity on the right hand side is the threshold. In this dissertation, we consider Bayes criterion for the design of decision rules. The minimax criterion [3] is a good alternative when the knowledge of a priori probabilities of each hypothesis is not available. Moreover, in many practical applications not only the *a priori* probabilities but also the cost assignments are difficult to make. For such cases, Neyman-Pearson test is employed that constrains the probability of false alarm ( $P_F$ ) while maximizing the probability of detection ( $P_D$ ). Then,  $P_F$  and  $P_D$  are defined as,

$$\begin{aligned} P_F &= P(\text{Decide } H_1 | H_0 \text{ is present}) = \int_{\mathcal{Z}_1} p(z|H_0) dz \\ P_D &= P(\text{Decide } H_1 | H_1 \text{ is present}) = \int_{\mathcal{Z}_1} p(z|H_1) dz \end{aligned}$$

A more detailed treatment of detection theory can be found in a wide variety of books such as [3] and [20].

### 1.1.2 Estimation Theory

Assume that a scalar parameter  $\theta$  to be estimated using the vector of measurements  $\mathbf{z}$ . Let  $p(\mathbf{z}; \theta)$  be the probability distribution function (pdf) of  $\mathbf{z}$  given  $\theta$  and it is assumed that  $p(\mathbf{z}; \theta)$  satisfies the regularity condition [21],

$$E \left[ -\frac{\partial \ln p(\mathbf{z}; \theta)}{\partial \theta} \right] = 0 \quad \text{for all } \theta \quad (1.8)$$

where the expectation is taken with respect to  $p(\mathbf{z}; \theta)$ . Then the variance of any unbiased estimator  $\hat{\theta}$  must satisfy,

$$\text{var}(\hat{\theta}) \geq \frac{1}{E \left[ -\frac{\partial^2 \ln p(\mathbf{z}; \theta)}{\partial \theta^2} \right]} \quad (1.9)$$

where the expectation is taken with respect to  $p(\mathbf{z}; \theta)$  and the derivative is evaluated at the true value of  $\theta$ .

Let  $\boldsymbol{\theta} = [\theta_1, \dots, \theta_N]$  be a vector parameter to be estimated. Then, the variance of each element  $\theta_i$  is found by the inverse of the Fisher information matrix defined as,

$$\text{var}(\hat{\theta}_i) \geq [F^{-1}(\boldsymbol{\theta})]_{i,i} \quad (1.10)$$

where  $i, j^{\text{th}}$  element of  $F(\boldsymbol{\theta})$  is defined as,

$$[F(\boldsymbol{\theta})]_{i,j} = E \left[ -\frac{\partial^2 \ln p(\mathbf{z}; \theta)}{\partial \theta_i \partial \theta_j} \right]$$

Let the vector of unknown parameters  $\boldsymbol{\theta}$  be a random vector with pdf  $p(\boldsymbol{\theta})$ . Then, the minimum mean squared error (MSE) estimator is the conditional mean estimator defined as,

$$\hat{\boldsymbol{\theta}}_{\text{MSE}} = \int \boldsymbol{\theta} p(\boldsymbol{\theta} | \mathbf{z}) d\boldsymbol{\theta} \quad (1.11)$$

In Bayesian estimation, the performance of any estimator  $\hat{\boldsymbol{\theta}}(\mathbf{z})$  can be bounded by the Posterior Cramér - Rao Lower Bound, under suitable regularity conditions [13]. The variance of each element  $\theta_i$  is found by the inverse of the Fisher information matrix  $\mathbf{J}$

whose  $i, j^{\text{th}}$  element is found according to,

$$\mathbf{J}_{i,j} = E \left[ -\frac{\partial^2 p(\mathbf{z}; \theta)}{\partial \theta_i \partial \theta_j} \right] \quad (1.12)$$

where the expectation is taken with respect to  $p(\mathbf{z}, \theta)$  which is the joint entropy between  $\mathbf{z}$  and  $\theta$ . The PCRLB states that,

$$\text{var}(\hat{\theta}_i) \geq [J^{-1}(\boldsymbol{\theta})]_{i,i} \quad (1.13)$$

with equality if and only if the a posteriori density  $p(\boldsymbol{\theta}|\mathbf{z})$  is a multivariate Gaussian density [13].

### 1.1.3 Information Measures

In this section, we give basic definitions and properties of several information measures such as entropy, conditional entropy and mutual information that we use in the dissertation. More details can be found in [22] and [23].

#### Entropy

The entropy or uncertainty of the discrete random variable  $X$  is,

$$H(X) = \sum_{a \in \text{sup}(P_X)} -P_X(a) \log_2 P_X(a) \quad (1.14)$$

where the support of a random variable  $X$  is the set,

$$\text{sup}(P_X) = \{a : a \in X, P_X(a) > 0\} \quad (1.15)$$

Alternatively we can write,

$$H(X) = E[-\log_2 P_X(X)] \quad (1.16)$$

## Conditional Entropy

The conditional entropy of  $X$  given the event  $Y = b$  with probability  $P(Y = b) > 0$  is,

$$\begin{aligned} H(X|Y = b) &= \sum_{a \in \text{sup}(P_{X|Y}(\cdot|b))} -P_{X|Y}(a|b) \log_2 P_{X|Y}(a|b) \\ &= E[-\log_2 P_{X|Y}(X|Y)|Y = b] \end{aligned} \quad (1.17)$$

The conditional entropy of  $X$  given  $Y$  is the average of the values (1.17), that is,

$$\begin{aligned} H(X|Y) &= \sum_{b \in \text{sup}(P_Y)} P_Y(b) H(X|Y = b) \\ &= \sum_{(a,b) \in \text{sup}(P_{X,Y})} -P_{X,Y}(a,b) \log_2 P_{X,Y}(a,b) \\ &= E[-\log_2 P_{X|Y}(X|Y)] \end{aligned} \quad (1.18)$$

## Mutual Information

The mutual information  $I(X, Y)$  between two random variables  $X$  and  $Y$  with respective discrete and finite alphabets is defined as,

$$I(X, Y) = H(X) - H(X|Y) \quad (1.19)$$

The name “mutual” describes the symmetry in the arguments of  $I(X, Y)$ , i.e.,

$$I(X, Y) = H(Y) - H(Y|X) \quad (1.20)$$

Note that,

$$\begin{aligned} I(X, Y) &\geq 0 \\ H(X|Y) &\leq H(X) \\ H(X, Y) &\leq H(X) + H(Y) \end{aligned}$$

with equality if and only if  $X$  and  $Y$  are statistically independent. The second inequality above means that conditioning reduces entropy.

#### 1.1.4 Monte Carlo Methods

Suppose, we want to compute the following integral for the test function  $\varphi$ .

$$I = \int_{\mathcal{Z}} \varphi(z) dz \tag{1.21}$$

where  $\mathcal{Z}$  represents the integration range of  $\varphi$ . To compute  $I$  numerically, Monte Carlo methods introduces a new pdf  $\pi(z)$  as

$$\begin{aligned} I &\approx I_A \\ I_A &= \int_{\mathcal{Z}} \varphi(z) \pi(z) dz \end{aligned} \tag{1.22}$$

where  $I_A$  is the numerical approximation of  $I$ . The Monte-Carlo method approximates  $\pi$  by the following point-mass measure.

$$\hat{\pi}(z) = \frac{1}{N} \sum_{i=1}^N \delta(z - z^i) \tag{1.23}$$

where  $N$  is the number of samples. Using (1.23) in (1.22) yields,

$$I_A = \int_{\mathcal{Z}} \varphi(z) \hat{\pi}(z) dz = \frac{1}{N} \sum_{i=1}^N \varphi(z^i) \tag{1.24}$$

Suppose we are interested in sampling from  $\pi(z)$  and assume that we are able to sample from another pdf  $q(z)$ . The importance sampling procedure allows us to sample from  $\pi$  using  $q(z)$ . Assuming that  $\pi(z) > 0$  and  $q(z) > 0$ , the following identity trivially holds,

$$\pi(z) = w(z)q(z) \tag{1.25}$$

where  $w(\cdot)$  is the importance weight given by,

$$w(z) = \frac{\pi(z)}{q(z)} \quad (1.26)$$

This suggests that if  $N$  samples  $\{z^i\}$  from  $q(\cdot)$  are available, then an approximation of this distribution is given by,

$$\hat{q}(z) = \frac{1}{N} \sum_{i=1}^N \delta(z - z^i) \quad (1.27)$$

Plugging this approximation in (1.25), we obtain

$$\hat{\pi}(z) = \frac{1}{N} \sum_{i=1}^N w(z^i) \delta(z - z^i) \quad (1.28)$$

Let  $w^i$  be the normalized importance weights as,

$$w^i = \frac{w(z^i)}{\sum_{j=1}^N w(z^j)} \quad (1.29)$$

Using importance sampling, (1.24) is computed as follows,

$$I_A = \sum_{i=1}^N w^i \varphi(z^i) \quad (1.30)$$

In Chapter 3, based on the available data  $\mathbf{z}$ , we are interested in approximating the posterior distribution

$$p(\boldsymbol{\theta}|\mathbf{z}) \propto p(\mathbf{z}|\boldsymbol{\theta})p(\boldsymbol{\theta})$$

Note that the initial probability of each sample  $p(\boldsymbol{\theta}^i) = 1/N$ . The weights are then selected as the likelihood of the received samples as

$$w(z^i) = p(\mathbf{z}|\boldsymbol{\theta}^i)$$

The details of Monte Carlo methods can be found in [24] and references there in.

### 1.1.5 Multiobjective Optimization

In this section, we briefly define the problem formulation of multiobjective optimization. The mathematical description of multiobjective optimization [25], [26], [27], [28], [29], [30], [31] can be given as follows:

$$\min_{\chi \in C} [f_1(\chi) \quad f_2(\chi) \quad \dots f_n(\chi)]^T \quad (1.31)$$

where  $\chi$  is a candidate solution to the multiobjective optimization problem (MOP). The number of objectives  $n \geq 2$  and the feasible set  $C$ ,

$$C : \{\chi : h(\chi) = 0, g(\chi) \leq 0, a \leq \chi \leq b\} \quad (1.32)$$

is subject to the equality and inequality constraints denoted as  $h(\chi)$  and  $g(\chi)$  respectively, and explicit variable bounds  $[a, b]$ . In a minimization problem, a solution  $\chi_1$  dominates another solution  $\chi_2$  ( $\chi_1 \gg \chi_2$ ) if and only if

$$\begin{aligned} f_u(\chi_1) &\leq f_u(\chi_2) \quad \forall u \in \{1, 2, \dots, n\} \\ f_v(\chi_1) &< f_v(\chi_2) \quad \exists v \in \{1, 2, \dots, n\} \end{aligned} \quad (1.33)$$

and a solution  $\chi^*$  is the Pareto optimal solution for the MOP if and only if there is no  $\chi \in C$  that dominates  $\chi^*$ . Pareto optimal points are also known as non-dominated points. A well known technique for solving MOPs is to minimize a weighted sum of the objective functions. Later in the dissertation we utilize several different methods to obtain the Pareto-optimal front.

## 1.2 Research Motivation and Approach

The distributed detection problem for WSNs has been studied extensively. If the fusion center receives the raw measurements of sensors, the problem is reduced to the classic hypothesis testing problem [32]. For quantized sensor data, in the temporal asymptotic regime, information theoretic frameworks have been developed [33], [34], [35] to find

the optimal decision rules based on the error exponent. For the case where the number of sensors goes to infinity, it is shown in [36] and [37] that an identical decision rule for all the sensors is asymptotically optimal.

For practical systems with limited number of sensors, the distributed detection problem can be decomposed into two inter-related problems. The first problem is to find the optimal decision rule at the fusion center. This is a relatively simple problem since the optimal fusion rule reduces to a likelihood ratio test (LRT) for binary and multi-bit sensor decisions [3]. The second problem is to obtain the decision rules at the sensors which is more complicated. Under the conditional independence assumption, the optimal decision rule at each sensor is expressed as an LRT [3]. Since the decision rules at distributed sensors and the fusion center are dependent on each other, person-by-person optimization (PBPO) is often used to obtain the optimal decision thresholds of sensors [38]. Many papers in the literature, assume ideal channels between the sensors and the fusion center. In [39], [4], [6], [40], non-ideal channels have been assumed between the sensors and the fusion center in the distributed detection context. Without the conditional independence assumption, the distributed detection problem becomes very hard [41], [42], [43].

In this dissertation, we first study the event detection problem for sensor networks under the isotropic signal emission model [44], [45] where the source location is assumed to be uniformly distributed in a given ROI. Given the source location and the assumption of independent identical noise distribution at each sensor, the optimal decision rules at the sensors and the fusion center are LRTs. When the source location is random and available only in terms of its probability distribution, the conditional independence assumption of sensor measurements and the optimality of LRT are no longer valid [45]. We assume that each sensor arrives at a binary decision about the event by comparing locally computed decision statistic with its decision threshold. The binary decision is then transmitted to the fusion center only if the presence of the event is decided [46], [47]. Therefore, the decision rules used at the sensors not only determine the decision error probability achieved by the WSN but also the total energy consumption of the WSN. In order to find the sensor decision thresholds, we formu-



late a multi-objective optimization problem (MOP) with two objectives, minimizing the probability of error at the fusion center and minimizing the total network energy consumption. Using a multi-objective optimization approach, we seek those solutions which provide significant energy savings as compared to the minimum error solution at potentially the cost of a slight increase in the best achievable probability of error of the network.

After the presence of the source emitting energy is detected by the WSN, an important task that needs to be performed is source localization, which is important for an accurate tracking of the target and higher level motion analysis. Under the isotropic signal emission model, the event (or source) location can be determined based on the energy readings of sensors [48], [44]. In [48] and [44], maximum likelihood (ML) based source localization approaches have been proposed by using analog and multi-bit ( $M$ -bit) sensor measurements respectively at the fusion center. Furthermore, in [49], the authors propose a joint detection and source localization scheme using the data received from all the sensors in the network. We call the source localization scheme which requires simultaneous data transmission from all the sensors to the fusion center as one-shot location estimation. One-shot location estimation introduces several challenges. First of all, the sensors that are far from the source location are not likely to carry useful information but they still consume energy to transmit information. Secondly, each sensor requires an independent channel for simultaneous data transmission to the fusion center. This assumption imposes a limitation on the number of sensors that the system can support in practice. In our approach, we assume that the source location is random and characterized by a multivariate Gaussian distribution whose covariance matrix is large so as to cover the entire ROI. In our model, rather than transmitting multi-bit data from all the sensors in the network to the fusion center, we first employ measurements from relatively few anchor sensors to detect the presence of the source and obtain a coarse location estimate. The non-anchor sensors do not transmit their measurements in the initial phase. Then, a few non-anchor sensors are activated at each step of an iterative procedure. Since only the most informative sensors about the source location are selected, the iterative algorithm is expected to provide significant

energy savings as compared to one-shot location estimation at the cost of some latency. Since source location is a random parameter which has a certain prior pdf, we consider posterior Cramer Rao lower bound (PCRLB) as the estimation benchmark for the mean squared error (MSE).

The lossless communication assumption between sensors and the fusion center is often not valid in practice. Since WSNs are resource constrained in terms of bandwidth and energy, increasing the transmission power of sensors or employing powerful error correction codes to ensure lossless communication may not always be feasible. Also, in a hostile environment, the power of the transmitted signal should be kept low to decrease the probability of interception or detection. Therefore, the iterative source localization method also helps in dealing with the channel impairments.

So far, we have assumed that the WSN is homogenous, i.e., the observation noise of each sensor is independent and identically distributed and all the sensors send the same amount of information to the fusion center. Next, we investigate a heterogenous WSN where the observation noise of each sensor is independent but not identically distributed [50] and depending on the quality of sensor observations, each sensor transmits different amount of data to the fusion center [51]. We consider a distributed random parameter estimation problem under a total bandwidth constraint. In the literature, the total rate-constrained distributed estimation problem has been investigated extensively (see [52] and references therein). Under rate constraints, the source coding problem has been studied by deriving the information theoretic achievable rate regions in [5]. If no prior is assumed for the estimation parameter, then the dynamic range of the parameter is assumed to be bounded within a certain interval. For such cases linear decentralized estimation schemes have been proposed for homogenous environments [53] and for heterogenous environments [51], [50]. Moreover, for 1-bit sensor data, [54] investigates the performance limit of distributed estimation systems where the dynamic range of the estimation parameter is assumed to be known. Also in [55], the authors assume that the sensor observations are bounded and they propose nonparametric distributed estimators based on the knowledge of the first  $N$  moments of sensor noises. Different from the papers discussed so far, we assume that the parameter to be estimated follows

a certain prior probability distribution, which requires a Bayesian estimator to be employed at the fusion center and PCRLB is computed as a benchmark for estimation. In a heterogenous network, the complexity to compute PCRLB is high, which motivates us to find another lower bound on the MSE. Since we assume that the total bandwidth is limited, so as not to exceed the total bandwidth, each sensor sending data at a specific data rate employs a certain transmission probability to send data to the fusion center. Using this approach of only requesting data from the sensors with more informative observations, we show that an estimation performance close to having all sensor data can be obtained. Similar probabilistic approaches for resource-constrained distributed estimation have been recently introduced in [56], [57]. In such approaches, each sensor measurement is transmitted to the fusion center with a certain probability so the total cost of information transmission from sensors to the fusion center does not exceed the available capacity. In [56], the authors have employed a channel-aware transmission control where the transmission probability of each sensor is chosen according to the quality of its local observation and transmission channels. In [57], the optimal transmission rates have been obtained by minimizing the posterior Cramer-Rao lower bound (PCRLB) under a total energy constraint. We follow a similar probabilistic scheme, and assume that each sensor transmits its data with a certain transmission probability per each quantization data rate. Given the number of sensors in the network and total available bandwidth, the transmission probabilities of each quantization rate minimizes the inverse of the Fisher information.

### 1.3 Major Contributions and Dissertation Organization

In this dissertation, resource aware distributed detection and estimation of random events in WSNs are investigated. We develop novel distributed detection and estimation schemes which can significantly save resources in terms of energy, communication and bandwidth while achieving a similar performance as the state-of-the-art detection/estimation methods at the cost of potentially slight increase in the probability of decision error, estimation latency and outage probability.

In Chapter 2, we study the distributed detection problem for WSNs where the

source location is assumed to be uniformly distributed in a ROI. We formulate a multi-objective optimization problem (MOP) with two conflicting objectives, minimizing the probability of error at the fusion center and minimizing the total network energy consumption. The decision thresholds at the sensors are selected as the optimization parameters of the MOP. We solve the MOP and generate the Pareto optimal solutions between these two conflicting objectives through Normal Boundary Intersection (NBI) [25] and Non Dominating Sorting Genetic Algorithm II (NSGA - II) [28]. Simulation results show that, instead of minimizing the global probability of error only, the proposed MOP approach provides a number of alternative solutions which are able to provide significant energy savings as compared to the minimum error solution at the cost of a slight increase in the minimum achievable probability of error of the network.

In Chapter 3, we study the source localization problem for a homogenous WSN where the observation noise is independent and identically distributed for each sensor and all the sensors send the same amount of information to the fusion center. The source location is random and modeled using a multivariate Gaussian distribution whose covariance matrix is large so as to cover the entire ROI. We present an iterative source localization method where rather than transmitting complete sensor data to the fusion center from all the sensors, the anchor sensors first detect the source and obtain a coarse source location estimate. Then, we develop and compare two different sensor selection schemes for static source localization. The first scheme iteratively activates the non anchor-sensors which maximize the mutual information between source location and the quantized sensor measurements. In the second sensor selection scheme, a number of non-anchor sensors are activated whose quantized data minimize the PCRLB at each iteration. Further, using the posterior probability distribution function of the source location, we compress the quantized data of each activated sensor using distributed data compression techniques. Simulation results show that the MI and PCRLB based sensor selection schemes, within a few iterations achieve similar estimation performance and get close to the PCRLB for the case when all the sensor data are used. The PCRLB-based sensor selection is better in terms of computational complexity when the number of non-anchor sensors selected at each iteration is greater than one. By selecting only

the most informative sensors about the source location, the iterative approach provides large energy savings as compared to one shot location estimation while introducing some latency.

In Chapter 4, we extend the iterative source localization method for the case where the channels between sensors and the fusion center are subject to Rayleigh fading. Considering phase coherent reception and using the channel gain statistics, we first derive the likelihood of the  $M$ -bit symbols of a sensor received over a fading channel. Simulation results show that source location estimation using the channel gain statistics yield performance that is quite close to the case where each sensor's channel gain is known exactly. We then extend the mutual information and PCRLB based sensor selection metrics that include channel fading. When the channel signal-to-noise ratio (SNR) is relatively high between sensors and the fusion center, the mean squared error of the iterative algorithm, in a few iterations gets close to the mean squared error when all  $N$  sensor data is available at the fusion center. On the other hand, if the channel SNR is low, then each selected sensor becomes less informative about the source location and the iterative sensor selection needs several iterations to reach the mean squared error of the case where data from all the  $N$  sensors is available.

In Chapter 5, we study a distributed parameter estimation problem for a heterogonous WSN where the observation noises of the sensors are Gaussian with non-identical statistics. The fusion center is unaware of the quality of the sensor observations and each sensor quantizes its measurement to the rate which improves its Fisher information per bit the most. For a heterogonous WSN, the complexity to compute the average PCRLB is high. To reduce the complexity associated with the PCRLB, we first show that the inverse of average Fisher information is a lower bound on the average PCRLB. From the previous chapter, we observe that the quantized sensor measurements become more informative as the wireless channel impairments are suppressed by increasing the energy per bit. In this chapter, we neglect the channel impairments for multi-bit sensor data, and assume the wireless channels between sensors and the fusion center are error free which can be provided by orthogonal channels with sufficient transmit power or powerful forward error correction. At the same time, we consider that the channels

between sensors and the fusion center can reliably transmit up to  $B$  bits information. So, not to exceed the total bandwidth ( $B$ ), the observation of each sensor quantized with the rate computed as above is sent to the fusion center with a certain transmission probability. To find the optimal transmission probabilities of each possible data rate of a sensor, we formulate a constrained optimization problem by minimizing the inverse of the average Fisher information while taking the total bandwidth and network utilization constraints into account. Under stringent constraint on available bandwidth, simulation results show that the proposed probabilistic scheme, overcomes the scheme where the total bandwidth is equally distributed among all sensors in the network. Instead of all sensors transmitting at high data rates which requires a large bandwidth, the proposed probabilistic bit transmission scheme obtains a similar MSE by requesting data at high rates only from the sensors with high SNR.

In Chapter 6, we summarize the main results of the dissertation and present suggestions for some future work.

#### 1.4 Notes

We make use of the standard notational conventions. Vectors and matrices are written in boldface and all vectors are column vectors. For a matrix  $\mathbf{A}$ ,  $\mathbf{A}^T$  indicates the transpose operation. The notation  $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  means that vector  $\mathbf{x}$  is Gaussian distributed with mean vector  $\boldsymbol{\mu}$  and covariance matrix  $\boldsymbol{\Sigma}$ . Also, throughout the dissertation, we denote the probability mass function of discrete variables by  $P(\cdot)$  and the probability density function of continuous variables by  $p(\cdot)$  or  $p(\cdot, \cdot)$  depending on the number of random variables.

Portions of the material in this dissertation have been presented at the 2008 IEEE Asilomar Conference on Signals, Systems, and Computers [58], the 2009 International Workshop on Computational Advances in Multi-Sensor Adaptive Processing [59], the 2010 Conference on Information Sciences and Systems [60] and accepted for presentation at the 2010 International Conference on Information Fusion [61]. Additionally, portions of the material have appeared in or have accepted to appear in the IEEE Transactions on Systems, Man, and Cybernetics, Part B [62] and IEEE Transactions on Signal

Processing [63].

## Chapter 2

# A Multi-objective Optimization Approach to Obtain Decision Thresholds for Distributed Detection

In this chapter, we study the detection problem where the objective of the WSN is to distinguish between two hypotheses, such as the absence (Hypothesis 0) or presence (Hypothesis 1) of a certain event. Such detection ability of a WSN is crucial for various applications. As an example, in a surveillance scenario the presence or absence of a target is usually determined, before attributes such as its position or velocity are estimated [37].

In distributed detection, by taking advantage of the limited onboard signal processing capabilities of sensors, the measurements are first preprocessed and a quantized version of the decision statistic is sent to the fusion center. For binary quantization and under different performance criteria (Bayes, Neyman-Pearson (NP)), the design of the optimal fusion rule is relatively straightforward but the evaluation of the decision thresholds at peripheral sensors is more complicated as a result of the distributed nature of the WSN. Therefore, obtaining local sensor decision rules is a major issue in the distributed detection problem [3].

For a given number of sensors and under the assumption of conditionally independent observations, the optimal decision rule at each sensor reduces to a likelihood ratio test (LRT) [3] for both Bayesian and NP criteria and different decision fusion



topologies such as parallel or serial. In parallel decision fusion, each sensor sends its decision directly to the fusion center whereas in serial decision fusion, all the sensors are connected in series. The routing path defines how these sensors are inter-connected and in this work we assume that it is known in advance. In the serial case, we assume that each sensor generates its decision by combining the decision coming from its predecessor with its own measurement. Then, the decision of the last sensor on the path is accepted as the final inference. Under decision fusion schemes for both topologies, the LRTs at each sensor are coupled with other sensor decisions and the fusion rule. Optimal values of the local sensor thresholds are typically found using Person by Person Optimization (PBPO) [3], where each sensor threshold is optimized iteratively by assuming a fixed fusion rule and decision rules at the other sensors. In the asymptotic regime where the number of sensors is very large, an identical decision rule for all the sensors is asymptotically optimal [36]. This result simplifies the design of decision rules considerably.

In this chapter, we assume ideal channels between the sensors and the fusion center (for recent work involving non-ideal channels, see [39], [6], [40]). Under the NP criterion and considering fading channels between sensors and the fusion center, an exhaustive search has been employed in [6] over all threshold selections to determine their optimal values. Computational complexity of such an approach increases exponentially with the number of local sensors and this approach for finding the optimal sensor thresholds is practical only with relatively few sensors. We assume that each sensor arrives at a binary decision about the event by comparing its decision statistic with a threshold. If the sensor decides positively about the presence of the event, it transmits one bit, otherwise it stays silent. To ensure perfect communication, each sensor decision should be transmitted with sufficient energy which is a function of the distance between the sensor and the fusion center [64]. Therefore, the thresholds of local sensors not only determine the network's probability of error, but also affect the total energy consumption.

A recent work [47] considers the design of local sensor decision rules that minimize the probability of error subject to a transmission rate constraint for each sensor. Under

conditionally independent observations, a constrained minimization problem is defined and the optimal thresholds are obtained using the well known PBPO procedure. Although conditional independence assumption simplifies the derivation of decision rules, it may not be valid in many realistic cases such as when the location of the event isn't known exactly. If the location of the event can only be described in terms of its probability density function, the received sensor decisions are no longer conditionally independent because of the unknown event location. Then the optimality of LRTs for local sensor decision making fails and the derivation of optimal sensor decision rules becomes complicated. In this chapter, we consider the case where the event has an isotropic signal emission with path loss [65], [66]. Then in the presence of the event, each sensor's measurement depends on the distance between the sensor and the event location. Each noisy sensor measurement then follows the same probability distribution with different means as long as the measurement noise is independent and identically distributed across sensors. The sensors in proximity of the event decide more likely to decide on the presence of the event. In other words, an isotropic signal source for the event implies a high degree of spatial correlation. A related work [67] proposes a collaborative detection scheme where a sensor close to the signal source requests collaboration and receives the decisions of the  $K_{max}$  sensors within its neighborhood. The authors showed that increasing  $K_{max}$ , namely including more sensors to the collaboration that are located far from the event degrades the detection performance considerably.

Sensor network design usually involves simultaneous consideration of multiple conflicting objectives [6], [68], such as maximizing the lifetime of the network or maximizing the detection capability, while minimizing the transmission costs. In a conventional WSN setting, one of the desired objectives is optimized while treating others as constraints of the problem or the problem is converted into a single objective problem by assigning weights to each objective function. In the constrained minimization case, one single solution is obtained based on available resource limitations and the solution has to be reevaluated for each time when the amount of resource has been changed. In the weighted sum approach, relative weights of the objectives are usually not known or difficult to determine. These drawbacks can be overcome via multi-objective optimization

methods [25], [26], [27], [28], [29], [30], [31] which optimize all the objectives simultaneously and generate a set of solutions at the same time reflecting different trade-offs between the objectives. Multiobjective optimization has recently been introduced for WSN design [69] where the mobile agent routing and sensor placement problems and the tradeoff solutions between the desired objectives were determined through the use of multi-objective optimization based on evolutionary algorithms.

In this chapter, we study the event detection problem for sensor networks under isotropic signal emission and the event location is only known in terms of its probability density function. Also, we assume that sensors employ the on-off keying scheme where they send one bit data to the fusion center only if they decide on the presence of the event. Then, sensor decision thresholds not only determine the probability of error but also determine the total energy consumption of the network. So, instead of having a single solution that minimizes the probability of error of the network, by using the multi-objective optimization approach, we seek several sensor threshold sets which deliver significant energy saving as compared to the energy consumption of the minimum probability of error solution without sacrificing probability of error too much. Thus, we are able to obtain a set of solutions which provide tradeoffs between energy consumption and probability of error performance.

Hence, we formulate a multi-objective optimization problem (MOP) with two objectives, minimizing the probability of error at the fusion center (global probability of error)  $P_e$  and minimizing the total network energy consumption (global energy consumption)  $E_T$  where the sensor decision thresholds are selected as the variables of the MOP. We solve the MOP and generate the Pareto optimal solutions between these two conflicting objectives through Normal Boundary Intersection (NBI) [25] and Non-Dominating Sorting Genetic Algorithm II (NSGA - II) [28]. In this chapter, we first study the problem for parallel decision fusion where each sensor performed binary quantization by comparing its measurement with its threshold. We then compare the results of parallel decision fusion with serial decision fusion. In the serial case, it is hard to evaluate the optimal decision rule of each sensor since the event location is known only in terms of its probability density function. Simulation results show that when

each sensor makes its decision based on the decision of its predecessor and its own observation, the performance is poor if the sensor is very far away from the event location. For this reason, motivated by the counting rule considered in [49], we use a heuristic decision rule at each sensor. Our decision statistic used for the serial case is the aggregation of sensor decisions from all the previous sensors and its own observation. In this work, we also compare the multi-objective optimization methods NBI and NSGA-II in detail by using the performance metrics, generational distance, domination and spacing metrics described in [29]. Finally, we compare the performance of the network both for different and identical sensor thresholds employed at each sensor.

The rest of the chapter is organized as follows. In Section 2.1, we state the WSN assumptions and describe each objective function under both parallel and serial decision fusion schemes. In Section 2.2, we review the fundamentals of MOP, describe NBI and NSGA-II methods. In Section 2.3, we present our simulation results and finally devote Section 2.4 to discussion of the results.

## 2.1 Problem Definition

In this section, we first state the wireless sensor network assumptions, then we define the mathematical models for both objective functions for parallel and serial decision fusion topologies.

### 2.1.1 Wireless Sensor Network Model and Statement of the MOP

A representative wireless sensor network consisting of  $N$  sensors,  $\{s_k, i = 1, 2, \dots, N\}$  with parallel decision fusion is shown in Figure 2.1. The distances between  $s_k$  and the fusion center and the event location  $(x, y)$  are denoted as  $d_{f,k}$  and  $d_k$  respectively. We assume the event location to be a random variable with an associated prior probability density function (pdf) and, therefore,  $d_k$  is a random variable.

Specifically, we assume that the location of the event is uniformly distributed with joint pdf,

$$p(x, y) = \frac{1}{A \times B}, \quad (2.1)$$

$$0 \leq x \leq A, \quad 0 \leq y \leq B$$

where the region of interest (ROI) is an area of size  $A \times B$ . Other pdfs can be employed in a similar manner. The average distance of  $s_k$  located at  $(x_k, y_k)$  to the event location  $(x, y)$  is then expressed as,

$$\bar{d}_k = \int_0^A \int_0^B \sqrt{(x - x_k)^2 + (y - y_k)^2} p(x, y) dy dx \quad (2.2)$$

Suppose that a signal that follows the power attenuation model such as an acoustic signal is radiated from an event source with energy  $P_0$  [65] and sensors  $s_k, i = 1, 2, \dots, N$  are deployed at positions  $(x_k, y_k), i = 1, 2, \dots, N$ . Then, the received energy ( $e_k$ ) observed at  $s_k$  is,

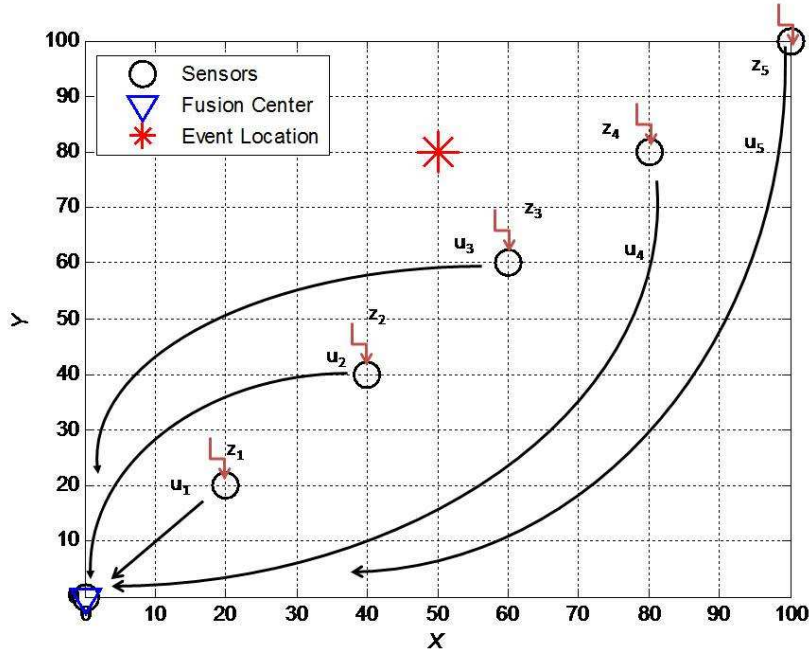


Figure 2.1: Wireless Sensor Network Model with Parallel Decision Fusion

$$e_k(x_k, y_k, x, y) = \begin{cases} P_0 & d_k \leq d_0 \\ P_0 \left(\frac{d_0}{d_k}\right)^n & \text{otherwise} \end{cases} \quad (2.3)$$

where  $n$  is the signal decay exponent and  $d_0$  is the reference distance where we select  $d_0 = 1m$ . When  $n = 2$ , the energy of the event decays at a rate inversely proportional to the square of the distance  $d_k = \sqrt{(x - x_k)^2 + (y - y_k)^2}$ . Then, under each hypothesis, the received measurement of each sensor ( $z_k$ ) can be expressed as,

$$z_k = n_k, \quad \text{under } H_0 \quad (2.4)$$

$$z_k = \sqrt{e_k(x_k, y_k, x, y)} + n_k, \quad \text{under } H_1$$

where  $n_k$  is the measurement noise that follows normal distribution at each sensor and it is assumed to be independent across the sensors.  $z_k$  then follows a normal distribution with parameters,

$$z_k \sim \begin{cases} N(0, \sigma^2) & \text{under } H_0 \\ N(\sqrt{e_k(x_k, y_k, x, y)}, \sigma^2) & \text{under } H_1 \end{cases} \quad (2.5)$$

Throughout the chapter, we assume that the noise variance,  $\sigma^2$  is unity. When  $z_k$  exceeds a certain threshold denoted as  $t_k$ , sensor  $s_k$  transmits a one bit decision ( $D_k = 1$ ) to the fusion center. Otherwise, it does not transmit anything.

The functions global probability of error  $P_e$  and global energy consumption  $E_T$  are functions of the local sensor thresholds  $t_k$  and constitute the objective functions of the MOP. The MOP considered here is formulated as follows,

$$\min_{t_1, t_2, \dots, t_N} \{P_e(t_1, t_2, \dots, t_N), E_T(t_1, t_2, \dots, t_N)\}, \quad (2.6)$$

$$t_{min} \leq t_k \leq t_{max} \quad i \in \{1, 2, \dots, N\}.$$

We first solve the above problem for  $N$  nonidentical decision thresholds  $\{t_1, t_2, \dots, t_N\}$  employed at each sensor. We also compare the performance of nonidentical decision thresholds with identical decision threshold at each sensor  $\{t = t_1 = \dots = t_N\}$  via simulation.

In the next subsections, we derive the objective functions for the global probability of error and the global energy consumption under parallel and serial decision fusion

models.

### 2.1.2 Parallel Decision Fusion

In this subsection, we derive mathematical expressions for the two objectives namely global probability of error and the global energy consumption for parallel decision fusion.

#### Global Probability of Error

Let  $D_0$  be the global decision at the fusion center about the presence or absence of an event, and  $P(H_0)$  and  $P(H_1)$  be the *a priori* probabilities of  $H_0$  and  $H_1$  respectively. The global probability of error is given by [3],

$$P_e = P(H_0)P_F + P(H_1)(1 - P_D) \quad (2.7)$$

where  $P_F = P(D_0 = 1|H_0)$  denotes the global probability of false alarm, and  $P_D = P(D_0 = 1|H_1)$  denotes the global probability of detection. Given the vector of local sensor decisions of size  $1 \times N$ ,  $\mathbf{D} = [D_1 \ D_2 \ \dots \ D_N]$  and  $D_k \in \{0,1\}$ , the probability of error is expressed as

$$P_e = P(H_0)P(D_0 = 1|H_0) + P(H_1)(1 - P(D_0 = 1|H_1)) \quad (2.8)$$

which can be written as,

$$P_e = P(H_1) + P(D_0 = 1|\mathbf{D})[P(H_0)P(\mathbf{D}|H_0) - P(H_1)P(\mathbf{D}|H_1)]$$

$P_e$  is minimized if,

$$P(D_0 = 1|\mathbf{D}) = 0 \text{ when } [P(H_0)P(\mathbf{D}|H_0) - P(H_1)P(\mathbf{D}|H_1)] > 0 \quad (2.9)$$

$$P(D_0 = 1|\mathbf{D}) = 1 \text{ when } [P(H_0)P(\mathbf{D}|H_0) - P(H_1)P(\mathbf{D}|H_1)] < 0$$

The above property leads to the following likelihood ratio test (LRT) at the fusion center [3],

$$\frac{P(\mathbf{D}|H_1)}{P(\mathbf{D}|H_0)} \underset{D_0=0}{\gtrless} \underset{D_0=1}{\frac{P(H_0)}{P(H_1)}} \quad (2.10)$$

By conditioning  $P_F$  over each possible incoming vector of decisions  $\mathbf{D}$  and then averaging over  $\mathbf{D}$ ,  $P_F$  is expressed as,

$$P_F = P(D_0 = 1|H_0) = \sum_{\text{all } \mathbf{D}} P(D_0 = 1|\mathbf{D})P(\mathbf{D}|H_0) \quad (2.11)$$

where according to the received decision vector  $\mathbf{D}$ ,  $P(D_0 = 1|\mathbf{D})$  is either zero or one based on the fusion rule expressed in Eq.(2.10). Since the noise samples are assumed to be independent and identically distributed,

$$P(\mathbf{D}|H_0) = \prod_{k=1}^N P(D_k|H_0) \quad (2.12)$$

where the false alarm probability of an individual sensor  $P_{F,k}$  is,

$$P_{F,k} = P(D_k = 1|H_0) = Q(t_k) \quad (2.13)$$

where  $Q(\cdot)$  is the complementary distribution function of the Gaussian defined as,

$$Q(t_k) = \int_{t_k}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \quad (2.14)$$

Since the event location is random,  $P(\mathbf{D}|H_1)$  can not be written directly as the product of individual decisions as in Eq.(2.12). Instead, the global probability of detection needs to be first conditioned on the location of the event, and then needs to be averaged over its probability density function. For a given event location  $(x, y)$ , the conditional global probability of detection  $CP_D$  is,

$$CP_D = P(D_0 = 1|x, y, H_1) = \sum_{\text{all } \mathbf{D}} P(D_0 = 1|\mathbf{D})P(\mathbf{D}|x, y, H_1) \quad (2.15)$$

and since the noise distribution is independent across sensors,



$$P(\mathbf{D}|x, y, H_1) = \prod_{k=1}^N P(D_k|x, y, H_1) \quad (2.16)$$

where the conditional probability of detection of an individual sensor  $CP_{D,k}$  under given event location  $(x, y)$  is expressed as,

$$CP_{D,k} = P(D_k = 1|x, y, H_1) = Q(t_k - \sqrt{e_k(x_k, y_k, x, y)}) \quad (2.17)$$

and  $P(D_k = 0|x, y, H_1) = 1 - P(D_k = 1|x, y, H_1)$ . Also, the error probability of an individual sensor  $P_{ind,k}(t_k)$  as a function of its decision threshold  $t_k$  can be expressed as,

$$P_{ind,k}(t_k) = P(H_0)P(D_k = 1|H_0) + P(H_1) \int_0^A \int_0^B P(D_k = 0|x, y, H_1) dy dx \quad (2.18)$$

The global detection probability  $P_D$ , is found by averaging  $CP_{D,k}$  over the probability density function of the event location as,

$$P_D = \int_0^A \int_0^B P(D_0 = 1|x, y, H_1) p(x, y) dy dx \quad (2.19)$$

Our first objective function, the probability of error, is given by (2.20),

$$P_e(t_1, t_2, \dots, t_N) = P(H_0) \sum_{all \mathbf{D}} P(D_0 = 1|\mathbf{D}) P(\mathbf{D}|H_0) + P(H_1) \left[ \int_0^A \int_0^B \sum_{all \mathbf{D}} P(D_0 = 0|\mathbf{D}) P(\mathbf{D}|x, y, H_1) p(x, y) dy dx \right] \quad (2.20)$$

## Global Energy Consumption

In this section, we employ an energy efficient on-off keying scheme where only the sensors that detect the event transmit their decision to the fusion center. We also

assume that the transmitted local decisions are delivered to the fusion center without any error. Then the energy consumption at sensor  $s_k$  for transmitting  $m$  bits perfectly to the fusion center over distance  $d_{f,k}$  defined as [64]

$$E_{TX}(m, d_{f,k}) = E_{elec} \times m + \epsilon_{amp} \times m \times d_{f,k}^2 \text{ [Joules]}. \quad (2.21)$$

According to this model, a sensor dissipates  $E_{elec} = 50$  nJ/bit to run the transmitter circuitry and  $\epsilon_{amp} = 100$  pJ/bit/ $m^2$  for the transmitter amplifier.

The energy consumption of the network is the total transmission energy of all single bit decisions transmitted to the fusion center. In other words, in (2.21),  $m$  becomes one if  $D_k = 1$  and  $m$  is zero (no transmission) if  $D_k = 0$ . An *individual sensor's* energy consumption can be expressed as,

$$E_{ind,k}(t_k) = E(1, d_{f,k})[P(D_k = 1|H_0)P(H_0) + P(D_k = 1|H_1)P(H_1)] \quad (2.22)$$

Transmission of the decision vector  $\mathbf{D}$  to the fusion center then requires,

$$E_C(\mathbf{D}) = \sum_{k=1}^N E_{TX}(D_k, d_{f,k}) \quad (2.23)$$

The energy consumption  $E_T$  of the network is then found by conditioning  $E_C(\mathbf{D})$  on all possible vector of decisions as,

$$\begin{aligned} E_T(t_1, t_2, \dots, t_N) &= \sum_{\text{all } \mathbf{D}} E_C(\mathbf{D})P(\mathbf{D}) \\ &= \sum_{\text{all } \mathbf{D}} E_C(\mathbf{D})(P(\mathbf{D}|H_0)P(H_0) + P(\mathbf{D}|H_1)P(H_1)) \end{aligned} \quad (2.24)$$

Using the relation,

$$P(\mathbf{D}|H_1) = \int_x \int_y \left[ \prod_{k=1}^N P(D_k|x, y, H_1) \right] p(x, y) dy dx \quad (2.25)$$

together with Eq.(2.12) and Eq.(2.24), our second objective, global energy consumption is obtained as,

$$E_T(t_1, t_2, \dots, t_N) = \sum_{\text{all } \mathbf{D}} E_C(\mathbf{D}) [P(H_0) \prod_{k=1}^N P(D_k|H_0) + P(H_1) \int_x \int_y [\prod_{k=1}^N P(D_k|x, y, H_1)] p(x, y) dy dx] \quad (2.26)$$

### 2.1.3 Serial Decision Fusion

In this section, we derive mathematical expressions for the probability of error and the total energy consumption for serial decision fusion. In the serial fusion scheme, as described earlier and shown in Figure 2.2, the decision of  $s_k$  is a function of its own measurement from the event  $z_k$  and the decisions of its predecessors  $\mathbf{D}_{k-1} = [D_1, \dots, D_{k-1}]$ . The aggregate decision  $D_k$  is then forwarded to the successor sensor together with the past decisions  $\mathbf{D}_{k-1}$ . The last sensor in the serial configuration takes the final decision which is a binary value that represents either of the two hypotheses. We assume that the routing path is known to all the sensors.

#### Probability of Error

In serial topology, the decision of the  $N^{\text{th}}$  sensor  $D_N$  is the decision of the entire WSN. Therefore, the probability of error is expressed as,

$$P_e = P(H_0)P(D_N = 1|H_0) + P(H_1) \int_x \int_y P(D_N = 0|x, y, H_1)p(x, y) dy dx \quad (2.27)$$

where  $P_F = P(D_N = 1|H_0)$  denotes the probability of false alarm, and  $P_D = 1 - P(D_N = 0|H_1)$  denotes the probability of detection. In order to calculate these two quantities, the decision of  $s_N$  should be both conditioned on the received measurement  $z_N$  and the decisions of all its predecessors  $\mathbf{D}_{N-1}$  as shown in Eq.(2.28).

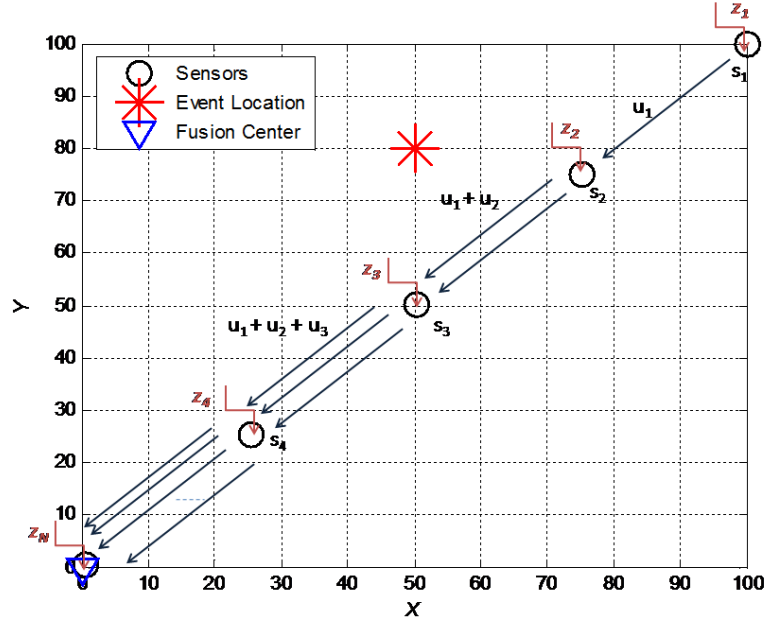


Figure 2.2: Wireless Sensor Network Model with Serial Decision Fusion

$$P_{F,N} = \sum_{\text{all } \mathbf{D}_{N-1}} \int_{z_N} P(D_N = 1 | \mathbf{D}_{N-1}, z_N, H_0) P(\mathbf{D}_{N-1}, z_N | H_0) dz_N \quad (2.28)$$

$$P_{D,N} = \int_x \int_y \left( \sum_{\text{all } \mathbf{D}_{N-1}} \int_{z_N} P(D_N = 1 | \mathbf{D}_{N-1}, z_N, x, y, H_1) \times \right. \\ \left. P(\mathbf{D}_{N-1}, z_N | x, y, H_1) dz_N \right) p(x, y) dy dx$$

At each sensor, we assume that the measurement is independent of the received incoming decisions so their joint probabilities can be expressed according to (2.29).

$$P(\mathbf{D}_{N-1}, z_N | H_0) = P(\mathbf{D}_{N-1} | H_0) p(z_N | H_0) \quad (2.29)$$

$$P(\mathbf{D}_{N-1}, z_N | x, y, H_1) = P(\mathbf{D}_{N-1} | x, y, H_1) p(z_N | x, y, H_1)$$

For simplicity, we only show the derivation of the probability of false alarm. Calculation of probability of detection is then quite straightforward except for an outer integration on the event location. Plugging (2.29) into (2.28) and using the fact that the sensor decision is independent of the underlying hypothesis we obtain (2.30)

$$P_{F,N} = \sum_{\text{all } \mathbf{D}_{N-1}} \left[ \int_{z_N} P(D_N = 1 | \mathbf{D}_{N-1}, z_N) p(z_N | H_0) dz_N \right] P(\mathbf{D}_{N-1} | H_0) \quad (2.30)$$

Given the event location  $(x, y)$  and independent and identically distributed noise at each sensor, the optimum decision rule at  $s_k$  is an LRT which uses the decisions of the previous sensors  $\mathbf{D}_{k-1}$  together with its own observation  $z_k$ . It is expressed as [3],

$$P(D_k = 1 | \mathbf{D}_{k-1}, z_k) = 0 \quad \text{if} \quad \frac{P(\mathbf{D}_{k-1} | H_1, x, y)}{P(\mathbf{D}_{k-1} | H_0)} \frac{p(z_k | H_1, x, y)}{p(z_k | H_0)} \leq t_k \quad (2.31)$$

$$P(D_k = 1 | \mathbf{D}_{k-1}, z_k) = 1 \quad \text{if} \quad \frac{P(\mathbf{D}_{k-1} | H_1, x, y)}{P(\mathbf{D}_{k-1} | H_0)} \frac{p(z_k | H_1, x, y)}{p(z_k | H_0)} > t_k$$

When the location of the event is a random variable, the sensor measurements from the event become correlated, the LRT shown in (2.31) is not necessarily optimal at the local sensors and the derivation of their optimal rules becomes a very hard problem. For this reason, motivated by the counting rule considered in [49] we use a heuristic decision statistic  $\delta_k$  at each sensor in the following form,

$$P(D_k = 1 | \mathbf{D}_{k-1}, z_N) = 0 \quad \text{if} \quad \delta_k = z_k + \sum_{p=1}^{k-1} D_p \leq t_k \quad (2.32)$$

$$P(D_k = 1 | \mathbf{D}_{k-1}, z_k) = 1 \quad \text{if} \quad \delta_k = z_k + \sum_{p=1}^{k-1} D_p > t_k$$

Basically, each sensor computes its decision statistic  $\delta_k$  by summing the number of 1s received from its predecessor sensors together with its own measurement. Then this decision statistic is compared with a certain threshold  $t_k$ . This heuristic rule works even when there is no prior information available in the network such as the location

of sensors or the location of the event.

The inner integration term in Eq.(2.30), can be written as shown in (2.33),

$$\int_{z_N} P(D_N = 1 | \mathbf{D}_{N-1}, z_N) p(z_N | H_0) dz_N = \quad (2.33)$$

$$\int_{t_N - (\sum_{k=1}^{N-1} D_k)}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z_N^2}{2}} dz_N = Q \left( t_N - \left( \sum_{k=1}^{N-1} D_k \right) \right)$$

Also, in Eq.(2.30) the probability mass function of the received decisions  $P(D_{N-1}, \dots, D_1 | H_0)$  needs to be iteratively conditioned on sensor decisions as shown in (2.34),

$$P(D_{N-1}, \dots, D_1 | H_0) = \int_{z_{N-1}} P(D_{N-1} | \mathbf{D}_{N-2}, z_{N-1}) p(z_{N-1} | H_0) dz_{N-1}$$

$$\dots \times \int_{z_1} P(D_1 | z_1) p(z_1 | H_0) dz_1 \quad (2.34)$$

In (2.34) depending on the local sensors decisions, each inner integral is replaced by appropriate  $Q(\cdot)$  or  $1 - Q(\cdot)$  function as defined in Eq.(2.33) based on the decision and the threshold of each local sensor.

Finally, the probability of error is found by averaging over all possible decisions in (2.35).

$$P_e = P(H_0) \left( \int_{z_N} P(D_N = 1 | \mathbf{D}_{N-1}, z_N) p(z_N | H_0) dz_N \quad (2.35)$$

$$\dots \int_{z_1} \sum_{D_1} P(D_1 | z_1) p(z_1 | H_0) dz_1 \right)$$

$$+ P(H_1) \left( 1 - \int_x \int_y \left[ \int_{z_N} P(D_N = 1 | \mathbf{D}_{N-1}, z_N, x, y) p(z_N | x, y, H_1) dz_N \right. \right.$$

$$\left. \left. \dots \int_{z_1} \sum_{D_1} P(D_1 | z_1, x, y) p(z_1 | x, y, H_1) dz_1 \right] p(x, y) dy dx \right)$$

## Energy Consumption

In serial decision fusion, a sensor's energy consumption depends not only on the distance between the source and the destination, but also the number of bits received and distance from its predecessors. Each sensor receives  $m - 1$  bits from its predecessors and transmits  $m$  bits to its next successor including its own decision. We define the distance between  $s_k$  and  $s_{k+1}$  as  $d_{k,k+1}$ , then  $E(i)$  the energy consumption of  $s_k$  is the sum of energy used for receiving  $m - 1$  bits from its predecessors  $E_{RX,k}$  and transmitting  $m$  bits to its successor over distance  $d_{k,k+1}$   $E_{TX,k}$  [64],

$$E_{RX,k}(m - 1) = E_{elec} \times (m - 1) \quad (2.36)$$

$$E_{TX,k}(m, d_{k,k+1}) = E_{elec} \times m + \epsilon_{amp} \times m \times d_{k,k+1}^2$$

$$E(i) = E_{RX,k} + E_{TX,k}$$

for  $i = \{2, 3, \dots, N - 1\}$  and  $E(1) = E_{TX,1}(D_1, d_{1,2})$  and  $E(N) = E_{RX}(\sum_{k=1}^{N-1} D_k)$  Joules.

Since the decision of the  $N^{th}$  sensor  $D_N$  is the final inference,  $D_N$  does not contribute to the energy consumption. Given the vector of past sensor decisions  $\mathbf{D}_{N-1}$ , energy consumption in the network is expressed as given in (2.37),

$$E_C(\mathbf{D}_{N-1}) = E_{TX}(D_1, d_{1,2}) + \quad (2.37)$$

$$\sum_{k=2}^{N-1} \left[ E_{TX,k} \left( \sum_{q=1}^k D_q, d_{q,q+1} \right) + E_{RX,k} \left( \sum_{q=1}^{k-1} D_q \right) \right] + E_{RX,N} \left( \sum_{k=1}^{N-1} D_k \right)$$

Finally, conditioning on all possible vector of decisions, the energy consumption  $E_T$  of the network is found according to (2.38),

$$E_T = \sum_{\text{all } \mathbf{D}_{N-1}} E_C(\mathbf{D}_{N-1}) P(\mathbf{D}_{N-1}) = \quad (2.38)$$

$$\sum_{\text{all } \mathbf{D}_{N-1}} E_C(\mathbf{D}_{N-1}) (P(\mathbf{D}_{N-1}|H_0)P(H_0) + P(\mathbf{D}_{N-1}|H_1)P(H_1))$$

where  $P(\mathbf{D}_{N-1}|H_0) = P(D_{N-1}, \dots, D_1|H_0)$  and  $P(\mathbf{D}_{N-1}|H_1) = P(D_{N-1}, \dots, D_1|H_1)$  are calculated as described in (2.34), and (2.35) respectively.

## 2.2 Multiobjective Optimization

In this section, we briefly summarize NBI and NSGA-II which are efficient methods to solve MOPs.

A well known technique for solving MOPs is to minimize a weighted sum of the objectives. Before describing the other MOP methods, in Figure 2.3, we minimize the weighted sum of the objectives  $P_e$  and  $E_T$  for parallel decision fusion with 5 sensors where the weights of each objective function are composed of evenly selected 10 points from the interval  $[0, 1]$ . In this problem, each solution to the MOP represents the set of local sensor thresholds  $\chi = [t_1, t_2, \dots, t_N]$ . As seen from the figure, minimizing the weighted sum of the objectives suffers from several drawbacks [25]. First of all, a uniform spread of weights rarely produces a uniform spread of points on the Pareto front. Some of the optimal design solutions are closely spaced which reduce the number of design alternatives. Secondly, if the Pareto optimal curve is not a convex function, the Pareto points on the concave parts of the actual Pareto optimal curve will be missed. Moreover, since it is up to the user to choose appropriate weights, decision on the preferences may not be clear to the user until the solution is generated. Similarly, when compared with other existing MOP algorithms such as Timmels Population Based Method (TPM) and Schaffers stochastic method (SSM) [30], the NBI method chosen in this study and explained below, is computationally efficient in locating Pareto optimal points. Therefore, we consider the application of NBI first to solve our MOP problem which is described briefly as follows:

### 2.2.1 Normal Boundary Intersection (NBI)

The NBI method [25] reduces the MOP to multiple number of single-objective constrained problems, called NBI subproblems. This method starts with finding the optimizers of each objective function separately. For the two-objective example illustrated in Figure 2.4, the shaded area represents the region of feasible design and the curve at



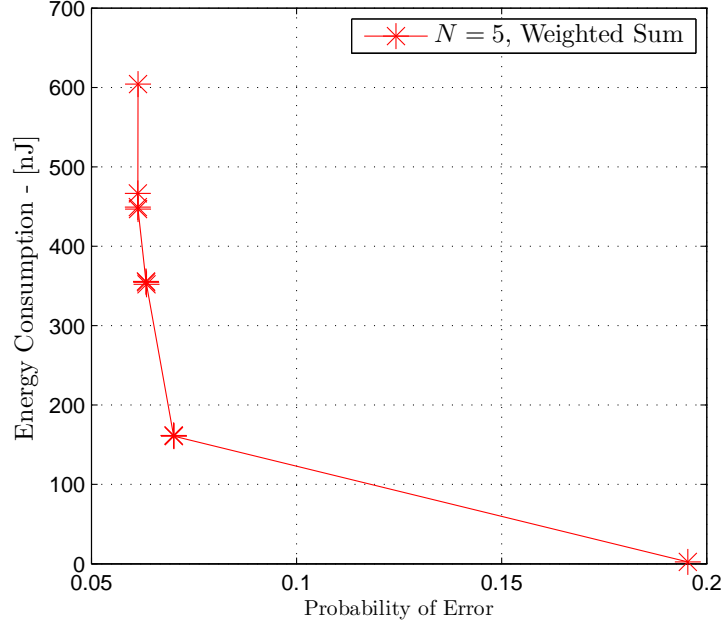


Figure 2.3: The Pareto optimal front found by minimizing the weighted sum of the objective functions,  $N = 5$

the lower boundary is the Pareto optimal front. The convex hull of individual minima (CHIM) is defined as the line segment AB. Any NBI problem is then specified by a reference point on the CHIM such as the point H. Let  $\chi_j^*$  be the minimizer of the  $j^{th}$  objective and  $F_j^* = F(\chi_j^*) = [f_1(\chi_j^*) \dots f_n(\chi_j^*)]^T$ , the payoff matrix  $\Phi$ , is an  $n \times n$  matrix whose  $j^{th}$  column is  $F_j^* - F^*$ .  $\beta$  represents one of the  $R_{NBI}$  evenly distributed points on the CHIM. As an example, if  $R_{NBI} = 11$ ,  $\beta$  is one of the following vectors.

$$\beta \in \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.9 \\ 0.1 \end{bmatrix}, \begin{bmatrix} 0.8 \\ 0.2 \end{bmatrix}, \dots, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$$

where for this example, we can denote  $\beta_1 \triangleq \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ ,  $\beta_2 \triangleq \begin{bmatrix} 0.9 \\ 0.1 \end{bmatrix}$  and so on.  $\Phi\beta$  then denotes the reference point H on the CHIM, and each NBI subproblem is defined as,

$$\max_{\chi, \tau} \quad \tau \tag{2.39}$$

$$s.t. \quad \Phi\boldsymbol{\beta} + \tau v = F(\chi)$$

$$h(\chi) = 0, g(\chi) \leq 0, a \leq \chi \leq b$$

The length of the line segment HP,  $\tau$ , represents the new variable introduced by the NBI subproblem. The new constraint given the NBI subproblem ensures that the point lies inside the feasible set  $C$ . The number of NBI subproblems,  $R_{NBI}$ , determines the resolution of the Pareto front. Clearly larger values for this parameter imply a better resolution of the Pareto front. If the Pareto set is disconnected, it is concluded that some of the subproblems have no solution [25]. Each NBI subproblem can be solved with any appropriate optimization method. Algorithm 1 summarizes the NBI method.

In Figure 2.5 and Figure 2.6, we show the contour plots of the global probability of error and global energy consumption for parallel and serial configurations. Since a closed form expression for either objective function (for the most general case of  $N$  sensors) is not available, the Hessian matrix composed of the  $2^{nd}$  derivatives of the objective functions with respect to sensor thresholds needed in a formal proof of unimodality can not be determined analytically. Hence any attempt to prove unimodality with the available information will only be approximate. For this reason, we choose to present numerical examples for the objective function's behavior rather than a formal proof. Simulation results show that even for a two sensor network the global probability of error is not a unimodal function of sensor thresholds. For such cases, gradient-based approaches may yield a local optimum instead of a global optimum. Hence the obtained results need to be compared with other global techniques such as an exhaustive search or genetic algorithm to ensure that the solutions really converge to the global optimum solution. For this purpose, next we describe an evolutionary algorithm for multiobjective optimization problems called NSGA-II.

**Algorithm 1 (The algorithm of NBI)**

$$\min_{\chi} f_1(\chi)$$

$$\min_{\chi} f_2(\chi)$$

*Generate CHIM*

$$r_{NBI} = 1$$

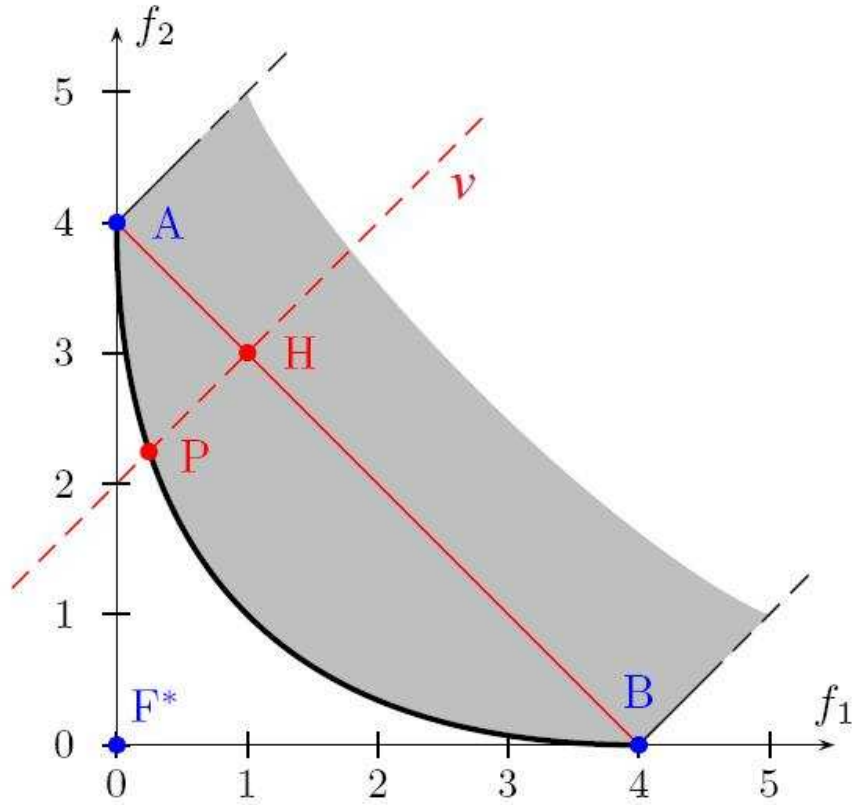


Figure 2.4: The point P is the solution of the single-objective constrained NBI subproblem outlined with the dashed line  $v$

**while**  $r_{NBI} \leq R_{NBI}$  **do**

*Solve the NBI subproblem*

$$\begin{aligned} \max_{\chi, \tau} \quad & \tau \\ \text{s.t.} \quad & \Phi \beta_{r_{NBI}} + \tau v = F(\chi) \\ & h(\chi) = 0, g(\chi) \leq 0, a \leq \chi \leq b \end{aligned}$$

$$r_{NBI} = r_{NBI} + 1$$

**end while**

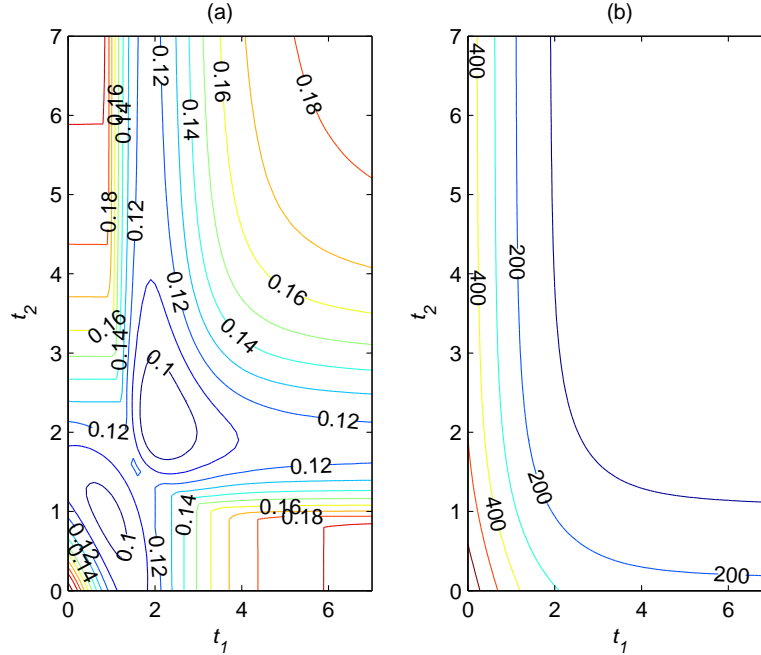


Figure 2.5: Contour lines of the objective functions with  $N = 2$  Sensors, parallel configuration (a) Probability of Error (b) Energy Consumption

### 2.2.2 Non-Dominating Sorting Genetic Algorithm - II (NSGA-II)

Non-Dominating sorting genetic algorithm-II (NSGA-II) [28] is a state of the art multiobjective evolutionary algorithm which simultaneously obtains  $M_{pop}$  Pareto optimal solutions in the  $n$  dimensional objective space. A solution in the population is represented as a sequence of decision variables, namely the sensor thresholds. Unlike NBI, using NSGA-II, the Pareto front (tradeoff curve) is found directly so there is no need to calculate the individual minimizers of each objective function separately. NSGA-II is an elitist algorithm where good solutions are preserved in the population.

NSGA-II is based on non-domination in each front. Each solution in the population is assigned a fitness and crowding distance value. The solutions with the same fitness are then re-sorted based on their crowding distance which is a closure measure of each solution to its neighbors. For each generation of the algorithm, the computational complexity  $O(n \times M^2)$  is governed by this nondominated sorting operation (See [28] for details). The mating population is subsequently generated by using binary tournament

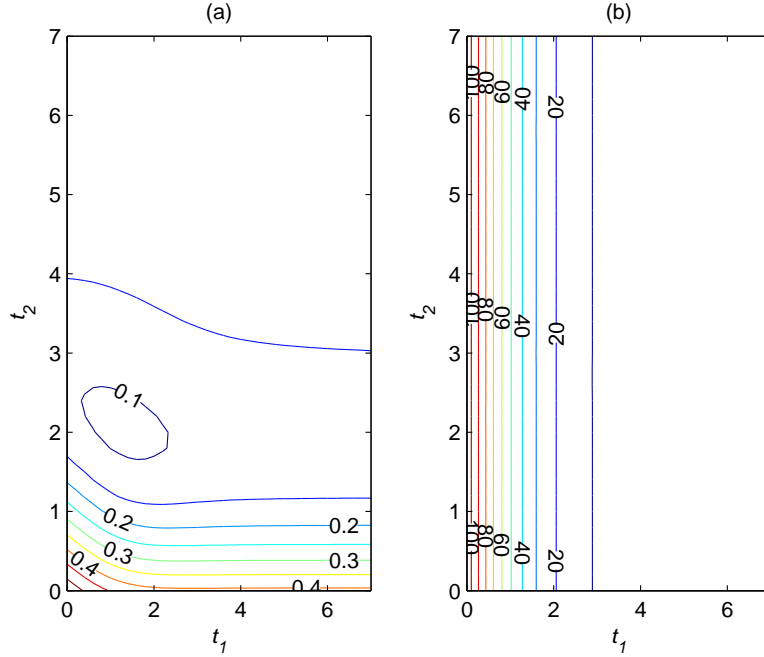


Figure 2.6: Contour lines of the objective functions with  $N = 2$  Sensors, serial configuration (a) Probability of Error (b) Energy Consumption

selection. If both of the solutions have the same fitness, the solution with larger crowding distance is selected. We use a real-parameter recombination operator, simulated binary crossover (SBX) which is used commonly in the evolutionary algorithm literature [28], [70]. The SBX has a parameter distribution index  $\eta_C$ , whose value determines the closeness of the offspring to their parents. Let  $p_1$  and  $p_2$  be two individual solutions obtained from binary tournament selection. In SBX, offspring solutions  $c_1$  and  $c_2$  are obtained from parent solutions  $p_1$  and  $p_2$  according to,

$$c_1 = \frac{1}{2}[(1 - \zeta)p_1 + (1 + \zeta)p_2] \quad (2.40)$$

$$c_2 = \frac{1}{2}[(1 + \zeta)p_1 + (1 - \zeta)p_2]$$

where  $\zeta \geq 0$  is a random number with probability density function [71],

$$p(\zeta) = \frac{1}{2}(\eta_C + 1)\zeta^{\eta_C} \quad 0 \leq \zeta \leq 1 \quad (2.41)$$

$$p(\zeta) = \frac{1}{2}(\eta_C + 1) \frac{1}{\zeta^{\eta_C+2}} \quad \zeta > 1$$

Along with the SBX, we use polynomial mutation that also makes use of a parameter distribution index,  $\eta_M$ . In polynomial mutation, the offspring solution  $c_l$  is obtained from the parent solution  $p_l$  according to,

$$c_l = p_l + (t_{max} - t_{min})\delta \quad (2.42)$$

where  $\delta$  is a small variation calculated from the density function [71],

$$p(\delta) = (2q)^{\frac{1}{\eta_M+1}} - 1 \quad q < 0.5 \quad (2.43)$$

$$p(\delta) = 1 - (2(1 - q))^{\frac{1}{\eta_M+1}} \quad q \geq 0.5$$

where  $q$  is a random number with uniform distribution between (0, 1).

The population is then updated by selecting the solutions starting from the first front. If the number of solutions in the last allowable front is larger than the available places in the population, the solutions with larger crowding distance is selected first. After several iterations, the entire population contains only of the solutions near or at the Pareto optimal front. Let  $G_T$  be the total number of NSGA-II iterations, then Algorithm 2 shows the high level description of NSGA-II.

### Algorithm 2 (The algorithm of NSGA-II)

*Randomly generate an initial population*

$g = 1$

**while**  $g \leq G_T$  **do**

- *Generate mating population*
- *Generate offsprings by simulated binary crossover or polynomial mutation*
- *Trim the new pool consisting of parents and offsprings to generate the population for the next iteration, with the primary criteria non-domination, secondary criteria crowding distance.*

$g = g + 1$

*end while*

### 2.2.3 Performance Metrics

For performance comparison between the solutions found with NBI and NSGA-II, we use three metrics: generational distance, domination metric and spacing metric [29], [72].

The generational distance (GD) [29],

$$GD(A, B) = \sqrt{\sum_{k=1}^M g_k^2} \quad (2.44)$$

measures the distance between the non-dominated solutions obtained by algorithms A and B.  $g_k$  is the Euclidean distance between the solution  $i \in A$  and the nearest solution in B.

Domination (Dom.) metric [29], is based on the number of solutions (obtained by one algorithm) dominated by each solution obtained by the other algorithm. The Dom metric is defined as:

$$Dom(A, B) = \frac{d(A, B)}{d(A, B) + d(B, A)} \quad (2.45)$$

where

$$d(X, Y) = \sum_x |\{y \in Y | x > y\}|$$

If each solution of algorithm A dominates every solution of algorithm B then  $Dom(A, B) = 1$  and  $Dom(B, A) = 0$  where  $Dom(B, A) = 1 - Dom(A, B)$ .

The Spacing metric [72], measures the uniformity of the solutions obtained in the Pareto-Optimal front. The  $S$  metric is defined as,

$$S(A) = \sqrt{\frac{1}{M-1} \sum_{k=1}^M (r_k - \bar{r})^2} \quad (2.46)$$

where  $M_{pop}$  is the number of nondominated solutions in the archive,  $r_k$  is the sum of the differences in objective function values between solution  $i$  and its two nearest

neighbors for each objective. The spacing metric approaches zero when the Pareto optimal solutions are near uniformly spaced.

## 2.3 Simulation Results

In this section, we first describe the simulation settings, then present the Pareto fronts obtained from NBI and NSGA-II algorithms and discuss the effects of non dominated solutions on WSN performance.

### 2.3.1 Simulation Settings

In our simulations, we use the WSN configuration described in Section 2.1.1. The solution of the MOP is illustrated with deterministic sensor placements where the sensors are equidistantly placed on the  $y = x$  line in the region of interest  $A \times B = 100m \times 100m$  as shown in Figure 2.1. As an example, boundary or pipeline surveillance requires placing the sensors on a straight line. The proposed MOP can be applied to any configurations as long as the sensor placements and the characteristics of the event of interest are known. The fusion center is located at the origin. According to our objective functions shown in Equations (2.20) and (2.26), adding an additional sensor doubles the number of possible vectors of received decisions. So the search space of both objectives increases exponentially with  $N$ , i.e. it is  $2^N$ . For this reason, we illustrate the proposed MOP with relatively few sensors. The *a priori* probabilities for  $H_0$  and  $H_1$  are selected as  $P(H_0) = 0.8$  and  $P(H_1) = 0.2$  respectively. The parameters of the event detection model are set as:  $P_0 = 5000$  and  $n = 2$ . The standard deviation of the measurement noise  $\sigma$  is set to 1. The minimum  $t_{min}$  and maximum  $t_{max}$  values for the thresholds are taken as 0 and 10 respectively. For NBI, individual minimizers of each objective function and each NBI subproblem are determined by using MATLAB ©'s `fmincon` routine. For the `fmincon` routine, all sensor thresholds are initialized at  $t_k^0 = 8$  where  $P_e^0 \approx 0.2$  and  $E_T^0 \approx 0$ , the algorithm termination tolerances of `fmincon` routine are all set to  $10^{-7}$ . The resolution of the Pareto-optimal front is selected as  $R_{NBI} = 11$ . For NSGA-II, we use a population of size  $M_{pop} = 100$ . Crossover and mutation probabilities are set at 0.9 and 0.1 respectively [73]. Parameter distribution indices of SBX and



Table 2.1: Generational Distance between NBI and NSGA-II, Spread Metric and Mean Execution Times (E.T.) for NSGA-II and NBI.

	GD (Mean)	GD (Std. Dev.)	S (Mean)	S (Std. Dev.)	Mean E.T. (seconds)
NSGA II: $G_T = 20$	603.9421	805.1765	19.2959	22.7605	0.2476e4
NSGA II: $G_T = 50$	253.0000	784.6029	4.5374	9.7953	0.6449e4
NSGA II: $G_T = 100$	5.0322	0.4002	1.6428	0.1827	1.2350e4
NSGA II: $G_T = 200$	5.1058	0.4269	1.7263	0.3049	2.3105e4
NSGA II: $G_T = 500$	5.0539	0.4498	1.9158	0.2094	5.5232e4
NBI: resolution 100			1.5594		1.7118e4

polynomial mutation are set to  $\eta_C = 20$  and  $\eta_M = 20$ , respectively. We observed that slight changes in these parameters do not change the results significantly. NSGA-II and NBI methods are implemented via available public codes in [73] and [74] respectively. All simulations are performed on a computer with a 3.2 GHz Pentium processor.

### 2.3.2 Performance Comparison of NBI and NSGA-II

For performance comparison between NBI and NSGA-II, we select both the Pareto front resolution of NBI and population size of NSGA-II as 100 with  $N = 4$  variables with parallel decision fusion. The generational distance (GD), domination metric (Dom) and spacing metric ( $S$ ) are averaged over 10 different NSGA-II trials. In Table 1, we vary the number of generations ( $G_T$ ) and measure the GD between 100 solutions of NSGA-II and NBI. Simulation results show that the average generational distance between NBI and NSGA-II is small after  $G_T = 100$  generations. Moreover, according to Table 1 after  $G_T = 100$  generations, the spacing metric of NSGA-II converges with small standard deviation. On the other hand, the solutions corresponding to NBI are more evenly spaced as compared to solutions of NSGA-II since NBI yields a smaller S-metric.

Table 2 reveals that the solutions obtained by NBI dominate the solutions of NSGA-II. Note that, the resolution of the Pareto optimal front is independent from the convergence of the NBI. So, 10 pareto-optimal solutions found with NBI clearly dominate every solution found with NSGA-II. So by solving a few subproblems, same Pareto optimal front can be achieved and evenly distributed trade off solutions can be obtained in a very short time as compared to NSGA-II. It should be pointed out that

Table 2.2: Domination Metric between NBI and NSGA-II.

Dom(A,B) A:NBI, B:NSGA-II	B: $G_T = 20$	B: $G_T = 50$	B: $G_T = 100$	B: $G_T = 200$	B: $G_T = 500$
A:res. 100,E.T.:1.7e4 s.	1	1	1	1	0.997
A:res. 20,E.T.:4.9e3 s.	1	1	1	1	1
A:res. 10,E.T.:2.3e3 s.	1	1	1	1	1

NBI is known to be better for two objective problems, but for problems with larger number of objectives, NSGA-II may be better [75].

### 2.3.3 Optimal Pareto Fronts

In this subsection, we present the Pareto optimal solutions for the parallel and serial decision fusion for the case of non-identical decision thresholds employed at each of the  $N$  sensors.

#### Pareto Optimal Fronts under Parallel Fusion

For parallel decision fusion and the WSN configuration as described in Section 2.1.2, Figure 2.7 shows the Pareto optimal fronts generated with NBI and NSGA-II where each solution is shown in terms of objective function pairs  $[P_e, E_T]$ . The number of decision variables is selected as  $N = 4, 5, 6$ . NSGA-II is executed with population size  $M_{pop} = 100$  and number of generations  $G_T = 100$  and NBI is executed with resolution 10. Simulation results show that NBI and NSGA-II yield Pareto optimal fronts that are fairly close to each other. Adding more sensors to the network decreases the error probability and NBI results in nearly equidistant points on the Pareto front. For  $N = 5$  sensors, if we only minimize  $P_e$ , the best achievable global probability of error is 0.061, which consumes  $497.5654nJ$ . By using the solution for the MOP, instead of selecting this solution, we may accept the neighboring solution on the Pareto optimal front with global error probability 0.0619, and global energy consumption 401.38nJ. Therefore, a 1.5% increase in the global probability of error, delivers 23% saving in global energy consumption. Similarly for the  $N = 6$  case, instead of operating on the minimum probability error solution  $[0.05, 536nJ]$ , selecting the solution  $[0.058, 246nJ]$  yields 53.9%

energy saving in exchange for a 15.11% increase in the global probability of error.

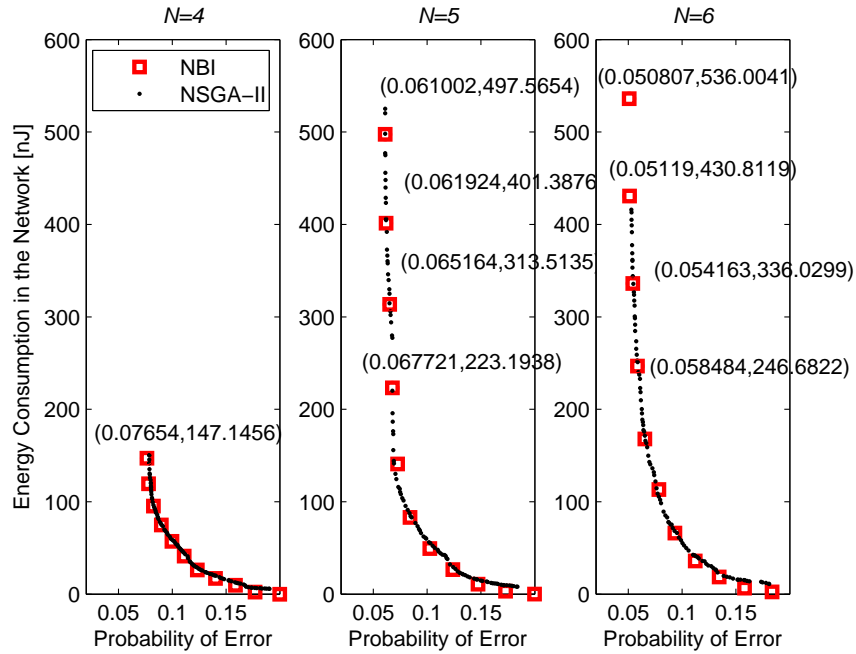


Figure 2.7: Pareto Optimal Solutions generated via NBI and NSGA-II methods for parallel fusion and non-identical decision thresholds at each sensor

### Pareto Optimal Fronts under Serial Fusion

Figure 2.8 shows the Pareto optimal solutions obtained with NBI and NSGA-II for the serial decision fusion case for  $N = 4, 5, 6$  sensors in the network according to the WSN configuration as shown in Figure 2.2. The heuristic decision rule of each sensor proposed in (2.30) yields the global probability of error that is slightly worse than the parallel configuration. As an example for  $N = 5$  and  $N = 6$  sensors the minimum achievable error probabilities for parallel decision fusion are 0.061 and 0.05 whereas serial case yields the minimum error probabilities 0.072 and 0.059 respectively. The global energy consumption of the serial configuration is determined by the distance between neighboring sensors and the number of received and transmitted bits of each sensor. For the  $N = 4$  case, the distance between two neighboring sensors is relatively large. The minimum error solution for the serial fusion consumes  $265nJ$  whereas the parallel configuration consumes  $147nJ$ . Although increasing the number of sensors

increases the number of bits for reception and transmission, the distance between sensors decreases significantly. Since energy consumption of the network is determined by the square of the inter sensor distance, increasing the number of sensors decreases the network's total energy consumption as compared to the parallel case. As an example, for  $N = 6$  sensors, under parallel network configuration, the minimum achievable global probability of error is about 0.050 with an energy consumption of 536nJ whereas under serial configuration the minimum achievable error probability is 0.059 with an energy consumption of 286nJ. In other words, deploying the network serially increases the probability of error by 18% but decreases the energy consumption of the network by 46%.

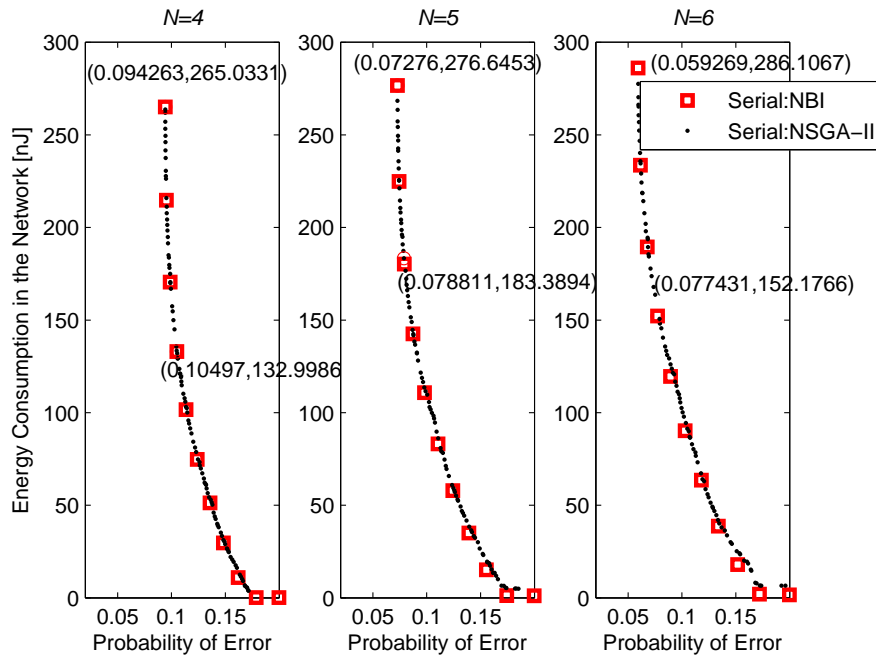


Figure 2.8: Pareto Optimal Solutions obtained by NBI and NSGA-II methods for serial fusion and non-identical decision thresholds at each sensor.

### 2.3.4 The Performance of WSN

In this subsection, we analyze the performance of WSNs based on the selected Pareto solution with  $N$  decision variables presented in the previous section. Under parallel decision fusion, we first determine the error probability of an individual sensor  $P_{ind}(t_k)$

as given in (2.18) as a function of its mean distance to the event location  $\bar{d}_k$ . We then calculate an individual sensor's energy consumption  $E_{ind}(t_k)$  as given in (2.22) as a function of its distance to the fusion center  $d_{f,k}$ . For serial decision fusion, we calculate the global probability of error and the global energy consumption of the network for a given number of sensor on the routing path.

### **WSN Performance under parallel decision fusion**

For the minimum global probability of error solutions, Figure 2.9 shows that the local sensor thresholds are assigned in such a way that the individual sensor error probability increases with the mean sensor distance to the event location. Due to this, a sensor transmits more frequently if it is close to the event and does not transmit that frequently if it is far. Then the error probability of a sensor far away from the mean event location is close to the prior probability  $P(H_1)$ . In terms of energy consumption, Figure 2.10 shows that energy consumption of a sensor increases with its distance to the fusion center. This is an expected result since the energy consumption of a sensor increases with the square of the distance to the fusion center. Figure 2.10 also shows that for the consecutive Pareto-optimal solutions with increased global probability of error and decreased global energy consumption, i.e  $N = 5 : [0.0619, 401\text{nJ}]$ ,  $N = 5 : [0.0651, 313\text{nJ}]$ ,  $N = 6 : [0.051, 430\text{nJ}]$ ,  $N = 6 : [0.054, 336\text{nJ}]$ , the energy consumption of the sensors that are far away from the fusion center decreases as their thresholds increase. Since these sensors are also relatively far from the expected event location, decrease in their transmission rate makes only a slight difference in the minimum achievable global probability of error. On the other hand, since delivering their decisions to the fusion center has much energy cost, decrease in their transmission rate provides significant savings in global energy consumption.

### **WSN Performance under serial decision fusion**

Figure 2.11 and Figure 2.12 show the global probability of error and the global energy consumption as a function of number of sensors (hops) on the routing path respectively. In these figures, sensor 1 ( $s_1$ ) in Figure 2.2 that is farthest away from the fusion center

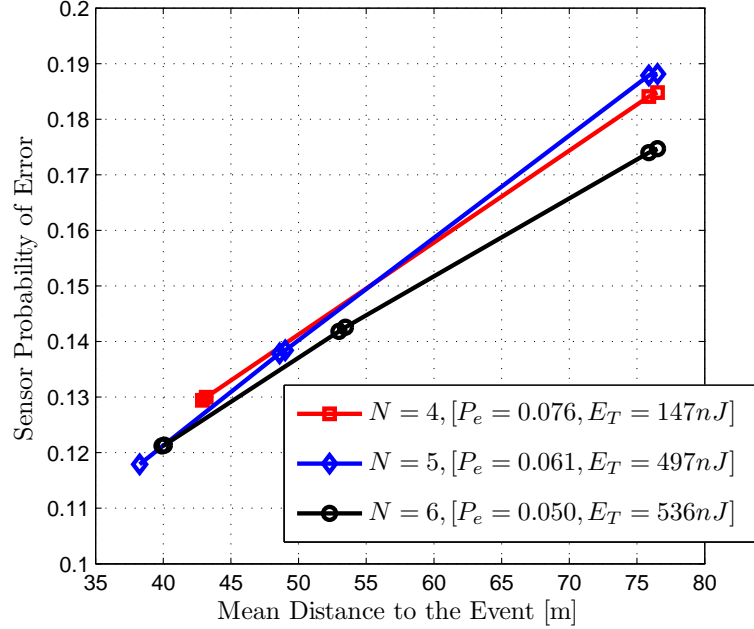


Figure 2.9: Parallel Decision Fusion - Local sensor error probability as a function of its mean distance to the event location

generates the first decision and transmits to  $s_2$ . At each sensor, we calculate the global probability of error and the global energy consumption. As a benchmark, we compare the performance of the proposed decision rule given in (2.32) with a simple rule where each sensor decision  $D_k$  is only the aggregation of the decision of the previous sensor and its measurement, that is  $D_k = z_k + D_{k-1}$ . In Figure 2.11 and Figure 2.12, since  $s_1$  has the farthest distance to the average event location, a higher threshold is assigned to this sensor and it is operating at a probability of error close to the prior probability  $P(H_1)$ . For the benchmark case, over consecutive sensors, the global probability of error decreases, since the sensors become much closer to the mean event location. On the other hand, the sensors near the fusion center are far away from the mean event location so the measurements of these sensors add uncertainty to the received decisions. That's why the global probability of error increases again when the distance to the fusion center is small (see the topmost curve in Figure 2.11). Our proposed rule considers not only the decision of the previous sensor but also the decisions of all predecessors. So increasing the number of sensors on the routing path increases the number of available

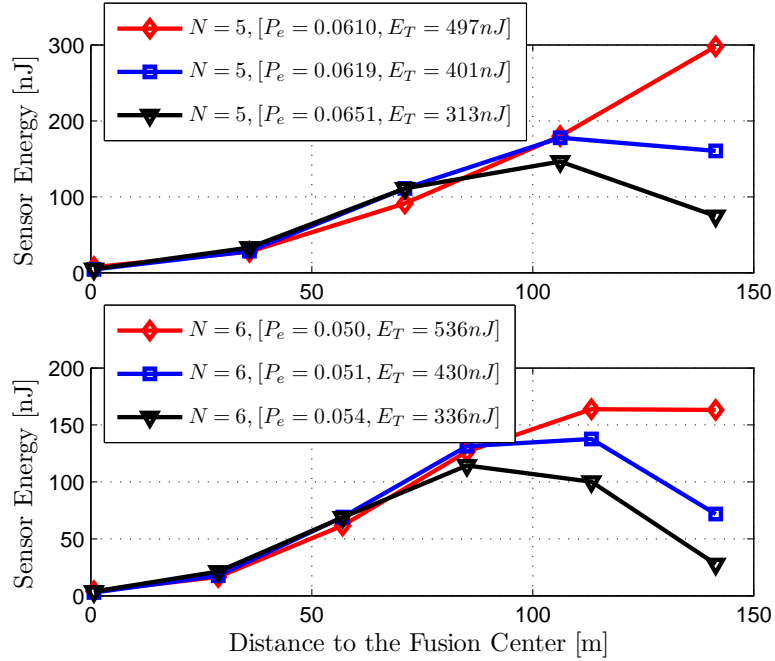


Figure 2.10: Parallel Decision Fusion - Local sensor energy consumption as a function of its distance to the fusion center

decisions about the event and the global probability of error decreases successively at each sensor. As shown in Figure 2.12, the global energy consumption of the network increases as a sensor close to the fusion center is included to the WSN. In the serial sensor topology, this is due to the fact that the cumulative increase in the transmitted and received bits at the sensors close to the fusion center.

### 2.3.5 Identical Decision Thresholds

Since the search space of both objectives increases exponentially with  $N$ , adding an additional sensor roughly doubles the computation time. In order to simplify the problem, we may constrain the decision rules to be identical at all the sensors. For parallel decision fusion, Figure 2.13 shows the optimal Pareto fronts for the case of identical and nonidentical decision thresholds. For  $N = 4, 5, 6$  sensors assuming an identical decision threshold for all the sensors, yields the objective function pair values with minimum global probability of error given by  $[0.080, 161nJ]$ ,  $[0.061, 563nJ]$ ,

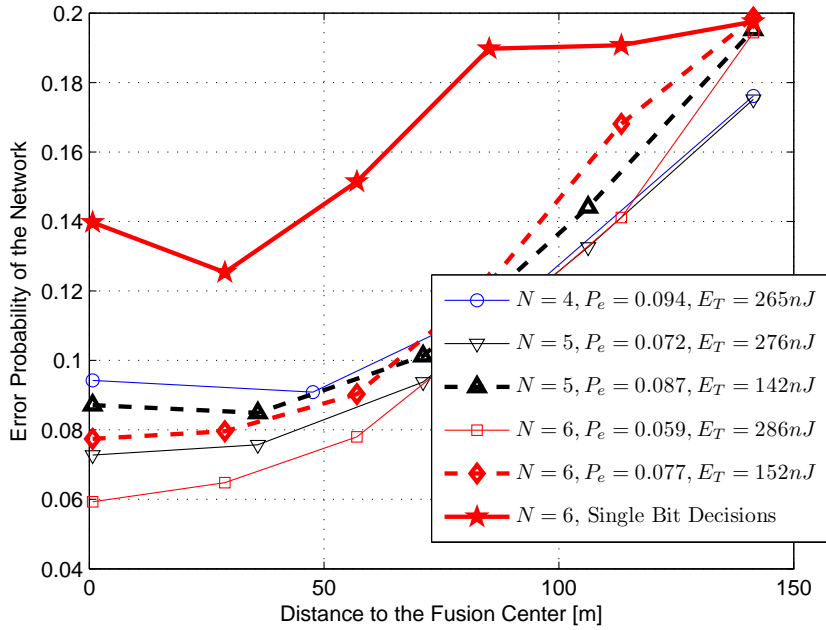


Figure 2.11: Serial Decision Fusion - Global error probability as a function of the hop count on the routing path

[0.051, 555nJ] whereas the non-identical threshold selection gives the objective function pair values [0.076, 147nJ], [0.061, 497nJ], [0.0508, 536nJ] respectively. Simulation results show that as the number of sensors in the network increases an identical decision threshold for all the sensors achieves nearly the same error probability as compared to non-identical threshold selection. In [36], it is shown that as the number of sensors grows to infinity, the probability of error goes to zero, for any reasonable set of decision thresholds. Therefore, fine adjustment of decision thresholds at each sensor becomes unnecessary if the number of sensors is large. So especially for a large number of sensors an identical decision threshold at all the sensors simplifies the problem and yet gives near optimal results. Note that the energy consumption of the network is slightly higher for the identical threshold case. This is due to the fact that under non-identical threshold assignment, the sensors that are far from the event are assigned a higher threshold which reduces the transmission rate of these sensors. On the other hand, for identical threshold selection these sensors are assigned a lower threshold value and transmit more frequently. Also, simulation results show that for identical threshold se-



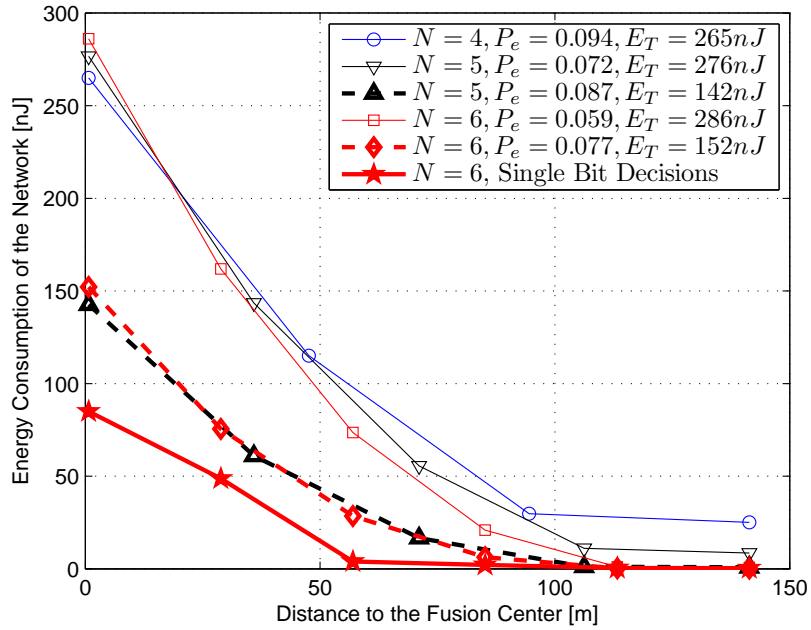


Figure 2.12: Serial Decision Fusion - Global energy consumption as a function of the hop count on the routing path

lection, the Pareto-optimal curve between  $E_T$  and  $P_e$  is not convex which implies that some of the candidate solutions on the Pareto front are still dominated. As an example, for  $N = 5$  case, the two solutions  $[0.077246, 293nJ]$  and  $[0.072728, 162nJ]$  are on the Pareto front but  $[0.077246, 293nJ]$  is already dominated by  $[0.072728, 162nJ]$ . On the other hand, non-identical sensor threshold selection yields a convex Pareto front where all the solutions are non dominated providing more alternatives to the designer. For serial decision fusion, Figure 2.14 shows the Pareto optimal solutions for the case of identical and nonidentical decision thresholds. Simulation results show that the best achievable error probability with identical threshold selection is slightly worse than the non-identical threshold selection. Similar to the parallel decision fusion system, the identical threshold scheme has a non convex trade-off curve and for a given global probability of error, energy consumption of identical threshold selection is higher than the non-identical threshold selection.

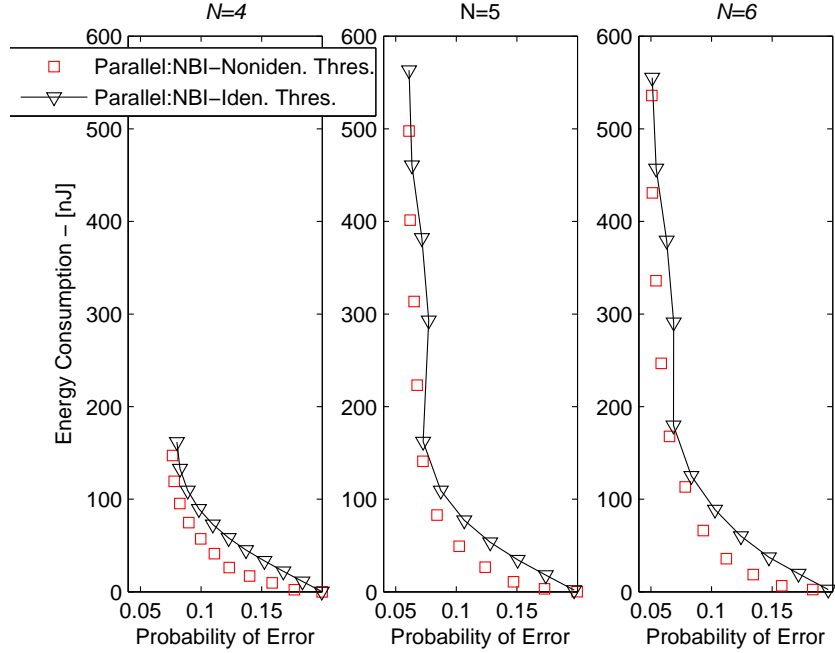


Figure 2.13: Parallel Decision Fusion, Pareto Optimal Solutions for identical and non-identical sensor thresholds

## 2.4 Discussion

In this chapter, we have studied the binary distributed detection problem. The event signal is represented by an isotropic emission model and the location of the event is statistically known where the decision thresholds that minimize the probability of error can not be determined using the existing methods such as PBPO. We formulated and solved a multi-objective optimization problem with two conflicting objectives: global probability of error and global energy consumption of the network where each solution of this problem corresponds to placing a different emphasis on the two objectives. The proposed MOP is solved by two different methods. NBI and NSGA-II yield Pareto optimal fronts that are very close to each other. Simulation results show that for our problem NBI provides better and more uniformly distributed solutions in a shorter time as compared to NSGA-II.

Under parallel decision fusion, the consecutive Pareto-optimal solutions decrease the global energy consumption significantly by allowing a slight increase in the minimum

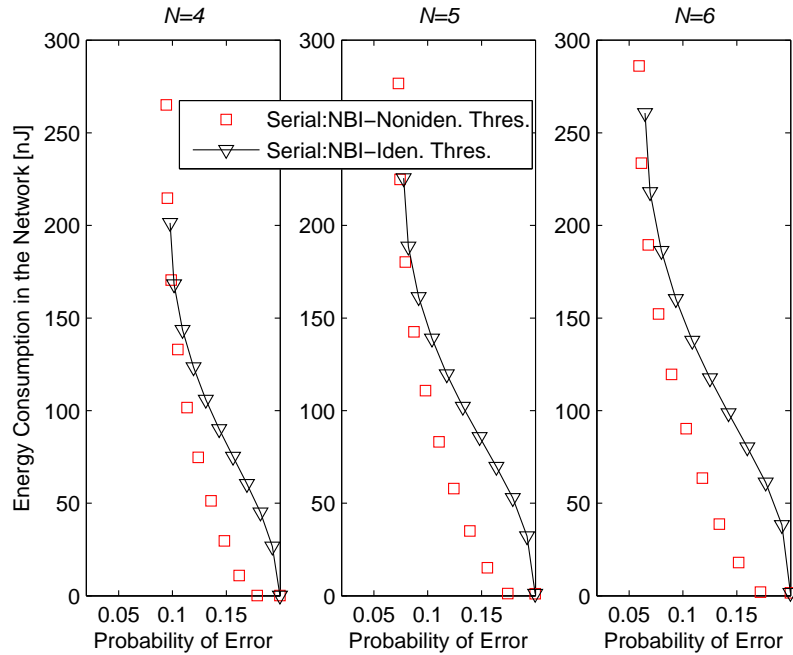


Figure 2.14: Serial Decision Fusion, Pareto Optimal Solutions for identical and non-identical sensor thresholds

achievable probability of error. Under serial decision fusion, increasing the number of sensors on the routing path, increases the amount of information about the event and the probability of error at each sensor decreases successively. We have also shown that an identical decision threshold for all the sensors achieves nearly the same error probability as compared to nonidentical threshold selections at each sensor as the number of sensors in the network increases. Therefore, especially for large number of sensors, an identical decision rule for all sensors can be employed to achieve nearly the best probability of error performance.

## Chapter 3

# Energy Aware Iterative Source Localization

In the previous chapter, we have studied the event detection problem where the source location is assumed to be random and uniformly distributed in a ROI. Accurate source localization is another important task to be performed by a WSN, the result of which has crucial role in accurate target tracking or higher level motion analysis. In this chapter, we study the static source localization problem where the aim is to estimate the coordinates of an energy emitting source (e.g. acoustic source).

In a region of interest (ROI), an accurate estimate of the source location can be obtained by using the energy readings of the sensors [48], [44]. In [48] and [44], maximum likelihood (ML) based approaches have been proposed by using analog and multi-bit ( $M$ -bit) sensor measurements respectively at the fusion center. In this work, we assume that each sensor measurement is quantized into  $M$ -bits and delivered to the fusion center over an error-free channel. Simultaneous transmission of all sensors'  $M$  bit data to the fusion center introduces some challenges. First of all, the sensors that are far from the source location are not likely to carry much useful information but they still consume energy to transmit information. Secondly, each sensor requires an independent channel for simultaneous data transmission to the fusion center. This assumption imposes a limitation on the number of sensors that the system can support in practice. Therefore, rather than transmitting multi-bit data from all the sensors, we first employ measurements from a relatively few anchor sensors to detect the source and obtain a coarse location estimate. In the literature, anchor sensors are utilized to find

the sensor node locations [76], [77]. In this work, we assume that sensor placements are known *a-priori* at the fusion center and try to estimate the source location. Our iterative algorithm starts when the anchor sensors send their multi-bit data to the fusion center. The non-anchor sensors do not transmit their measurements in the initial phase. A few non-anchor sensors are activated at each step of an iterative procedure. Now the problem is to select the set of non-anchor sensors at each iterative step which improve the accuracy of the source location estimate the most. These activated sensors send their multi-bit measurement data to the fusion center to refine the location estimate. Distributed compression of measurement data prior to transmission is also employed at the non-anchor sensors to further reduce the energy consumption. Thus, we achieve significant energy savings in source localization at the cost of tolerating some delay.

The sensor selection problem in sensor networks has been widely studied in the literature. For sensor management, information based measures have recently been proposed as objective functions to choose the sensing action that maximizes the expected gain in information [78], [79], [14,15], [16,17], [80,81]. In [78], a sensor selection approach has been proposed which chooses the sensors having maximum mutual information with source location based on the analog sensor measurements. In [79], authors focus on using the expected change in Shannon entropy when tracking a single target. In [14,15], authors have compared several sensor selection approaches involving entropy and relative entropy. Kreucher et al [16,17] have proposed sensor management schemes that maximize the Rényi divergence between the current target state probability density and the density after a new measurement arrives. In [80,81], sensors are selected to maximize the mutual information between the sensor measurements and the target state.

The PCRLB is a very important tool because it provides a theoretical performance limit for a Bayesian estimator. In [82], Tichavsky *et. al.* derived an elegant recursive approach to calculate the sequential PCRLB for a general multi-dimensional discrete-time nonlinear filtering problem. In [83], based on the PCRLB, a sensor deployment approach is developed to achieve better tracking accuracy while at the same time it uses the limited sensor resources more efficiently. Such approaches are extended in [18]

to incorporate sensor deployment and motion uncertainties. For single target tracking, a subset of sensors are selected in a bearing-only sensor network to minimize the PCRLB on the estimation error, where the selected sensors transmit analog data [84] or quantized data [85] to the fusion center. Further, the PCRLB based criterion has been employed to manage sensor arrays for multi-target tracking problems [19,86]. Another related work is reported in [87], where a PCRLB based adaptive radar waveform design method for target tracking has been presented. In this chapter we show that, one problem with the approaches based on information theoretic approaches is that the complexity to compute mutual information is large, especially when the number of sensors to be selected,  $A$ , is large. If the sensors provide quantized data, we show in this chapter that the computational complexity of the mutual information is exponential in  $A$ , whereas the complexity of the PCRLB is linear in  $A$ . This fact makes the sensor management based on information theoretic measures impractical when  $A$  is large.

First, we extend the mutual information based sensor selection scheme presented in [81] for quantized sensor measurements. Then, we define another metric for sensor selection based on the PCRLB. Note that in [88] the recursive approach presented in [82] is utilized to calculate the PCRLB. After that, we re-formulate the PCRLB for static source location estimation. We approximate the posterior pdf of the source location using an importance sampling based Monte-Carlo method [24]. Using this posterior pdf approximated by Monte-Carlo methods, a number of non-anchor sensors are activated in an iterative manner which minimize the PCRLB. Simulation results show that, within a few iterations, the mean squared error approaches the PCRLB of a Bayesian estimate based on all the sensor data. Since the fusion center is not likely to request multi-bit data of the non-informative sensors; which are typically far away from the source location, the proposed iterative algorithm is expected to provide large energy savings.

When sensors are densely deployed in a region of interest (ROI), the sensor measurements are likely to be spatially correlated and this correlation can be utilized to compress the quantized measurements of each sensor prior to transmission to further reduce energy consumption [89], [90]. Given the multi-bit data received during previous

iterations and the posterior pdf of the source location, the fusion center calculates the conditional entropy of the sensors to be activated during an iteration and it requests a compressed version of sensor's multi-bit data. Simulation results show that for the first few iterations, the uncertainty about the source location is very high which implies a high conditional entropy for the sensor to be activated. In such circumstances, data compression does not have much effect and each sensor measurement is sent to the fusion center using almost  $M$ -bits. Including new data at each iteration reduces the uncertainty about the source location and the conditional entropy of each activated sensor gets smaller at each iteration. After the most informative sensors about the source location have been selected, the conditional entropy for each activated sensor becomes very small and only a small number of bits are requested by the fusion center. Hence, data compression yields significant energy savings.

The rest of the chapter is organized as follows. In Section 3.1, we introduce the system model and review the ML location estimation method as presented in [44]. In Section 3.2, we present the iterative source location estimation algorithm. We first give a brief overview of the algorithm. We then explain the estimation approach for source location and approximation of the posterior pdf of source location using Monte Carlo methods. Using this posterior pdf later in this chapter we describe the sensor selection methodology. We extend the mutual information based sensor selection method for quantized sensor data and also present the PCRLB based sensor selection method. In Section 3.3, we discuss data compression using the distributed source coding approach. In Section 3.4, we compare the two sensor selection schemes in terms of computation time and give numerical examples to show their estimation performance. In Section 3.4, we also study the trade-off between estimation performance and communication cost. We first define a stopping criterion to terminate the algorithm and show that as  $A$ , the number of non-anchor sensors to be activated at each iteration, increases, the algorithm terminates much faster, at the cost of increased total number of bits transmitted to the fusion center. Finally, Section 3.5 is devoted to discussion of the results.

### 3.1 System Model

We consider a WSN consisting of  $N$  sensors  $\{s_k, k = 1, 2, \dots, N\}$ . We assume that a signal (e.g., an acoustic signal) is radiated from a location  $(x, y)$  that follows an isotropic power attenuation model. In this chapter, we assume that the source is based on flat ground and all the sensors and source have the same height so that a 2-D model is sufficient to formulate the problem. As an example, an acoustic event on the ground can be analyzed using a 2-D scenario as shown in Fig. 3.1. In this chapter, we assume that  $N$  sensors are deployed in a grid layout and the WSN uses a parallel architecture where the quantized measurements of each sensor are directly delivered to the fusion center. The assumption of grid layout is not necessary. Source localization based on sensor readings can be performed for an arbitrary network layout if sensor placements are known in advance. The location of each sensor ( $s_k$ ) is represented by  $(x_k, y_k)$ . Then, the distance between  $s_k$  and the source location  $(x, y)$  is  $d_k = \sqrt{(x - x_k)^2 + (y - y_k)^2}$ . The received source energy  $a_k^2$  at  $s_k$  is expressed as [44],

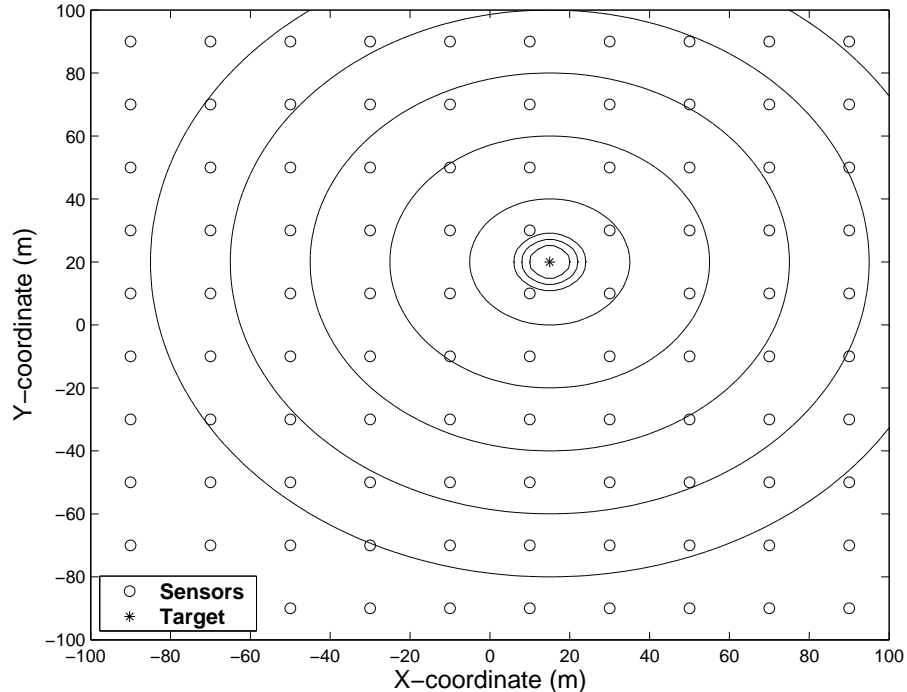


Figure 3.1: The signal intensity contours of a source located in a sensor field.



$$a_k^2 = P_0 \left( \frac{d_0}{d_k} \right)^n \quad (3.1)$$

where  $P_0$  is the signal power measured at a reference distance  $d_0$  (In this chapter, we set  $d_0 = 1m.$ ),  $a_k$  is the received signal amplitude at sensor  $s_k$  and  $n$  is the signal decay exponent. At each sensor, the received signal amplitude  $a_k$  is corrupted by an additive Gaussian noise:

$$z_k = a_k + n_k \quad (3.2)$$

where  $z_k$  is the noisy signal measurement at sensor  $s_k$ . Here, we assume that the noise  $n_k$  is independent and identically distributed across sensors with Gaussian distribution  $\mathcal{N}(0, \sigma)$  with  $\sigma = 1$ .

We assume the same set of quantization thresholds at all the sensors  $\boldsymbol{\eta} = \boldsymbol{\eta}_1 = \boldsymbol{\eta}_2 = \dots = \boldsymbol{\eta}_N$  and  $\boldsymbol{\eta} = [\eta_0, \eta_1, \dots, \eta_L]^T$ . Then  $D_k$  is obtained from  $z_k$  according to (5.4). The optimal quantization rules for  $M$ -bit sensor data are given in [44]. Such rules mostly affect the performance when the number of decision intervals  $L = 2^M$  is small (e.g. for the cases when  $M = 1$  or  $M = 2$ ). Since we are interested in a larger number of quantization levels such as  $M \geq 3$ , the optimal design of decision thresholds becomes less crucial. Therefore, we assume that each sensor in the field uses the same decision thresholds and employs a simple quantization rule that is,  $L-1$  points which evenly partition the interval  $[0, \sqrt{P_0}]$ , are selected as the thresholds. The sensor measurements less than 0 and more than  $\sqrt{P_0}$  are mapped to 0 and  $L-1$  respectively.

In this chapter, we assume that the source location  $\boldsymbol{\theta} = [x, y]^T$  follows a prior pdf  $p_0(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$  where  $\boldsymbol{\mu}_0$  is the center of the ROI and  $\boldsymbol{\Sigma}_0 = \begin{bmatrix} \sigma_{x,0}^2 & 0 \\ 0 & \sigma_{y,0}^2 \end{bmatrix}$  is the covariance matrix which is very coarse so that its 99% confidence region covers the whole ROI. Note that our proposed approach does not require the prior pdf to be Gaussian and will work with other prior pdf's also.

Under the Gaussian noise assumption, the probability that  $D_k$  takes a specific value  $l$  is,

$$p(D_k = l | \boldsymbol{\theta}) = Q \left( \frac{\eta_l - a_k}{\sigma} \right) - Q \left( \frac{\eta_{(l+1)} - a_k}{\sigma} \right) \quad (3.3)$$

where  $Q(\cdot)$  is the complementary distribution function of the standard Gaussian distribution, and

$$Q(t) \triangleq \int_t^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \quad (3.4)$$

### 3.1.1 Source detection using multi-bit sensor data

Recall from Section 2.1.2 that  $D_0$  be the global decision at the fusion center about the presence or absence of an event, and  $P(0)$  and  $P(1)$  be the *a priori* probabilities of  $H_0$  and  $H_1$  respectively. Instead of using all  $N$  sensors in the network,  $K$  anchor sensors are utilized to detect the event first. The probability of error is given by [3],

$$P_e = P(0)P_F + P(1)(1 - P_D) \quad (3.5)$$

where  $P_F = P(D_0 = 1|H_0)$  denotes the probability of false alarm, and  $P_D = P(D_0 = 1|H_1)$  denotes the probability of detection. Given the vector of quantized sensor data of  $K$  anchor sensors  $\mathbf{D}_K = [D_1, D_2, \dots, D_K]$  and  $D_k \in \{0, 1, \dots, L-1\}$ , the probability of error is expressed as

$$P_e = P(0)P(D_0 = 1|H_0) + P(1)(1 - P(D_0 = 1|H_1)) \quad (3.6)$$

which can be written as,

$$P_e = P(1) + P(D_0 = 1|\mathbf{D}_K)[P(0)P(\mathbf{D}_K|H_0) - P(1)P(\mathbf{D}_K|H_1)]$$

$P_e$  is minimized if,

$$P(D_0 = 1|\mathbf{D}_K) = 0 \text{ when } [P(0)P(\mathbf{D}_K|H_0) - P(1)P(\mathbf{D}_K|H_1)] > 0 \quad (3.7)$$

$$P(D_0 = 1|\mathbf{D}_K) = 1 \text{ when } [P(0)P(\mathbf{D}_K|H_0) - P(1)P(\mathbf{D}_K|H_1)] < 0$$

The above property leads to the following likelihood ratio test (LRT) at the fusion center [3],

$$\frac{P(\mathbf{D}_K|H_1)}{P(\mathbf{D}_K|H_0)} \underset{D_0=0}{\underset{D_0=1}{\geq}} \frac{P(0)}{P(1)} \quad (3.8)$$

Under  $H_1$ , the source location follows a certain prior probability density function ( $p_0(\boldsymbol{\theta})$ ), then (3.8) is computed according to,

$$\frac{1}{P(\mathbf{D}_K|H_0)} \int_{\boldsymbol{\theta}} P(\mathbf{D}_K|\boldsymbol{\theta}, H_1) p_0(\boldsymbol{\theta}) d\boldsymbol{\theta} \underset{D_0=0}{\underset{D_0=1}{\geq}} \frac{P(0)}{P(1)} \quad (3.9)$$

### 3.1.2 Source location estimation using multi-bit sensor data

After the fusion center determines the presence of the source, then the fusion center estimates the location of the source. Let  $\mathbf{D} = [D_1, D_2, \dots, D_N]^T$  represent the collected data from all  $N$  sensors. Given the source location  $\boldsymbol{\theta}$ , the quantized sensor measurements are conditionally independent. Therefore, the likelihood function at the fusion center has the form [44],

$$p(\mathbf{D}|\boldsymbol{\theta}) = \prod_{k=1}^N \prod_{l=0}^{L-1} p(D_k = l|\boldsymbol{\theta})^{\delta(D_k-l)} \quad (3.10)$$

where  $\delta(l)$  is the Kronecker delta function and is defined as,

$$\delta(t) = \begin{cases} 1 & t = 0 \\ 0 & t \neq 0 \end{cases} \quad (3.11)$$

The log-likelihood function of the source location then has the form,

$$\ln p(\mathbf{D}|\boldsymbol{\theta}) = \sum_{k=1}^N \sum_{l=0}^{L-1} \delta(D_k - l) \ln[p(D_k = l|\boldsymbol{\theta})] \quad (3.12)$$

The maximum likelihood estimate (MLE) of the source location  $\hat{\boldsymbol{\theta}}^{MLE}$  is the solution

of the following optimization problem:

$$\hat{\boldsymbol{\theta}}^{MLE} = \arg \max_{\boldsymbol{\theta}} \ln p(\mathbf{D}|\boldsymbol{\theta}) \quad (3.13)$$

Assuming the existence of an unbiased estimator  $\hat{\boldsymbol{\theta}}(\mathbf{D})$ , the CRLB is given by [13],

$$E\{[\hat{\boldsymbol{\theta}}(\mathbf{D}) - \boldsymbol{\theta}][\hat{\boldsymbol{\theta}}(\mathbf{D}) - \boldsymbol{\theta}]^T\} \geq \mathbf{J}^{-1}(\boldsymbol{\theta}) \quad (3.14)$$

in which  $\mathbf{J}(\boldsymbol{\theta})$  is the  $2 \times 2$  Fisher information matrix (FIM) and the source location estimate  $\hat{\boldsymbol{\theta}}(\mathbf{D})$  is a function of  $\mathbf{D}$ . Given an unknown constant source location  $\boldsymbol{\theta}$ ,  $\mathbf{J}(\boldsymbol{\theta})$  can be found as follows:

$$\mathbf{J}(\boldsymbol{\theta}) = n^2 \sum_{k=1}^N \kappa_k a_k^2 d_k^{-4} \begin{bmatrix} (x_k - x)^2 & (x_k - x)(y_k - y) \\ (x_k - x)(y_k - y) & (y_k - y)^2 \end{bmatrix} \quad (3.15)$$

where

$$\kappa_k = \frac{1}{8\pi\sigma^2} \sum_{l=0}^{L-1} \frac{\gamma_{k,l}}{p(D_k = l|\boldsymbol{\theta})}$$

and

$$\gamma_{k,l} = \left[ e^{-\frac{(\eta_l - a_k)^2}{2\sigma^2}} - e^{-\frac{(\eta_{l+1} - a_k)^2}{2\sigma^2}} \right]^2$$

The derivation of the FIM provided in (3.15) can be found in [44].

Since  $\boldsymbol{\theta}$  is treated as a random parameter with a certain prior pdf, we consider PCRLB as the estimation benchmark for the mean squared error (MSE) which is defined later in the chapter.

### 3.2 Iterative Source Location Estimation Method

Fig. 3.2 depicts an example WSN where each black point represents a sensor and the proposed iterative source localization algorithm is illustrated in Fig. 3.3. At step 1, the algorithm starts with the collection of  $M$ -bit quantized data from each of the  $K$  anchor sensors (represented with blue squares in Fig. 3.2). For notational simplicity, let  $\mathbf{W}_i = [D_1, D_2, \dots, D_{K+iA}]^T$  denote the collected sensor data until and including the

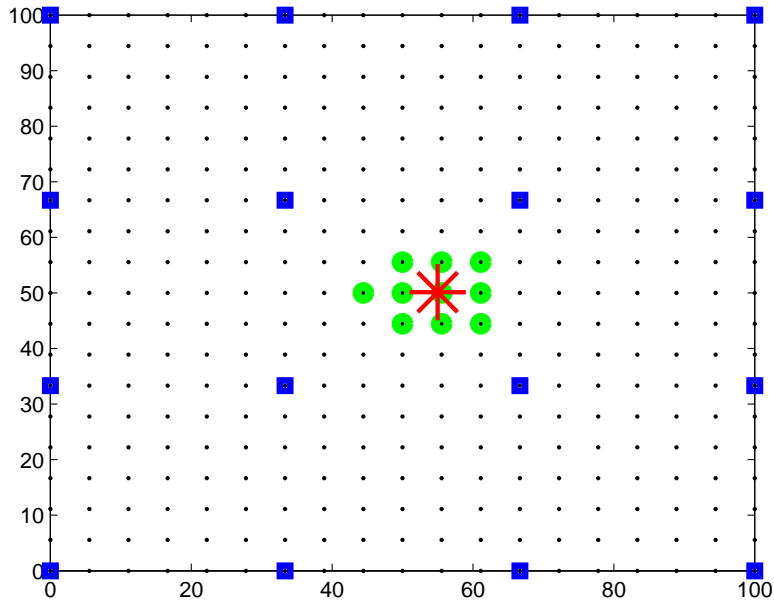


Figure 3.2: Wireless Sensor Network Model. Black Points: Sensor Locations; Blue Squares: Anchor Sensors used for initial iteration; Green Circles: Activated Sensors after 10 iterations for the example considered in Section V; Red Star: Source.  $A = 1$  sensor is activated / iteration

$i^{th}$  iteration where  $i \in \{0, 1, \dots\}$  and  $A$  is the number of non-anchor sensors activated at each iteration (activated sensors are represented with green circles in Fig. 3.2 for the example considered in Section V). Note that, at iteration 0, only the anchor sensor data are received at the fusion center. Let  $p(\boldsymbol{\theta}|\mathbf{W}_i)$  denote the posterior pdf of the source location based on the currently available sensor data  $\mathbf{W}_i$  at the  $i^{th}$  iteration. At step 2 of the algorithm shown in Fig. 3.3, the fusion center finds the source location estimate  $\hat{\boldsymbol{\theta}}$  using the posterior pdf  $p(\boldsymbol{\theta}|\mathbf{W}_i)$ . The algorithm starts the next iteration ( $i = i + 1$ ) at step 3 of the algorithm. Note that the posterior pdf of the source location is based on the previously received data until the end of iteration  $i - 1$  and  $p(\boldsymbol{\theta}|\mathbf{W}_{i-1})$  serves as the prior pdf of source location for the  $i^{th}$  iteration, which is denoted as  $p_i(\boldsymbol{\theta})$ ,

$$p_i(\boldsymbol{\theta}) = p(\boldsymbol{\theta}|\mathbf{W}_{i-1}) \quad (3.16)$$

At step 4 of the algorithm, the fusion center activates  $A$  non-anchor sensors. In

this work, we present two sensor selection strategies. The first one selects the sensors that maximize the mutual information (MI) between the source location and sensors to be selected. The second one chooses the sensors that minimize the PCRLB. These two approaches will be compared in terms of computation complexity and mean squared error performance later in the chapter. Finally, at step 5, using the already available sensor data as side information, the  $M$ -bit data of each activated sensor is locally compressed using standard distributed source coding techniques. We will show later through simulations that as the amount of information about the source location increases and the most informative (based on either the MI criterion or PCRLB criterion) sensors about the source are selected, the estimation error on the source location decreases quickly.

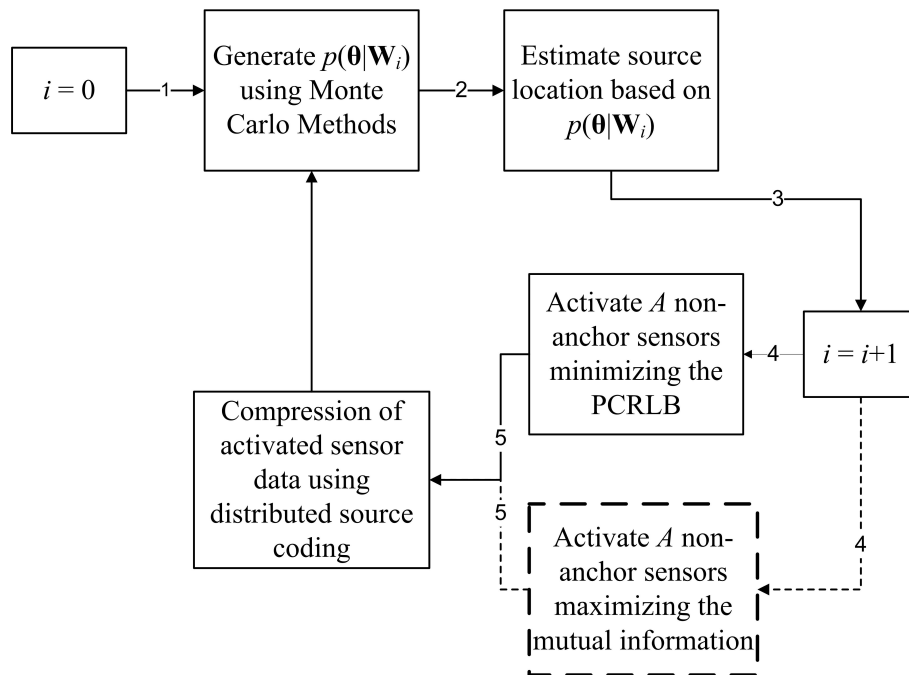


Figure 3.3: The flow chart of the algorithm. The dashed blocks represent the state-of-the-art Mutual Information based sensor selection method. The entire set of solid blocks represent the PCRLB based algorithm.

### 3.2.1 Source Location Estimation Based on Monte Carlo Methods

In signal processing, Monte Carlo methods are used to obtain simulation based estimates [24], [91], [88]. Monte Carlo-based methods are currently used to compute integrals, evaluate marginal distributions and optimization. Basically, Monte Carlo methods can be classified into three categories (see [24] and references therein). The acceptance-rejection methods are used to sample from the probability distribution of interest. Importance sampling methods are used to approximate the posterior distribution  $p(\boldsymbol{\theta}|D_1, D_2, \dots, D_N)$  of the state  $\theta$  given the available data  $\{D_1, D_2, \dots, D_N\}$ . Finally, Markov Chain Monte Carlo methods are suitable for sampling from the posterior which includes multiple states  $p(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_N|D_1, D_2, \dots, D_N)$ . Since we are interested in estimating a single parameter  $\theta$  which is the source location, we employ an importance sampling based Monte Carlo method [24].

At each iteration of the algorithm, the fusion center gathers the  $M$ -bit data (or its compressed version) from additional  $A$  non-anchor sensors. Let  $p(\boldsymbol{\theta}|\mathbf{W}_i)$  be the posterior pdf of the source location given the available data  $\mathbf{W}_i$  for iteration  $i$ ,  $i \geq 0$  (at step 1 in Fig. 3.3). In this chapter, we approximate  $p(\boldsymbol{\theta}|\mathbf{W}_i)$  using an importance sampling based Monte Carlo method as,

$$p(\boldsymbol{\theta}|\mathbf{W}_i) = \sum_{m=1}^{N_s} w^{m,i} \delta(\boldsymbol{\theta} - \boldsymbol{\theta}^{m,i}) \quad (3.17)$$

where the posterior distribution of source location is represented with  $N_s$  particles. The particles  $\boldsymbol{\theta}^{m,i} = [\theta_x^{m,i} \theta_y^{m,i}]^T$  ( $m = 1, 2, \dots, N_s$ ) are drawn from the distribution  $p_0(\boldsymbol{\theta})$  with equal weights  $w^{m,0} = 1/N_s$ . Let  $\tilde{w}^{m,i}$  be the weight of particle  $\boldsymbol{\theta}^{m,i}$  which is obtained according to [24],

$$\tilde{w}^{m,i} \propto p(\mathbf{W}_i|\boldsymbol{\theta}^{m,i})w^{m,0} \quad (3.18)$$

The initial weight of each particle is then multiplied with the likelihood function of the sensor data received up to the current iteration. Since the sensor decisions are conditionally independent,  $p(\mathbf{W}_i|\boldsymbol{\theta}^{m,i}) = p(D_1|\boldsymbol{\theta}^{m,i}) \times \dots \times p(D_{K+iA}|\boldsymbol{\theta}^{m,i})$ , and the likelihood function  $p(\mathbf{W}_i|\boldsymbol{\theta}^{m,i})$  can be computed from (3.3) and (3.10). The particle

weights are further normalized as,

$$w^{m,i} = \frac{\tilde{w}^{m,i}}{\sum_{m=1}^{N_s} \tilde{w}^{m,i}} \quad (3.19)$$

Then at the end of the  $i^{th}$  iteration, the Monte Carlo approach yields the source location estimate  $\hat{\boldsymbol{\theta}}_i$  as,

$$\hat{\boldsymbol{\theta}}_i = \sum_{m=1}^{N_s} w^{m,i} \boldsymbol{\theta}^{m,i} \quad (3.20)$$

For the next iteration, the particles are generated from the prior  $p_0(\boldsymbol{\theta})$  and weights are updated according to (3.18), using  $\mathbf{W}_{i+1}$ . Namely, we employ an importance-sampling based Monte Carlo method independently at each iteration using the entire received data to approximate the posterior distribution and update the source location. Having represented the posterior pdf of the source location, we can now describe the sensor selection methods.

### 3.2.2 Sensor Selection Methods

Let  $\mathbf{s}_{\mathbf{A}}^i = \left\{ \mathbf{s}_{\mathbf{A}}^{(i,1)}, \mathbf{s}_{\mathbf{A}}^{(i,2)}, \dots, \mathbf{s}_{\mathbf{A}}^{(i, C_A^{N-K-(i-1)A})} \right\}$  be the collection of all distinct  $A$ -element subsets of  $N - K - (i - 1)A$  remaining non-anchor sensors at iteration  $i$ .  $C_A^{N-K-(i-1)A}$  is the combination operation defined as,

$$C_A^{N-K-(i-1)A} = \frac{(N - K - (i - 1)A)!}{A!(N - K - iA)!} \quad (3.21)$$

Let  $\mathbf{s}_{\mathbf{A}}^{(\nu,i)}$  be the set of  $A$  non-anchor sensors activated at the  $i^{th}$  iteration according to the sensor selection strategy  $\nu$ . Then,  $\mathbf{s}_{\mathbf{A}}^{(\nu,i)} = \{s_1^{\nu,i}, s_2^{\nu,i}, \dots, s_A^{\nu,i}\}$  where  $s_k^{\nu,i}$  ( $k \in \{1, \dots, A\}$ ) is the  $k^{th}$  activated non-anchor sensor according to  $\nu$  at iteration  $i$  and  $\mathbf{D}_{\mathbf{A}}^{(\nu,i)} = \{D_1^{\nu,i}, D_2^{\nu,i}, \dots, D_A^{\nu,i}\}$  are the quantized measurements of  $\mathbf{s}_{\mathbf{A}}^{(\nu,i)}$ . Now, the objective is to find the optimal sensor selection strategy  $\nu^*$  which activates the set  $\mathbf{s}_{\mathbf{A}}^{(\nu^*,i)} = \{s_1^{\nu^*,i}, s_2^{\nu^*,i}, \dots, s_A^{\nu^*,i}\}$ . Corresponding to  $i^{th}$  iteration,  $\mathbf{s}_{\mathbf{A}}^{(\nu^*,i)}$  minimizes a certain cost



function  $\Psi_i(\cdot)$  as

$$\nu^* = \arg \min_{\substack{\nu \\ \mathbf{s}_A^{(\nu,i)}}} \Psi_i(\nu), \quad \nu \in \{1, 2, \dots, C_A^{N-K-(i-1)A}\} \quad (3.22)$$

In this work, we first select  $\Psi_i(\nu)$  as the negative of the mutual information between source location and the sensors to be selected and then we select  $\Psi_i(\nu)$  as the trace of the PCRLB matrix. Before describing the sensor selection methods in detail, we first discuss the relationship between mutual information and Fisher information briefly in the following subsection.

### The relationship between mutual information and Fisher Information

The relationship between mutual information and Fisher information was investigated in [92]. For ease of presentation, in this subsection we review this analysis for a scalar parameter  $\theta$ . The details of the analysis for vector-valued parameter is straightforward and can be found in [92].

Let  $I(\theta, \mathbf{D})$  represent the mutual information between the observations  $\mathbf{D}$  and state  $\theta$  and suppose there exists an unbiased estimator  $\hat{\theta}$  with mean  $\theta$  and variance  $\frac{1}{J(\theta)}$ . The amount of information gained about  $\theta$  in the computation of  $\hat{\theta}$  is,

$$I(\theta, \hat{\theta}) = H(\hat{\theta}) - H(\hat{\theta}|\theta) \quad (3.23)$$

where  $H(\hat{\theta})$  is the entropy of the estimator. Since the MMSE estimator is asymptotically efficient and unbiased [93], for each  $\theta$ , the latter term is smaller than the entropy of a Gaussian distribution with the same variance  $\frac{1}{J(\theta)}$ . This implies,

$$I(\theta, \hat{\theta}) \geq H(\hat{\theta}) - \int \frac{1}{2} \ln \left( \frac{2\pi e}{J(\theta)} \right) p(\theta) d\theta \quad (3.24)$$

where  $p(\theta)$  is the prior probability of  $\theta$ . Since processing can not increase the information,  $I(\theta, \mathbf{D}) \geq I(\theta, \hat{\theta})$ . In the limit, according to [92], the distribution of the estimator is sharply peaked around its mean value and the entropy of the estimator becomes identical to the entropy of the state ( $H(\hat{\theta}) \approx H(\theta)$ ). Then (3.24) becomes,

$$I(\theta, \mathbf{D}) \geq H(\theta) - \int \frac{1}{2} \ln \left( \frac{2\pi e}{J(\theta)} \right) p(\theta) d\theta \quad (3.25)$$

which shows that asymptotically the lower bound of the mutual information is a function of the Fisher information.

### Mutual information based sensor selection

An entropy based sensor selection method using particle filters was presented in [81] where sensor data are assumed to be analog. In this chapter, we extend the approach presented in [81] to deal with quantized sensor data. Let  $p_i(\boldsymbol{\theta})$  be the prior pdf of the source location as defined in (3.16). Besides the prior pdf of the source location, we also need to know the locations of non-anchor sensors and the sensing models of candidate sensors  $p(\mathbf{D}_{\mathbf{A}}^{(\nu,i)}|\boldsymbol{\theta})$ . Now, for iteration  $i$ , the objective is to find the optimal sensor activation scheme  $\nu^*$  which activates  $A$  sensors out of  $N - K - (i - 1)A$  remaining non-anchor sensors whose data minimize the expected conditional entropy of the posterior source location distribution,

$$\nu^* = \arg \min_{\nu} H(\boldsymbol{\theta}|\mathbf{D}_{\mathbf{A}}^{(\nu,i)}) \quad (3.26)$$

Let  $I(\boldsymbol{\theta}, \mathbf{D}_{\mathbf{A}}^{(\nu,i)}) = H(\boldsymbol{\theta}) - H(\boldsymbol{\theta}|\mathbf{D}_{\mathbf{A}}^{(\nu,i)})$  be the mutual information between the source location  $\boldsymbol{\theta}$  and the measurements of the activated sensors according to the activation scheme  $\nu$ . The sensor selection problem now turns into,

$$\nu^* = \arg \max_{\nu} I(\boldsymbol{\theta}, \mathbf{D}_{\mathbf{A}}^{(\nu,i)}) \quad (3.27)$$

$I(\boldsymbol{\theta}, \mathbf{D}_{\mathbf{A}}^{(\nu,i)})$  can also be expanded as,

$$I(\boldsymbol{\theta}, \mathbf{D}_{\mathbf{A}}^{(\nu,i)}) = H(\mathbf{D}_{\mathbf{A}}^{(\nu,i)}) - H(\mathbf{D}_{\mathbf{A}}^{(\nu,i)}|\boldsymbol{\theta}) \quad (3.28)$$

To compute (3.28) using Monte Carlo approximation, we start with writing the entropy

of  $\mathbf{D}_{\mathbf{A}}^{(\nu,i)}$ ,

$$H(\mathbf{D}_{\mathbf{A}}^{(\nu,i)}) = - \sum p(\mathbf{D}_{\mathbf{A}}^{(\nu,i)}) \log p(\mathbf{D}_{\mathbf{A}}^{(\nu,i)}) \quad (3.29)$$

$p(\mathbf{D}_{\mathbf{A}}^{(\nu,i)})$  can be decomposed as,

$$p(\mathbf{D}_{\mathbf{A}}^{(\nu,i)}) = \int_{\boldsymbol{\theta}} \left( \prod_{k=1}^A p(D_k^{\nu,i} | \boldsymbol{\theta}) \right) p_i(\boldsymbol{\theta}) d\boldsymbol{\theta} \quad (3.30)$$

where  $p_i(\boldsymbol{\theta})$  is the prior pdf of the source location and  $p(D_1^{\nu,i} | \boldsymbol{\theta}), p(D_2^{\nu,i} | \boldsymbol{\theta}), \dots, p(D_A^{\nu,i} | \boldsymbol{\theta})$  are the likelihood functions. Using (3.16) and (3.17) in (3.30) results in,

$$\begin{aligned} p(\mathbf{D}_{\mathbf{A}}^{(\nu,i)}) &= \int_{\boldsymbol{\theta}} \left( \prod_{k=1}^A p(D_k^{\nu,i} | \boldsymbol{\theta}) \right) \left( \sum_{m=1}^{N_s} w^{m,i-1} \delta(\boldsymbol{\theta} - \boldsymbol{\theta}^{m,i-1}) \right) d\boldsymbol{\theta} \\ &= \sum_{m=1}^{N_s} w^{m,i-1} \left( \prod_{k=1}^A p(D_k^{\nu,i} | \boldsymbol{\theta}^{m,i-1}) \right) \end{aligned} \quad (3.31)$$

Then using (3.31),  $H(\mathbf{D}_{\mathbf{A}}^{(\nu,i)})$  defined in (3.29) is rewritten as follows,

$$\begin{aligned} H(\mathbf{D}_{\mathbf{A}}^{(\nu,i)}) &= - \sum_{l_1=0}^{L-1} \dots \sum_{l_A=0}^{L-1} \left\{ \sum_{m=1}^{N_s} w^{m,i-1} \left( \prod_{k=1}^A p(D_k^{\nu,i} = l_k | \boldsymbol{\theta}^{m,i-1}) \right) \right\} \\ &\quad \times \log \left\{ \sum_{m=1}^{N_s} w^{m,i-1} \left( \prod_{k=1}^A p(D_k^{\nu,i} = l_k | \boldsymbol{\theta}^{m,i-1}) \right) \right\} \end{aligned} \quad (3.32)$$

Now let us compute the second term of (3.28). First we have,

$$H(\mathbf{D}_{\mathbf{A}}^{(\nu,i)} | \boldsymbol{\theta}) = - \int_{\boldsymbol{\theta}} \sum_{l_1, l_2, \dots, l_A} p(\mathbf{D}_{\mathbf{A}}^{(\nu,i)}, \boldsymbol{\theta}) \log p(\mathbf{D}_{\mathbf{A}}^{(\nu,i)} | \boldsymbol{\theta}) d\boldsymbol{\theta} \quad (3.33)$$

. Since  $p(\mathbf{D}_{\mathbf{A}}^{(\nu,i)}, \boldsymbol{\theta}) = p(\mathbf{D}_{\mathbf{A}}^{(\nu,i)} | \boldsymbol{\theta}) p_i(\boldsymbol{\theta})$ , we have

$$H(\mathbf{D}_{\mathbf{A}}^{(\nu,i)} | \boldsymbol{\theta}) = - \sum_{l_1=0}^{L-1} \dots \sum_{l_A=0}^{L-1} \left( \int_{\boldsymbol{\theta}} \left( \prod_{k=1}^A p(D_k^{\nu,i} = l_k | \boldsymbol{\theta}) \right) p_i(\boldsymbol{\theta}) \right) \quad (3.34)$$

$$\times \log \left( \prod_{k=1}^A p(D_k^{\nu,i} = l_k | \boldsymbol{\theta}) \right) d\boldsymbol{\theta}$$

Then using the Monte Carlo approximation of the prior source location pdf,  $H(\mathbf{D}_{\mathbf{A}}^{(\nu,i)} | \boldsymbol{\theta})$  becomes,

$$\begin{aligned} H(\mathbf{D}_{\mathbf{A}}^{(\nu,i)} | \boldsymbol{\theta}) &= - \sum_{k=1}^A \sum_{l=0}^{L-1} \left[ \int_{\boldsymbol{\theta}} p(D_k^{\nu,i} = l | \boldsymbol{\theta}) \left\{ \sum_{m=1}^{N_s} w^{m,i-1} \delta(\boldsymbol{\theta} - \boldsymbol{\theta}^{m,i-1}) \right\} \log(p(D_k^{\nu,i} = l | \boldsymbol{\theta})) d\boldsymbol{\theta} \right] \\ &= - \sum_{k=1}^A \sum_{l=0}^{L-1} \left[ \sum_{m=1}^{N_s} w^{m,i-1} p(D_k^{\nu,i} = l | \boldsymbol{\theta}^{m,i-1}) \log(p(D_k^{\nu,i} = l | \boldsymbol{\theta}^{m,i-1})) \right] \end{aligned} \quad (3.35)$$

Finally using (3.32) and (3.35), the mutual information function  $I(\boldsymbol{\theta}, \mathbf{D}_{\mathbf{A}}^{(\nu,i)})$  expressed in (3.28) is calculated as,

$$\begin{aligned} I(\boldsymbol{\theta}, \mathbf{D}_{\mathbf{A}}^{(\nu,i)}) &= - \sum_{l_1=0}^{L-1} \dots \sum_{l_A=0}^{L-1} \left[ \left\{ \sum_{m=1}^{N_s} w^{m,i-1} \left( \prod_{k=1}^A p(D_k^{\nu,i} = l_k | \boldsymbol{\theta}^{m,i-1}) \right) \right\} \times \right. \\ &\quad \left. \log \left\{ \sum_{m=1}^{N_s} w^{m,i-1} \left( \prod_{k=1}^A p(D_k^{\nu,i} = l_k | \boldsymbol{\theta}^{m,i-1}) \right) \right\} \right] + \\ &\quad \left[ \sum_{k=1}^A \sum_{l=0}^{L-1} \left\{ \sum_{m=1}^{N_s} w^{m,i-1} p(D_k^{\nu,i} = l | \boldsymbol{\theta}^{m,i-1}) \log(p(D_k^{\nu,i} = l | \boldsymbol{\theta}^{m,i-1})) \right\} \right] \end{aligned} \quad (3.36)$$

The quantity  $-I(\boldsymbol{\theta}, \mathbf{D}_{\mathbf{A}}^{(\nu,i)})$  is employed as  $\Psi_i(\nu)$  for sensor selection in (3.22).

### PCRLB based Sensor Selection

Let  $p(\mathbf{D}, \boldsymbol{\theta})$  be the joint probability density of the pair of  $(\mathbf{D}, \boldsymbol{\theta})$ . Then, the PCRLB of the estimation error has the form [13], [82],

$$E\{[\hat{\boldsymbol{\theta}}(\mathbf{D}) - \boldsymbol{\theta}][\hat{\boldsymbol{\theta}}(\mathbf{D}) - \boldsymbol{\theta}]^T\} \geq \mathbf{J}^{-1} \quad (3.37)$$

where  $\mathbf{J}$  is the  $2 \times 2$  Fisher information matrix (FIM)

$$\mathbf{J} = E[-\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p(\mathbf{D}, \boldsymbol{\theta})] \quad (3.38)$$

In (3.38),  $\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}}$  is the second derivative operator,

$$\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \triangleq \nabla_{\boldsymbol{\theta}} \nabla_{\boldsymbol{\theta}}^T \quad (3.39)$$

where  $\nabla_{\boldsymbol{\theta}}$  is the gradient operator with respect to  $\boldsymbol{\theta}$ .

Using the equality  $p(\mathbf{D}, \boldsymbol{\theta}) = p(\mathbf{D}|\boldsymbol{\theta})p_0(\boldsymbol{\theta})$ , an alternative expression for the Fisher information matrix can be written as,

$$\begin{aligned} \mathbf{J} &= E[-\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p(\mathbf{D}|\boldsymbol{\theta})] + E[-\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p_0(\boldsymbol{\theta})] \\ &= \mathbf{J}_d + \mathbf{J}_p \end{aligned} \quad (3.40)$$

In (3.40),  $\mathbf{J}_p \triangleq E[-\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p_0(\boldsymbol{\theta})]$  represents the *a priori* information, and  $\mathbf{J}_d \triangleq E[-\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p(\mathbf{D}|\boldsymbol{\theta})]$  is the standard FIM given in (3.15) further averaged over the prior pdf of the source location.

After initialization via the use of  $K$  anchor sensors, during each iteration the fusion center requests data from  $A$  non-anchor sensors that minimize the PCRLB. At iteration  $i$ , given available data  $\mathbf{W}_{i-1}$ , the PCRLB of  $A$  non-anchor sensors is expressed as,

$$E\{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T | \mathbf{W}_{i-1}\} \geq F^{-1}(\mathbf{D}_{\mathbf{A}}^{(\nu,i)} | \mathbf{W}_{i-1}) \quad (3.41)$$

where  $\mathbf{J}_c(\nu) = F(\mathbf{D}_{\mathbf{A}}^{(\nu,i)} | \mathbf{W}_{i-1})$  is the Fisher information matrix (FIM) of the random variable  $\boldsymbol{\theta}$  contained in  $\mathbf{D}_{\mathbf{A}}^{(\nu,i)}$  given available data  $\mathbf{W}_{i-1}$ . Then  $\mathbf{J}_c(\nu)$  is expressed as,

$$\mathbf{J}_c(\nu) = F(\mathbf{D}_{\mathbf{A}}^{(\nu,i)} | \mathbf{W}_{i-1}) \triangleq E\left\{ \left[ -\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p(\boldsymbol{\theta}, \mathbf{D}_{\mathbf{A}}^{(\nu,i)} | \mathbf{W}_{i-1}) \right] | \mathbf{W}_{i-1} \right\} \quad (3.42)$$

$$= - \int_{\boldsymbol{\theta}} \sum_{l_1, l_2, \dots, l_A} \left[ \Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p(\boldsymbol{\theta}, \mathbf{D}_{\mathbf{A}}^{(\nu,i)} | \mathbf{W}_{i-1}) \right] p(\boldsymbol{\theta}, \mathbf{D}_{\mathbf{A}}^{(\nu,i)} | \mathbf{W}_{i-1}) d\boldsymbol{\theta}$$

where we take expectation over all possible source locations  $\boldsymbol{\theta}$  and all quantized sensor measurements  $\{l_1, l_2, \dots, l_A\}$ .

Using Bayesian decomposition, the joint probability density function  $p(\boldsymbol{\theta}, \mathbf{D}_{\mathbf{A}}^{(\nu,i)} | \mathbf{W}_{i-1})$  of source location  $\boldsymbol{\theta}$  and new quantized measurements  $\mathbf{D}_{\mathbf{A}}^{(\nu,i)}$  is written as,

$$p(\boldsymbol{\theta}, \mathbf{D}_{\mathbf{A}}^{(\nu,i)} | \mathbf{W}_{i-1}) = p(\mathbf{D}_{\mathbf{A}}^{(\nu,i)} | \boldsymbol{\theta}) p(\boldsymbol{\theta} | \mathbf{W}_{i-1}) \quad (3.43)$$

where the identity  $p(\mathbf{D}_{\mathbf{A}}^{(\nu,i)} | \mathbf{W}_{i-1}, \boldsymbol{\theta}) = p(\mathbf{D}_{\mathbf{A}}^{(\nu,i)} | \boldsymbol{\theta})$  has been used. Using the above decomposition, (3.42) can be written as,

$$\mathbf{J}_c(\nu) = - \int_{\boldsymbol{\theta}} \sum_{l_1, l_2, \dots, l_A} \Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log \left\{ p(\mathbf{D}_{\mathbf{A}}^{(\nu,i)} | \boldsymbol{\theta}) p(\boldsymbol{\theta} | \mathbf{W}_{i-1}) \right\} p(\mathbf{D}_{\mathbf{A}}^{(\nu,i)} | \boldsymbol{\theta}) p(\boldsymbol{\theta} | \mathbf{W}_{i-1}) d\boldsymbol{\theta} \quad (3.44)$$

which can be decomposed as,

$$\begin{aligned} \mathbf{J}_c(\nu) = & - \left[ \int_{\boldsymbol{\theta}} \sum_{l_1, l_2, \dots, l_A} \left\{ \Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p(\mathbf{D}_{\mathbf{A}}^{(\nu,i)} | \boldsymbol{\theta}) \right\} p(\mathbf{D}_{\mathbf{A}}^{(\nu,i)} | \boldsymbol{\theta}) p(\boldsymbol{\theta} | \mathbf{W}_{i-1}) d\boldsymbol{\theta} \right. \\ & \left. + \int_{\boldsymbol{\theta}} \left\{ \Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p(\boldsymbol{\theta} | \mathbf{W}_{i-1}) \right\} p(\boldsymbol{\theta} | \mathbf{W}_{i-1}) \left\{ \sum_{l_1, l_2, \dots, l_A} p(\mathbf{D}_{\mathbf{A}}^{(\nu,i)} | \boldsymbol{\theta}) \right\} d\boldsymbol{\theta} \right] \end{aligned} \quad (3.45)$$

Since  $p(\mathbf{D}_{\mathbf{A}}^{(\nu,i)} | \boldsymbol{\theta}) = \prod_{k=1}^A p(D_k^{\nu,i} | \boldsymbol{\theta})$  and  $\sum_{l_1, l_2, \dots, l_A} p(\mathbf{D}_{\mathbf{A}}^{(\nu,i)} | \boldsymbol{\theta}) = 1$ , (3.45) becomes,

$$\begin{aligned} \mathbf{J}_c(\nu) = & - \left[ \int_{\boldsymbol{\theta}} \sum_{l_1=0}^{L-1} \dots \sum_{l_A=0}^{L-1} \left( \left\{ \sum_{k=1}^A \Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p(D_k^{\nu,i} = l_k | \boldsymbol{\theta}) \right\} \prod_{k=1}^A p(D_k^{\nu,i} | \boldsymbol{\theta}) \right) p(\boldsymbol{\theta} | \mathbf{W}_{i-1}) d\boldsymbol{\theta} \right. \\ & \left. + \int_{\boldsymbol{\theta}} \Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log \{ p(\boldsymbol{\theta} | \mathbf{W}_{i-1}) \} p(\boldsymbol{\theta} | \mathbf{W}_{i-1}) d\boldsymbol{\theta} \right] \end{aligned} \quad (3.46)$$

Finally (4.42) reduces to,

$$\begin{aligned} \mathbf{J}_c(\nu) = & - \left\{ \int_{\boldsymbol{\theta}} \sum_{k=1}^A \sum_{l=0}^{L-1} \left\{ [\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p(D_k^{\nu,i} = l | \boldsymbol{\theta})] p(D_k^{\nu,i} = l | \boldsymbol{\theta}) \right\} p(\boldsymbol{\theta} | \mathbf{W}_{i-1}) d\boldsymbol{\theta} \right. \\ & \left. + \int_{\boldsymbol{\theta}} [\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p(\boldsymbol{\theta} | \mathbf{W}_{i-1})] p(\boldsymbol{\theta} | \mathbf{W}_{i-1}) d\boldsymbol{\theta} \right\} \end{aligned} \quad (3.47)$$

Note that the result of the double summation term,  $\sum_{k=1}^A \sum_{l=0}^{L-1} [\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p(D_k^{\nu,i} = l | \boldsymbol{\theta})] p(D_k^{\nu,i} = l | \boldsymbol{\theta}) =$

$l|\boldsymbol{\theta}$ ) has been derived in [44] provided by (3.15). For the first term of (3.47), we use (3.16) and (3.17) to approximate  $p(\boldsymbol{\theta}|\mathbf{W}_{i-1})$ . The second term requires the second derivative of  $p(\boldsymbol{\theta}|\mathbf{W}_{i-1})$ . Since  $p(\boldsymbol{\theta}|\mathbf{W}_{i-1})$  has a non-parametric representation by a set of random samples with associated weights, it is very difficult to express the exact form of its second order derivatives. We propose two alternative methods to compute the FIM of the prior as follows:

### Numerical Computation for the FIM of the prior pdf

Let us define,

$$\boldsymbol{\Gamma}_i \triangleq \int_{\boldsymbol{\theta}} [\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p(\boldsymbol{\theta}|\mathbf{W}_{i-1})] p(\boldsymbol{\theta}|\mathbf{W}_{i-1}) d\boldsymbol{\theta} \quad (3.48)$$

We can calculate the (1, 1) element of  $\boldsymbol{\Gamma}_i$  first.

$$\boldsymbol{\Gamma}_i(1, 1) = - \int_x \int_y \frac{1}{p(\boldsymbol{\theta}|\mathbf{W}_{i-1})} \left( \frac{\partial p(\boldsymbol{\theta}|\mathbf{W}_{i-1})}{\partial x} \right)^2 dy dx$$

Let  $\mathcal{A}^2$  be the area of the region of interest (ROI). We partition the ROI into  $G$  equal size cells where the area of each cell is  $\delta^2 = \mathcal{A}^2/G$  and  $\delta$  is the distance between the centers of each neighboring cell.

Let  $p(\boldsymbol{\theta} \in \{c, q\}|\mathbf{W}_{i-1})$  be the probability of a particular cell specified by the cell indices  $c$  and  $q$ , then

$$p(\boldsymbol{\theta} \in \{c, q\}|\mathbf{W}_{i-1}) = p \{x^{m,i} \in [c\delta, (c+1)\delta] \& y^{m,i} \in [q\delta, (q+1)\delta]\}$$

where

$$c \in \{0, 1, \dots, \sqrt{G} - 1\}, q \in \{0, 1, \dots, \sqrt{G} - 1\}$$

Denote  $V_{c,q}$  as the total number of particles inside the cell specified by  $c$  and  $q$  where each particle has the weight  $w_v^{c,q}$ . Then,

$$p(\boldsymbol{\theta} \in \{c, q\}|\mathbf{W}_{i-1}) = \sum_{v=1}^{V_{c,q}} w_v^{c,q} \quad (3.49)$$

Then  $\Gamma_i(1, 1)$  can be approximated as follows,

$$\Gamma_i(1, 1) \approx - \sum_{c=0}^{\sqrt{G}-1} \sum_{q=0}^{\sqrt{G}-1} \frac{1}{p(\boldsymbol{\theta} \in \{c, q\} | \mathbf{W}_{i-1})} \left( \frac{p(\boldsymbol{\theta} \in \{c+1, q\} | \mathbf{W}_{i-1}) - p(\boldsymbol{\theta} \in \{c, q\} | \mathbf{W}_{i-1})}{\delta} \right)^2 \delta^2 \quad (3.50)$$

Using the above procedure  $\Gamma_i(1, 2)$ ,  $\Gamma_i(2, 1)$  and  $\Gamma_i(2, 2)$  can be computed similarly. Note that calculation of  $\Gamma_i$  is independent of the number of sensors to be selected ( $A$ ).

Using the approximations presented in (3.17) and the numerical approximation for the FIM of the prior pdf (3.50), (3.47) is rewritten as follows,

$$\mathbf{J}_c(\nu) = - \left[ \left\{ \sum_{m=1}^{N_s} w^{m,i-1} \sum_{k=1}^A \sum_{l=0}^{L-1} [\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p(D_k^{\nu,i} = l | \boldsymbol{\theta}^{m,i-1})] p(D_k^{\nu,i} = l | \boldsymbol{\theta}^{m,i-1}) \right\} - \Gamma_i \right] \quad (3.51)$$

### Gaussian approximation for the FIM of the prior pdf:

The posterior pdf represented with the particles and their associated weights can be also approximated as a Gaussian distribution as,

$$\begin{aligned} \log p(\boldsymbol{\theta} | \mathbf{W}_{i-1}) &\approx -\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{\theta} - \boldsymbol{\mu}_i) \\ \nabla_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p(\mathbf{W}_{i-1} | \boldsymbol{\theta}) &\approx -\boldsymbol{\Sigma}_i^{-1} \end{aligned} \quad (3.52)$$

where

$$\begin{aligned} \boldsymbol{\mu}_i &= \sum_{m=1}^{N_s} w^{m,i} \boldsymbol{\theta}^{m,0} \\ \boldsymbol{\Sigma}_i &= \sum_{m=1}^{N_s} w^{m,i} (\boldsymbol{\theta}^{m,0} - \boldsymbol{\mu}_i)(\boldsymbol{\theta}^{m,0} - \boldsymbol{\mu}_i)^T \end{aligned} \quad (3.53)$$

Using the approximations presented in (3.17) and the Gaussian approximation of the



prior pdf (3.52), (3.47) is rewritten as follows,

$$\mathbf{J}_c(\nu) = - \left[ \left\{ \sum_{m=1}^{N_s} w^{m,i-1} \sum_{k=1}^A \sum_{l=0}^{L-1} [\Delta_{\theta}^{\theta} \log p(D_k^{\nu,i} = l | \boldsymbol{\theta}^{m,i-1})] p(D_k^{\nu,i} = l | \boldsymbol{\theta}^{m,i-1}) \right\} - \boldsymbol{\Sigma}_i^{-1} \right] \quad (3.54)$$

The result of the two inner summations in the first term of (3.51) and (3.54) is basically the FIM defined in (3.14) which is then averaged over the prior distribution of the source location represented by the particles. For the activation strategy  $\nu$ , we calculate its corresponding FIM  $\mathbf{J}_c(\nu)$  as defined in (3.51). The fusion center then decides on the optimal sensor activation strategy  $\nu^*$  that minimizes the trace of  $\mathbf{J}_c^{-1}(\nu)$  which is the PCRLB corresponding to the summation of the MSE of the estimates of  $x$  and  $y$ .

Note that the mutual information function defined in (3.36) requires  $N_s \times L^A + A \times L \times N_s$  summations. In comparison, the FIM function defined in (3.51) requires  $A \times L \times N_s$  summations. In other words, since (3.36) requires an  $A$ -fold summation, the computational complexity of the mutual information based sensor selection scheme increases exponentially with  $A$  while the computational complexity of PCRLB based sensor selection scheme increases linearly with  $A$ .

### Nearest Sensor Selection

For performance comparison, we also consider selecting the  $A$  nearest sensors to the estimated source location  $\hat{\boldsymbol{\theta}}$ . Let  $\mathbf{x}_{\mathbf{A}}^{\nu,i} = [x_1^{\nu,i}, x_2^{\nu,i}, \dots, x_A^{\nu,i}]^T$  and  $\mathbf{y}_{\mathbf{A}}^{\nu,i} = [y_1^{\nu,i}, y_2^{\nu,i}, \dots, y_A^{\nu,i}]^T$  represent the coordinates of the  $\nu^{th}$  sensor set to be activated ( $\mathbf{s}_{\mathbf{A}}^{\nu,i}$ ) at iteration  $i$ . The fusion center activates  $A$  non-anchor set according to  $\nu^*$  which minimizes,

$$\nu^* = \arg \min_{\mathbf{s}_{\mathbf{A}}^{\nu,i}} \left( \sum_{k=1}^A \sqrt{(\hat{\theta}_x - x_k^{\nu,i})^2 + (\hat{\theta}_y - y_k^{\nu,i})^2} \right) \quad (3.55)$$

where  $\hat{\theta}_x$  and  $\hat{\theta}_y$  are the coordinates of the estimated source location  $\hat{\boldsymbol{\theta}}$ .

### 3.3 Sensor Data Compression

In this section, distributed source coding techniques are discussed which use the posterior pdf of the source location to further compress the data transmitted by the activated sensors. Let  $s_k^{\nu^*,i}$ , ( $k \in \{1, 2, \dots, A\}$ ) be a non-anchor sensor which is activated according to the sensor selection strategy  $\nu^*$  at iteration  $i$ . Using the Monte Carlo approximation of the posterior pdf of the source location, probability of receiving a certain data  $l$  from  $s_k^{\nu^*,i}$  is  $p(D_k^{\nu^*,i} = l)$  and expressed as,

$$\begin{aligned} p(D_k^{\nu^*,i} = l) &= \int_{\theta} p(D_k^{\nu^*,i} = l, \theta) d\theta \\ &= \int_{\theta} p(D_k^{\nu^*,i} = l | \theta) p_i(\theta) d\theta \\ &= \sum_{m=1}^{N_s} w^{m,i-1} p(D_k^{\nu^*,i} = l | \theta^{m,i-1}) \end{aligned} \quad (3.56)$$

Let  $H(D_k^{\nu^*,i})$  [94] be the conditional entropy of a non-anchor sensor  $D_k^{\nu^*,i}$  which is defined as,

$$H(D_k^{\nu^*,i}) = - \sum_{l=0}^{L-1} p(D_k^{\nu^*,i} = l) \log_2 p(D_k^{\nu^*,i} = l) \quad (3.57)$$

The fusion center requests the  $M$ -bit data of each non-anchor sensor to be activated in  $B_k$  bits which has to satisfy,

$$B_k \geq H(D_k^{\nu^*,i}) \quad k \in \{1, 2, \dots, A\}. \quad (3.58)$$

As an example, in Fig. 3.4, we present the conditional entropies of  $N - K = 345$  non-anchor sensors for the first iteration given the decisions of  $K = 16$  anchor sensors as depicted in Fig. 3.2 and the source is located at  $[75m.75m.]$ . Simulation results show that the sensors close to the actual source location have high entropies. As the sensor distance from the source location increases, quantized observations of the sensors tend to zero and no matter what the side information is, conditional entropy of such a sensor decreases and goes to zero. Note that for the  $M = 6$  bit case, the asymptotic entropy of each non-anchor sensor far away from the source is around 1 due to the noise fluctuations. This means that only a small subset of the sensors contain information of the source location.

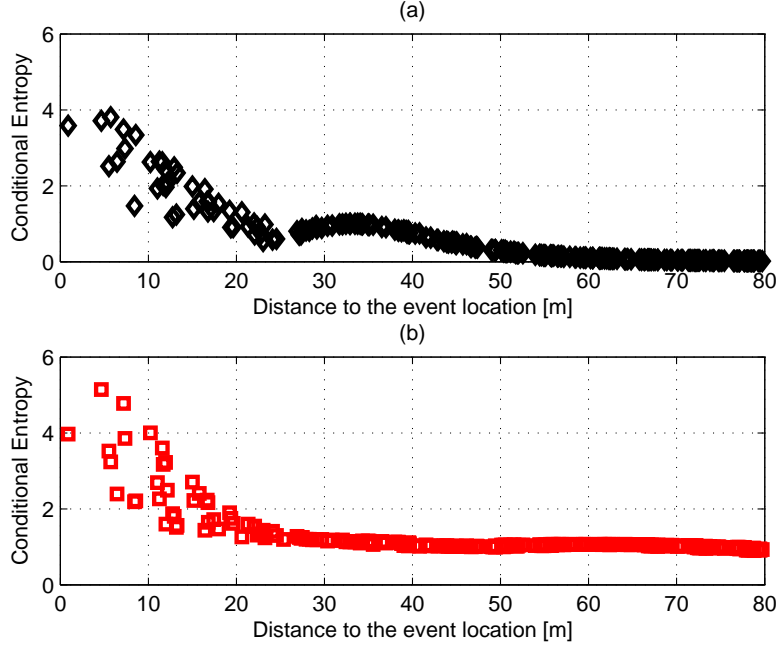


Figure 3.4: Conditional Entropy of non-anchor sensors in the field given the multi-bit decisions of the anchor sensors at the beginning of the first iteration. (a)  $M = 5$  bit, (b)  $M = 6$  bit

In this work, we select,

$$B_k = \min \left\{ \left( \lceil H(D_k^{\nu^*, i}) \rceil + 1 \right), M \right\} \quad (3.59)$$

where  $\lceil \cdot \rceil$  is the round towards next integer operator or the ceiling function. Using an approximate posterior pdf for the source location makes the conditional entropy of each sensor defined in (3.57) also approximate. According to the structure of our iterative method, any decoding error at a particular iteration may cause error propagation at the subsequent iterations. Therefore, in order to ensure lossless data compression, we include an extra guard bit to the approximated entropy of each sensor to be activated.

Let  $u_k$  be the  $B_k$ -bit compressed sensor data which is obtained from its actual  $M$ -bit sensor observation  $D_k^{\nu^*, i}$  according to [89] as,

$$u_k = D_k^{\nu^*, i} \bmod 2^{B_k} \quad (3.60)$$

where we assume that  $u_k$  is delivered to the fusion center without any error.

The fusion center generates the decision vector  $L_u$  which includes all the possible multi-bit decisions  $l$  which yield  $u_k$  as a remainder after the modulo  $2^{B_k}$  operation.

$$L_u = [l : \forall l, u_k = l \bmod 2^{B_k}] \quad (3.61)$$

Using the past information  $\mathbf{W}_{i-1}$  as side information, the multi-bit decision of each sensor is recovered with a simple maximum a posteriori probability (MAP) rule,

$$l^* = \arg \max_{l \in L_u} p(D_k^{\nu^*, i} = l | \mathbf{W}_{i-1}) \quad (3.62)$$

where  $p(D_k^{\nu^*, i} = l | \mathbf{W}_{i-1})$  is calculated according to (3.56). As an example, suppose that  $M = 4$  and  $L = 2^4$ . Let the quantized data of an activated sensor be  $D_k = 12$ . If the quantized data of the activated sensor is requested in  $B_k = 3$  bits, then  $u_k = 12 \bmod 2^3$ . Fusion center receives  $u_k = 4$ , finds out that  $L_u = \{4, 12\}$  and computes the following probabilities,  $p(D_k = 4 | \mathbf{W}_{i-1})$  and  $p(D_k = 12 | \mathbf{W}_{i-1})$  according to (3.56). The fusion center then picks either 4 or 12, depending on which has the largest probability.

After recovering the decision of each activated sensor,  $D_k^{\nu^*, i} = l^*$ , at iteration  $i$ , the fusion center updates the new posterior pdf  $p(\boldsymbol{\theta} | \mathbf{W}_i)$  using the procedure described in Section 3.2.

### 3.4 Simulation Results

In this section, we first study the detection performance for different anchor sensor layouts. We compare the computational cost of the two sensor selection schemes presented, then we give some illustrative examples to show their source location estimation performances. Trade-off between estimation performance and communication cost is finally studied.

### 3.4.1 Source Detection Performance

In our examples, we consider the source energy and signal decay exponent as  $P_0 = 25000$  and  $n = 2$  respectively. The prior pdf of the source location  $p_0(\cdot)$  is assumed to be a Gaussian with  $\mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$  where  $\boldsymbol{\mu}_0 = [50 \text{ m. } 50 \text{ m.}]^T$  and  $\boldsymbol{\Sigma}_0 = \begin{bmatrix} \sigma_{x,0}^2 & 0 \\ 0 & \sigma_{y,0}^2 \end{bmatrix}$  where  $3\sigma_{x,0} = 3\sigma_{y,0} = 50 \text{ m.}$  and  $\mathbf{J}_p = \boldsymbol{\Sigma}_0^{-1}$ . The selection of  $K$  anchor sensors determine the event detection performance. We assume the  $K$  anchor sensors are deployed in a  $100 \times 100 \text{ m}^2$  field in a grid layout as shown in Fig. 3.5 where the layout is specified by the inter-sensor distance (ISD) and the distance of the nearest sensor to the mean source location (D-NS-MSL). We fix the placement of the bottom-left sensor to the point  $(0, 0)$  and consider 11 different layouts by varying the ISD. The ISD and D-NS-MSL properties of each layout is shown in Table 3.1

Table 3.1: The sensor layouts to evaluate the detection performance.

Layout	$K$	ISD (m.)	D-NS-MSL (m.)
1	16	33.333	23.57
2	9	38.889	15.713
3	9	44.444	7.8567
4	9	50	0
5	4	55.556	7.8567
6	4	61.111	15.713
7	4	66.667	23.57
8	4	72.222	31.427
9	4	77.778	39.284
10	4	83.333	47.14
11	4	88.889	54.997

We assume that the two hypotheses  $H_0$  and  $H_1$  are equally likely, that is  $P_0 = P_1 = 0.5$ . Then the decision threshold at the fusion center becomes  $\log P_0/P_1 = 0$ . For 1000 different trials under  $H_1$ , Fig. 3.6 shows the detection performance corresponding to the anchor sensor layouts presented in Table 3.1.

Notice that in layout 4, one sensor is placed exactly at the mean source location, so for  $M \geq 4$  the source is detected with  $P_D \approx 1$ . Layout 3 and layout 5 also contain a

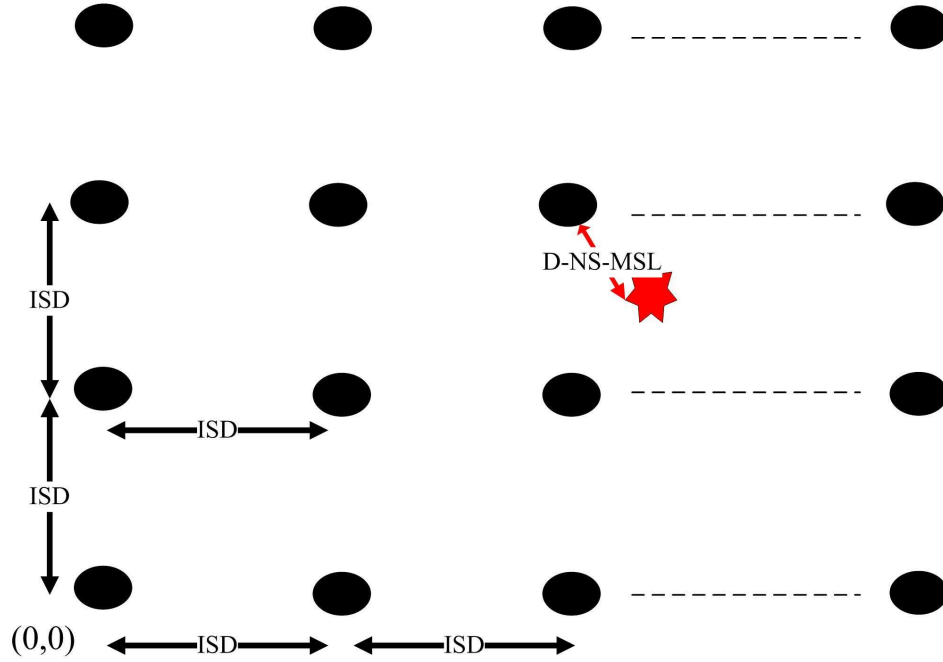


Figure 3.5: Grid Sensor Layout specified by inter-sensor distance (ISD) and distance of the nearest sensor to the mean source location (D-NS-MSL). (Black points: Sensors, Red Star: Source)

sensor that is very close to the mean source location ( $\approx 7 m.$ ), so the source is detected with very high probability of detection. Decreasing  $M$ , reduces the information gathered from the network, so for a given layout, the detection performance decreases gradually. For the layouts between 7 and 11, the ISD, hence the D-NS-MSL increases. So for  $K = 4$  with ISD greater than 66.667 meters, due to the isotropic signal emission model none of the sensors hear from the source, and the detection performance gets worse. Similarly, for  $M = 3$  and  $M = 4$ , the detection performance of layout 1 and 2 are worse than that of layout 3 and 4 since anchor sensors are far away from the mean source location. On the other hand, for layouts 1 and 2, as  $M$  increases, several sensors in the region become informative due to the increased number of quantization levels.

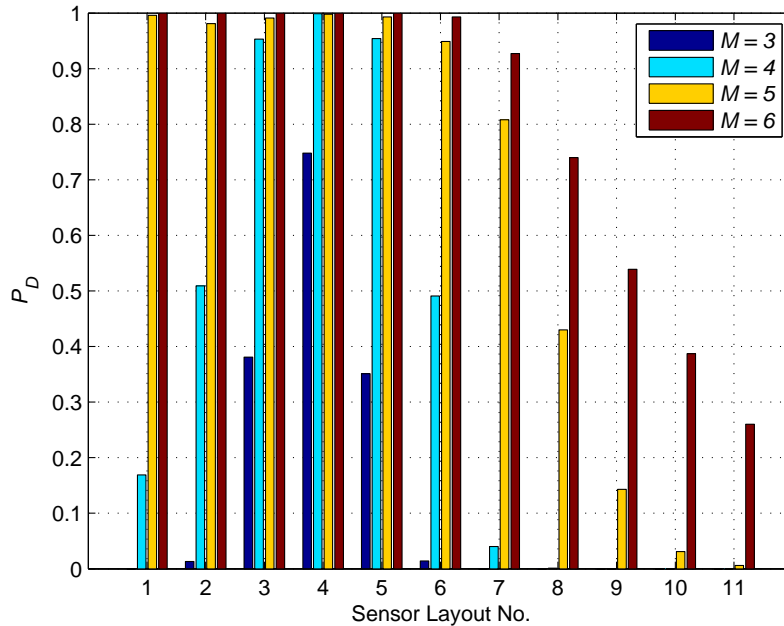


Figure 3.6: Detection performance of anchor sensor layouts

Therefore, the source is detected with  $P_D \approx 1$ .

In our simulations, we also computed  $P_F$ . After averaging over 1000 trials, we obtain  $P_F \approx 0$  for all 11 layouts.

### 3.4.2 Computational Cost

In this subsection, we compare the computation time of two sensor selection schemes. The mutual information based sensor selection method uses (3.36) to evaluate the mutual information between the source location and the sensor measurements. The PCRLB based sensor selection method uses (3.51) and calculates the trace of the PCRLB matrix. We use MATLAB's `cputime` function to calculate the computation times of functions (3.36) and (3.51). Table 3.2 and Fig. 3.7 show the average computation times of the two methods. The results are averaged over 100 different executions of each function. The CPU times are obtained on a computer with 2.1 GHz processor.

For  $A = 1$ , (3.36) is much simpler than (3.51) so (3.36) is computed faster than (3.51). On the other hand, for  $A \geq 2$ , the computational complexity of MI increases

Table 3.2: Mean CPU times of MI and PCRLB

	PCRLB	MI
$A = 1$	0.26 s.	0.17 s.
$A = 2$	0.51 s.	9.49 s.
$A = 3$	0.75 s.	371.79 s.
$A = 4$	1.00 s.	15544.00 s.

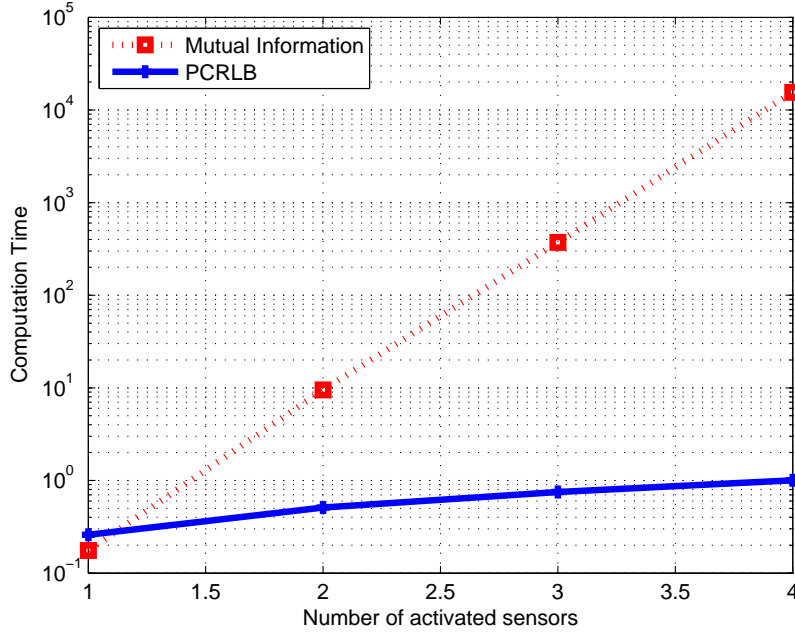


Figure 3.7: Mean computation times of objective functions

exponentially with  $L^A$  while the computation time of PCRLB increases linearly with  $A$  as  $L \times A$ . Note that for the  $i^{th}$  iteration of the algorithm, the selection of  $A$  optimal sensors has a search set of size  $C_A^{N-K-(i-1)A}$  which is the same for the two sensor selection schemes. In a dense network, activating a large number of sensors may result in a large search space and it may take a long time to find the optimal sensor selection strategy.

### 3.4.3 Iterative Location Estimation Performance

In this section,  $N = 19 \times 19 = 361$  sensors are deployed in a  $100 \times 100m^2$  field and the sensors are deployed in a grid where the location of each sensor is assumed to be



known. We initialize the iterative algorithm with  $K = 4 \times 4 = 16$  anchor sensors deployed in a grid layout where the distance between two-anchor sensor is  $33.5m$  as shown in Fig. 3.2. We select  $N_s = 10000$  particles and the particles  $\boldsymbol{\theta}^{m,i}$  are also drawn from  $\mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$  where  $w^{m,0} = 1/N_s$ . The mean squared error (MSE) matrix of the estimation is calculated as follows,

$$\text{MSE} = \frac{1}{Z} \sum_{z=1}^Z (\hat{\boldsymbol{\theta}}_z - \boldsymbol{\theta}_z)(\hat{\boldsymbol{\theta}}_z - \boldsymbol{\theta}_z)^T \quad (3.63)$$

We tested our algorithm over  $Z = 100$  different source locations  $\boldsymbol{\theta}_z$  drawn from the prior distribution  $\mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ .

### Calculation of the FIM of the prior

In Fig. 3.8, we present  $\Gamma(1, 1)$  and  $\Gamma(2, 2)$  according to (3.50) as a function of  $\delta$  which is the distance between the centers of each neighboring cell as defined in (3.50). While computing (3.51), the numerical approximation of the FIM of the prior should not manipulate the overall metric. Thus, we require a stable region of  $\delta$  where  $\Gamma$  needs to be obtained consistently. Fig. 3.8 shows that delta is stable within the interval  $\delta \in \{0.5, 2\}$ . Thus, in our experiments we select  $\delta = 1$ .

We employ two approaches to compute the FIM of the prior, namely by using numerical computation and the Gaussian approximation. The Gaussian approximation is computationally simpler. In Fig. 3.9, we present the MSE performance of PCRLB based sensor selection using both methods for different values of  $M$ . For  $M = 3$ , the numerical computation of  $\int_{\boldsymbol{\theta}} [\Delta_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p(\boldsymbol{\theta} | \mathbf{W}_{i-1})] p(\boldsymbol{\theta} | \mathbf{W}_{i-1}) d\boldsymbol{\theta}$  performs better than the Gaussian approximation. Simulation results show that, for  $M = 4$  and  $M = 5$  cases both schemes give similar MSE performance at each iteration of the algorithm.

### Estimation Performance

In Fig. 3.10, we present the mean squared error (MSE) of the estimation using MI and PCRLB based sensor selection methods without data compression at each activated sensor. The experimental MSE obtained above is also compared with the PCRLB

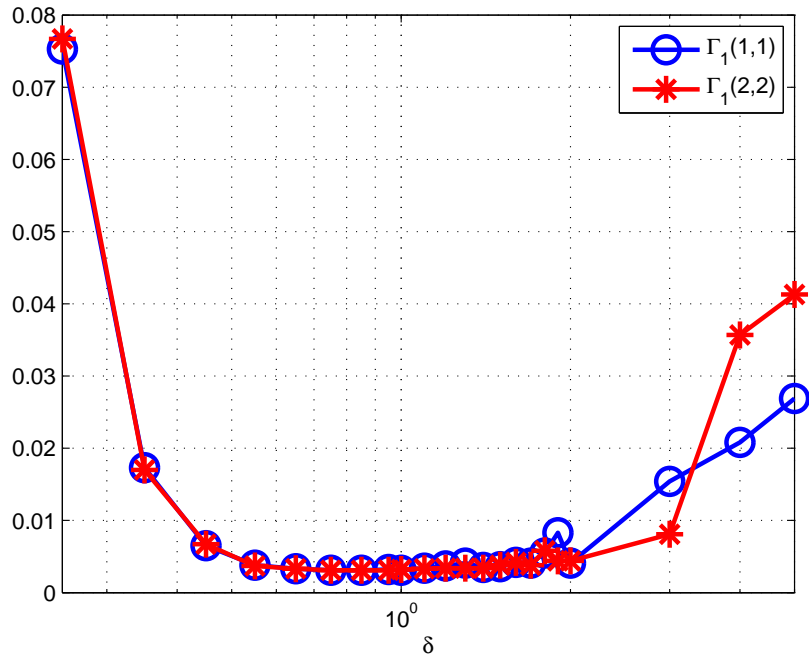


Figure 3.8:  $\Gamma(1,1)$  and  $\Gamma(2,2)$  at the first iteration as a function of  $\delta$  ( $\sqrt{G} = 100$ ,  $M = 5$ ).

found when all the  $N = 361$  sensors send their  $M$ -bit quantized data to the fusion center as defined in (3.40) and an iterative method which selects  $A$  non-anchor sensors nearest to the estimated source location as defined in (3.55). In our simulations, we activate  $A = 1$  sensor at a time after the initialization via anchor sensors. Simulation results show that, when  $M = 3$ , the MI and PCRLB based sensor selection schemes are the optimal sensor selection scheme and outperform the nearest sensor selection scheme in terms of MSE. For  $M = 4$ , PCRLB and MI based sensor selection are slightly better than nearest sensor selection. As  $M$  further increases, measurements of each activated sensor become more informative and the nearest, MI and PCRLB based sensor selection schemes achieve similar performance. For  $M = 5$  and  $M = 6$ , the MSE gets close to the PCRLB of  $N$  sensor data in about 5 iterations by activating only the most informative sensor about the source location during each iteration. Therefore, instead of using  $N = 361$  sensors, 21 sensors are enough to achieve a performance close to that when all the  $N = 361$  sensors send their data to the fusion center. Moreover, increasing  $A = 1$  to  $A = 2$  decreases the MSE of each iteration since we have more sensor data. In Figure

3.11(a), we present the MSE of estimation using MI and PCRLB based sensor selection methods. The experimental MSE is also compared with the PCRLB found when all  $N = 361$  sensors send their  $M = 5$ -bit quantized data to the fusion center. The PCRLB and MI-based sensor selection gives similar MSE performance for  $A = 1$  and  $A = 2$ . Furthermore, as shown in Fig. 3.11(b), for the  $A = 2$  and  $A = 3$  cases, simulation results show that the MSE gets close to the PCRLB computed with all sensor data more quickly in about 3 and 2 iterations respectively.

### Data Compression Performance

In Table 3.3, we compare the MSE at the end of the 9<sup>th</sup> iteration obtained by using location estimator based on compressed data to that based on data without compression. Source localization with compressed data achieves almost the same performance as that without data compression, which implies that the compressed sensor measurements are decoded almost perfectly at each iteration. For performance evaluation, we define two metrics: Compression Gain ( $CG_M(i)$ ) is the ratio between the average number of bits saved and the fixed number of bits  $M$  for iteration  $i$ ,

$$CG_M(i) = 1 - \frac{\bar{B}_k(i)}{M} \quad (3.64)$$

where  $\bar{B}_k(i)$  is the average number of bits transmitted to the fusion center at the  $i^{\text{th}}$  iteration. The overall compression gain ( $OCG_M$ ) is then defined as the average total number of bits saved by the proposed compression scheme and the fixed number of  $M$  bits until the end of the 9<sup>th</sup> iteration.

$$OCG_M = 1 - \frac{\sum_{i=1}^9 \bar{B}_k(i)}{9M} \quad (3.65)$$

Results presented in Table 3.4 show that, for  $M > 4$ , about 40% of the bits are saved by compression. At the beginning of the algorithm, there is a relatively large uncertainty about the source location so the measurements of the sensors selected at the beginning of the algorithm are transmitted to the fusion center in almost  $M$ -bits so that is why CG is small during the first few iterations. For the particular case illustrated

in Fig. 3.10 after the 3<sup>rd</sup> iteration, the MSE of the algorithm decreases rapidly which is the time when most of the informative sensors about the source location are selected. Then there is no need to send full  $M$ -bit information to the fusion center. As the fusion center learns more about the source location and the most informative sensors are selected, the uncertainty regarding source location gets smaller, and the conditional entropy defined in (3.57) becomes very small. After the most informative sensors have been selected, the CG increases to around 50% for  $M > 3$ .

Table 3.3: Final MSE at the end of the 9<sup>th</sup> iteration.  $A = 1$ , PCRLB based sensor selection.

	$M = 3$ x axis	$M = 3$ y axis	$M = 4$ x axis	$M = 4$ y axis
No Data Compression	0.3888	0.2993	0.0495	0.0427
Data Compression	0.3369	0.1775	0.0500	0.0520
PCRLB of $N$ sensor data	0.0381	0.0396	0.0183	0.0182
	$M = 5$ x axis	$M = 5$ y axis	$M = 6$ x axis	$M = 6$ y axis
No Data Compression	0.0270	0.0296	0.0155	0.0145
Data Compression	0.0314	0.0259	0.0117	0.0169
PCRLB of $N$ sensor data	0.0089	0.0091	0.0048	0.0048

Table 3.4:  $A = 1$ , Average number of bits used to represent the  $M = 3$ ,  $M = 4$ ,  $M = 5$  and  $M = 6$  bit sensor data.

Iteration	$M = 3$	$CG_3$	$M = 4$	$CG_4$	$M = 5$	$CG_5$	$M = 6$	$CG_6$
1	3.0000	0	3.9700	0.0075	4.5000	0.1000	5.8673	0.0221
2	2.9900	0.0033	3.9800	0.0050	4.9000	0.0200	5.8469	0.0255
3	2.8600	0.0467	3.5800	0.1050	4.1600	0.1680	4.2653	0.2891
4	2.5300	0.1567	2.7600	0.3100	2.9700	0.4060	3.2245	0.4626
5	2.2200	0.2600	2.2700	0.4325	2.2900	0.5420	2.9694	0.5051
6	2.1000	0.3000	2.0500	0.4875	2.0900	0.5820	2.9184	0.5136
7	2.0500	0.3167	2.0100	0.4975	2.0200	0.5960	2.9388	0.5102
8	2.0000	0.3333	2.0000	0.5000	2.0100	0.5980	2.8980	0.5170
9	2.0000	0.3333	2.0000	0.5000	2.0000	0.6000	2.8367	0.5272
$OCG$	-	0.1944	-	0.3161	-	0.4013	-	0.3748

### 3.4.4 The trade-off between estimation performance and communication cost

In order to make the proposed iterative algorithm useful in practice, we introduce a stopping criterion to terminate the iterations. Since the sensor placements are known and the prior distribution of source location is available, the fusion center can compute the PCRLB of the source location estimate. Let  $tr\{\text{MSE}(N)\}$  be the trace of the MSE matrix when data from all the  $N$  sensors are assumed to be received and let  $tr\{\text{MSE}(K + iA)\}$  be the MSE after data from  $K + iA$  sensors are received. Then  $\mathcal{SM}(i)$  is defined as the stopping metric at iteration  $i$ , that is, the iterative algorithm terminates after the following criterion is met.

$$\mathcal{SM}(i) = \frac{tr\{\text{MSE}(K + iA)\} - tr\{\text{MSE}(N)\}}{tr\{\text{MSE}(N)\}} \leq \epsilon \quad (3.66)$$

where  $\epsilon$  is the desired accuracy.

#### Offline evaluation of stopping metric

The stopping metric (3.66) can be computed offline using the initial prior pdf ( $\mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ ). It can be used by the fusion center to coarsely determine how many and which non-anchor sensors should be selected to meet the stopping criterion in advance. Since PCRLB is a lower bound on the MSE and the MSE gets very close to its PCRLB for large sensor data,  $tr\{\text{MSE}(N)\}$  can be approximated by its PCRLB as  $tr\{\text{MSE}(N)\} \approx tr\{\text{PCRLB}(N)\}$ . At each iteration, similarly we assume that  $tr\{\text{MSE}(K + iA)\} \approx tr\{\text{PCRLB}(K + iA)\}$ . Given the prior distribution of the source location ( $\mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ ), appropriate selection of the number and locations of sensors in the network, the number and locations of the anchor sensors yield significant communication savings as compared to one-shot location estimation. As shown in Fig. 3.12,  $\mathcal{SM}(1)$  intersects the threshold at about 9. Therefore, 9 sensors should be selected to meet the stopping criterion at the first iteration.

## Online evaluation of stopping metric

We next evaluate the number of iterations and the communication cost by evaluating the stopping metric (3.66) online. We select  $A$  non-anchor sensors at each iteration based on the PCRLB-based sensor selection metric.  $\mathcal{SM}(i)$  for the selected sensors at iteration  $i$  is computed online using the iteratively refined prior pdf. To compute the MSE of all sensor data, we use the approximation  $tr\{\text{MSE}(N)\} \approx tr\{\text{PCRLB}(N)\}$ .  $tr\{\text{MSE}(K + iA)\}$  is approximated using the iteratively refined prior as,

$$tr\{\text{MSE}(K + iA)\} \approx tr\{\boldsymbol{\Sigma}_i\}$$

where  $\boldsymbol{\Sigma}_i$  is defined as,

$$\boldsymbol{\Sigma}_i = \sum_{m=1}^{N_s} w^{m,i} (\boldsymbol{\theta}^{m,i} - \boldsymbol{\mu}_i)(\boldsymbol{\theta}^{m,i} - \boldsymbol{\mu}_i)^T$$

and

$$\boldsymbol{\mu}_i = \sum_{m=1}^{N_s} w^{m,i} \boldsymbol{\theta}^{m,i}$$

Fig. 3.13-(a) shows the average number of iterations that is required for the stopping criterion (3.66) to be satisfied versus  $A$ . For  $A = 1$  and  $M = 5$ , the algorithm terminates in about 5 iterations which is consistent with Fig. 3.10 (c). According to offline computation of  $\mathcal{SM}(i)$ , 9 sensors need to be selected in order for the MSE to get very close to the PCRLB of  $N$  sensor data. Therefore, the offline computation of  $\mathcal{SM}(i)$  yields a loose estimate on the required number of iterations. The results presented in Fig. 3.13-(b) show the average number of bits used by the sensors for transmission until the end of the iterations by activating the non-anchor sensors based on iteratively updated posterior pdf of the source location and using distributed source coding as discussed in Section 3.3. As  $A$  increases, the algorithm terminates much faster, at the cost of increased total number of bits transmitted to the fusion center. As an example, for  $M = 5$  and  $A = 1$ , the algorithm terminates in about 5 iterations and on the average 20 bits are transmitted to the fusion center. For  $M = 5$  and  $A = 2$ , the algorithm converges in about 3-4 iterations and on the average 25 bits are transmitted to the fu-

sion center. For  $M = 6$ , the fusion center has much more information about the source location at each iteration as compared to the  $M = 5$  case, so for the  $M = 6$  case, the algorithm terminates faster as compared to the  $M = 5$  case. Note that when  $A$  is large the fusion center has to select a large number of sensors using coarse information at the first iteration. Together with the use of distributed source coding,  $A = 1$  yields the minimum number of bits transmitted to the fusion center until the end of the iterations.

### 3.5 Discussion

In this chapter, we have presented an iterative source localization method, where a coarse source location estimate is first obtained through the use of anchor sensors. Then, the posterior probability density function of the source location is approximated using a Monte Carlo method. We have developed and compared two different sensor selection schemes for static source localization. The first scheme iteratively activates those non anchor-sensors which maximize the mutual information between source location and the quantized sensor measurements. In the second sensor selection scheme, at each iteration a number of non-anchor sensors are activated whose quantized data minimize the PCRLB. Simulation results show that the MSE of the proposed iterative scheme gets close to the PCRLB for the case when all the sensor data are used, within a few iterations by selecting only the most informative sensors while significantly decreasing the communication requirements.

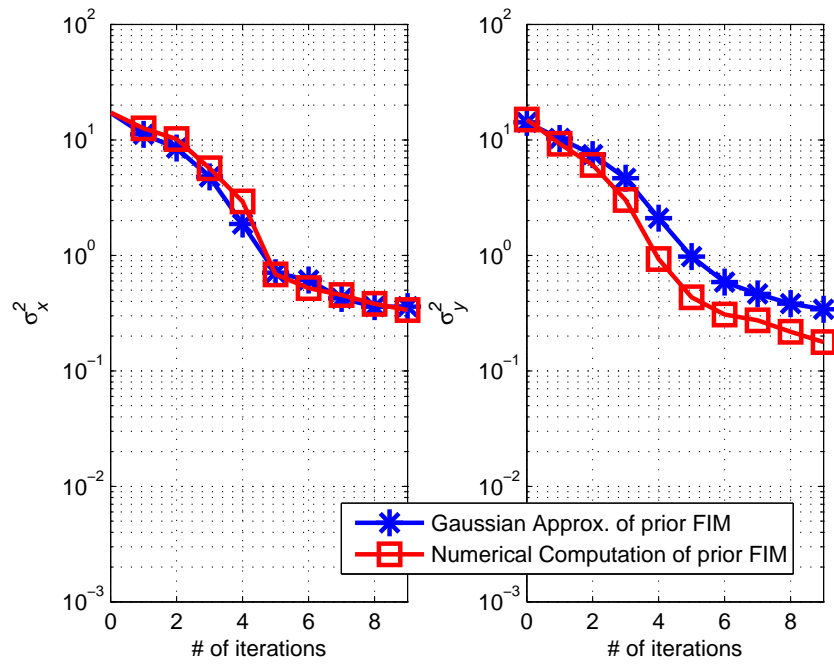
Simulation results show that the MI and PCRLB based sensor selection schemes achieve similar estimation performance and significantly outperform the selection of sensors which are nearest to the estimated source location when  $M$  is small. PCRLB based sensor selection is better in terms of computational complexity. It has been shown that the computational complexity of MI based sensor selection increases exponentially with the number of activated sensors per iteration; while the computational complexity of PCRLB based sensor selection increases linearly with the number of activated sensors per iteration.

The Monte Carlo-based posterior pdf of the source location is further used to compress the data of each activated sensor using distributed source coding techniques.

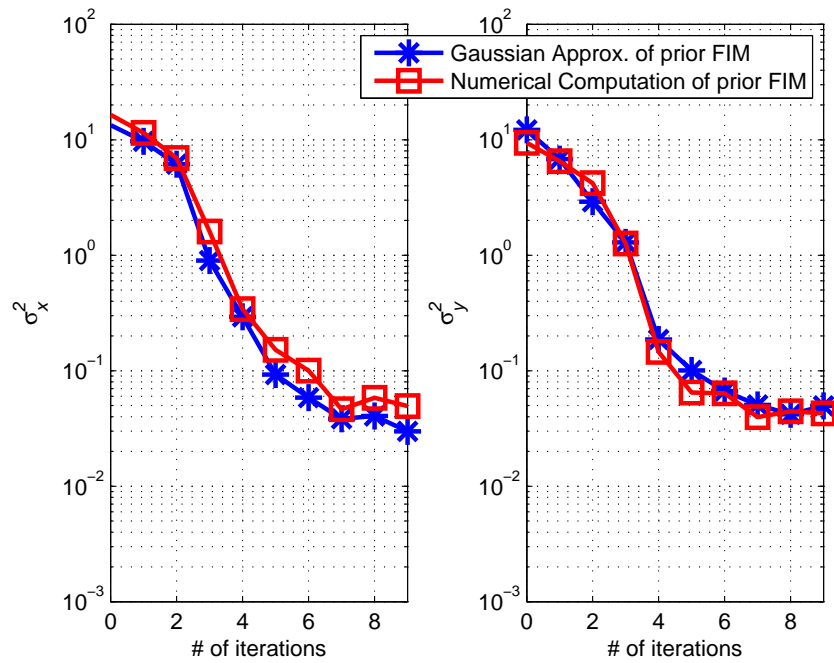
As the uncertainty about the source location decreases, the conditional entropy of each activated sensor becomes small and their  $M$ -bit data can be compressed significantly.

In this chapter, we considered that multi-bit sensor measurements are perfectly received at the fusion center. Next chapter will include channel fading and noise between sensors and the fusion center.

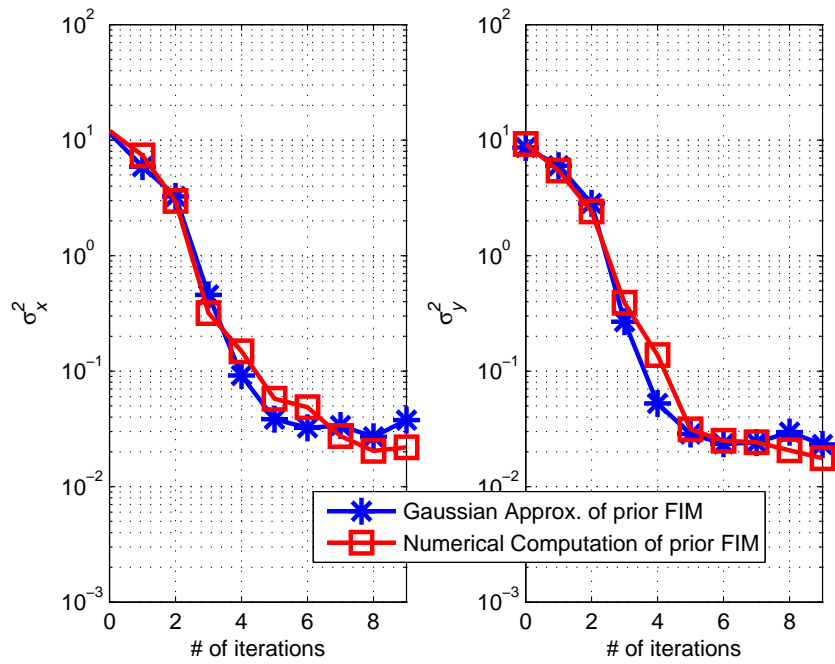




(a)

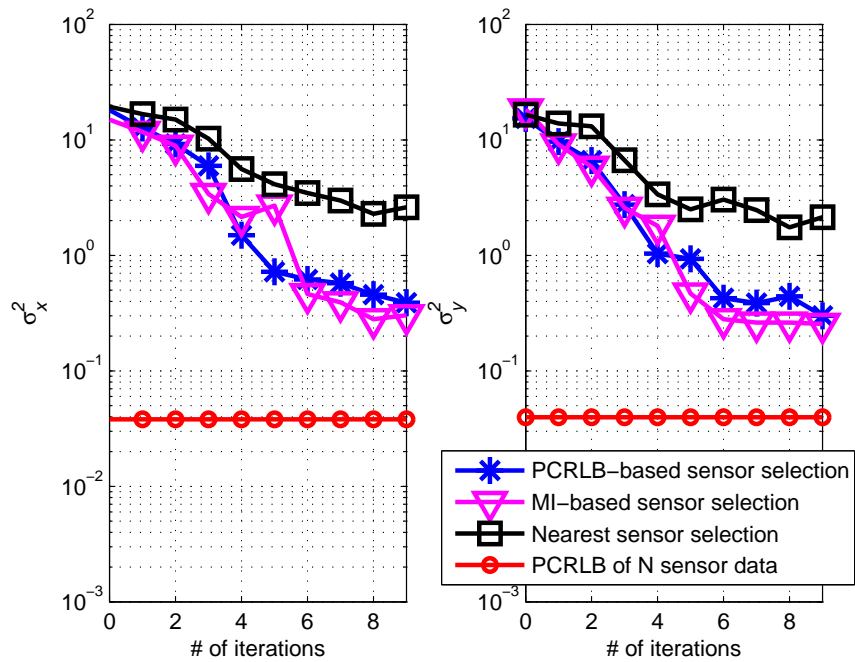


(b)

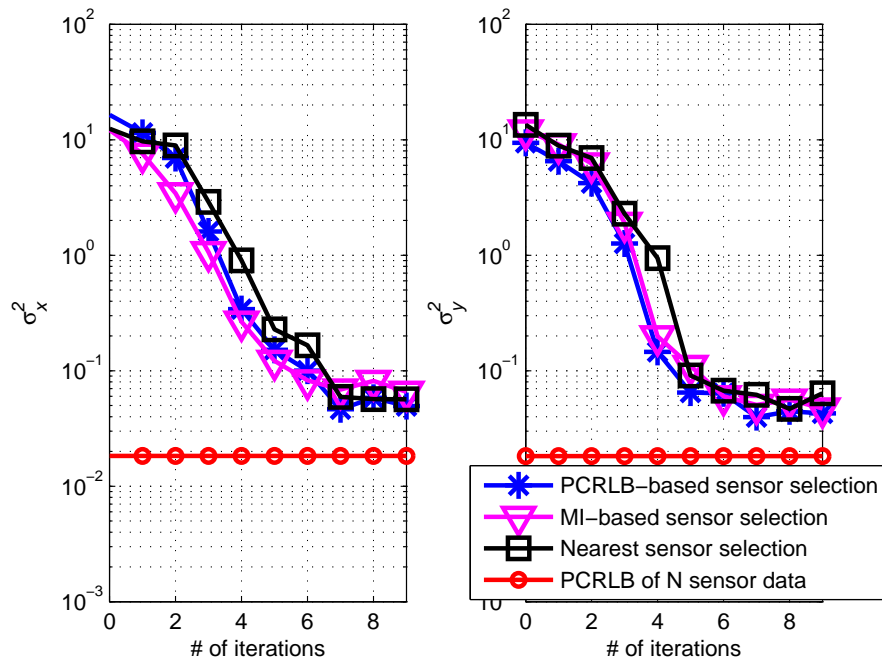


(c)

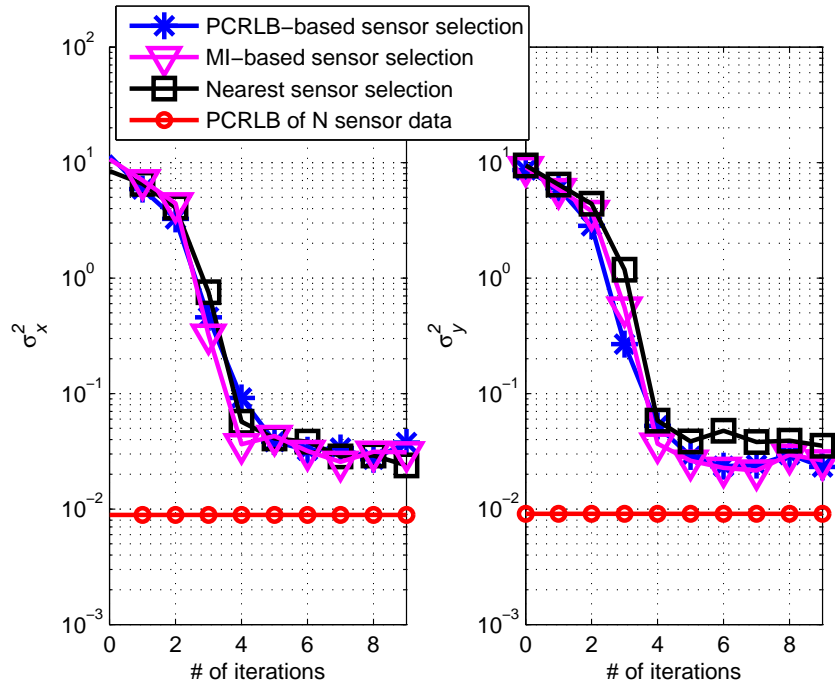
Figure 3.9: The MSE performance of the PCRLB based sensor selection. Comparison of Numerical computation and Gaussian approximation for the FIM of the prior. (a)  $M = 3$ , (b)  $M = 4$ , (c)  $M = 5$



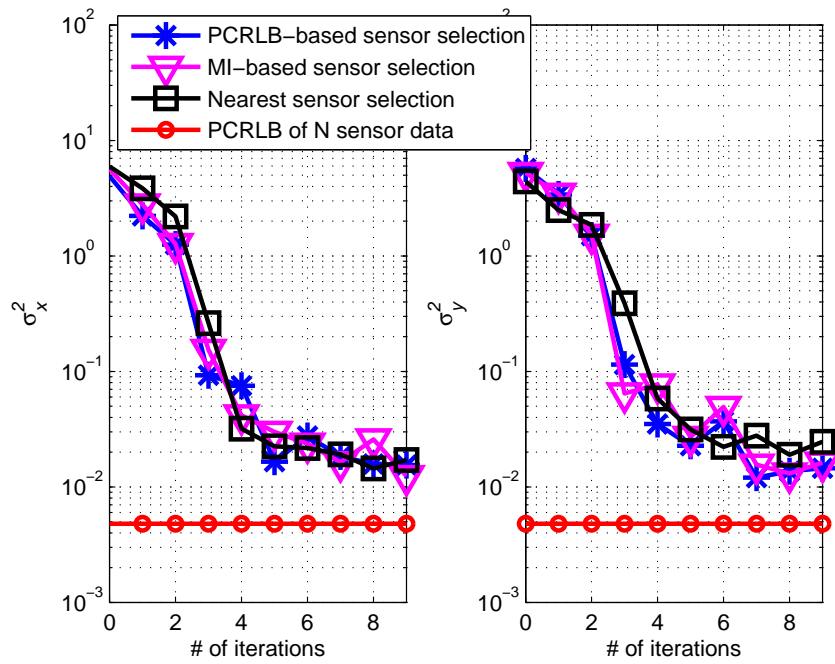
(a)



(b)

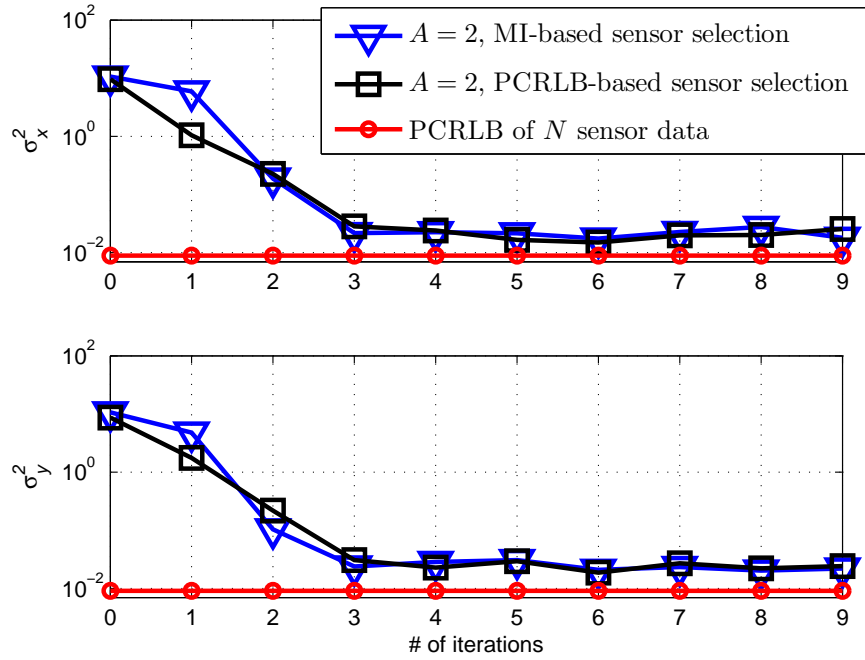


(c)

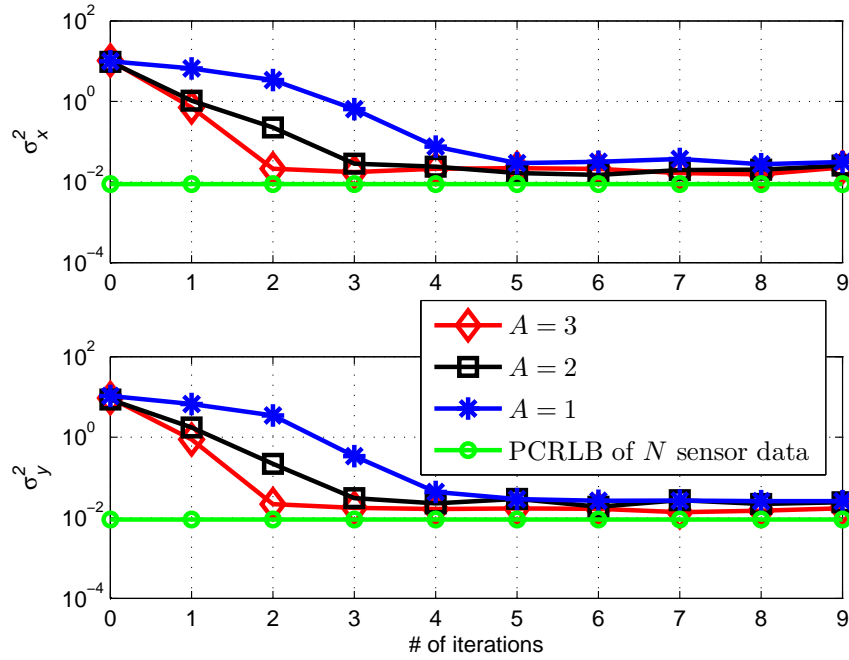


(d)

Figure 3.10: MSE at each iteration sensor selection is based on MI, PCRLB and nearest sensor to the estimated source location. (a)  $M = 3$ , (b)  $M = 4$ , (c)  $M = 5$ , (d)  $M = 6$  bits quantization



(a)



(b)

Figure 3.11: (a) MSE performance of MI and PCRLB based sensor selection schemes.  $N = 361$ ,  $K = 16$ ,  $M = 5$ ,  $A = 2$ ; (b)  $M = 5$  bit quantization of each sensor measurement, MSE performance of source localization with PCRLB based sensor selection and data compression.  $A = 1$ ,  $A = 2$  and  $A = 3$  sensor activations / iteration.

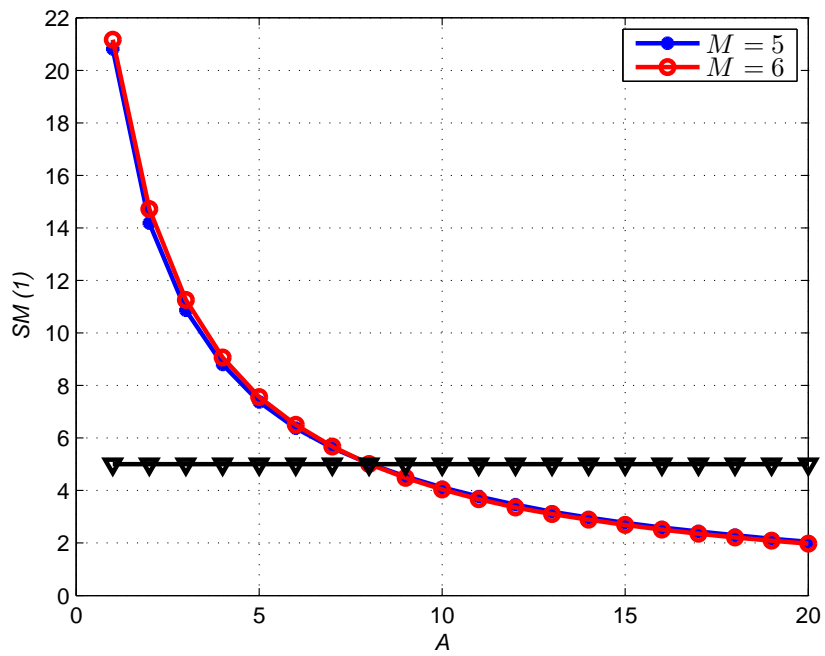
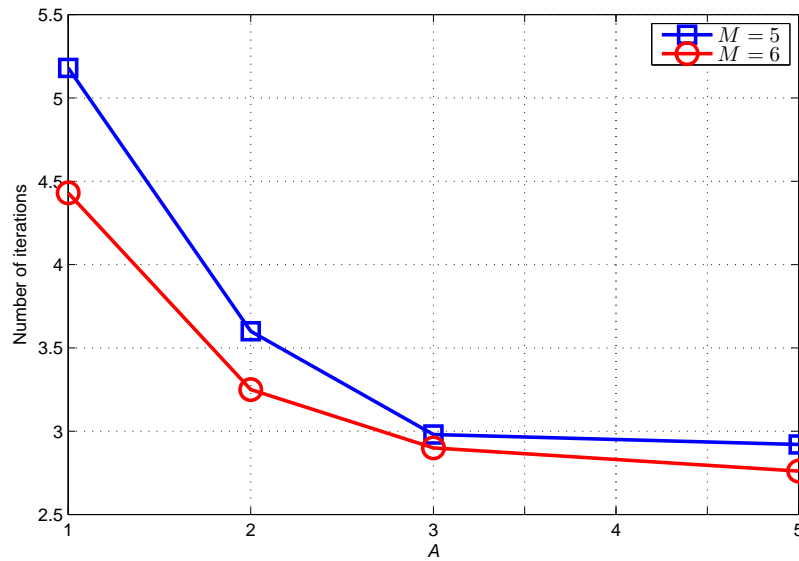
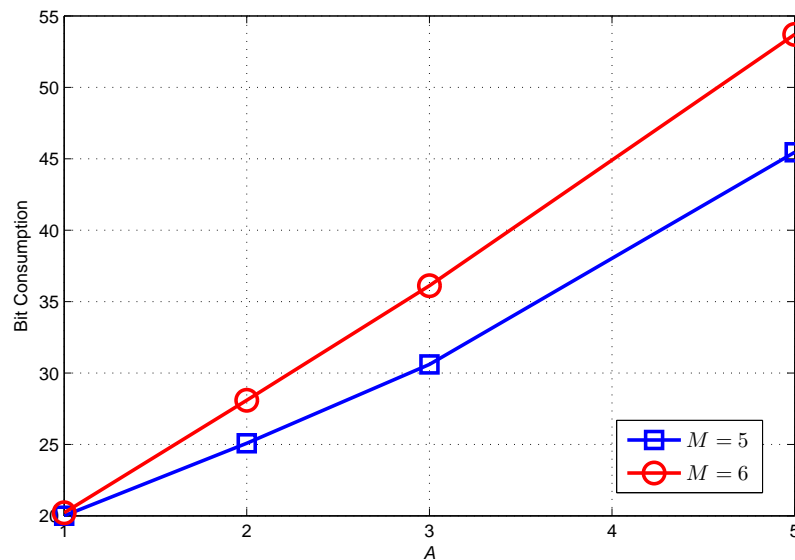


Figure 3.12: Stopping metric vs. the number of sensors to be selected. The black line with triangle markers indicates the accuracy threshold. ( $i = 1, \epsilon = 5$ )



(a)



(b)

Figure 3.13: (a) Average number of iterations until the termination of the algorithm. (b) Average number of bits transmitted to the fusion center until the termination of the algorithm. ( $M = 5$ ,  $M = 6$ ,  $\epsilon = 5$ , 100 different trials.)

## Chapter 4

# Channel Aware Iterative Source Localization

In the previous chapter, we have presented an iterative source localization method under the perfect communication channels assumption. Instead of requesting multi-bit decisions from all the sensors in the WSN, we first employed a small number of anchor sensors to obtain a coarse location estimate. Then, a few non-anchor sensors were activated at a time to refine the location estimate in an iterative manner. In this chapter, we extend our method for the case where the channels between sensors and the fusion center are subject to fading and noise. For the source localization problem, a related work [95] considers imperfect channels but their analysis is only limited to 1-bit transmission. In this work, we generalize the source location estimation approaches both given in [95] and [44] and consider  $M$ -bit sensor data. Assuming phase coherent reception, we consider channel impairments of two different types. In the first case, we assume that complete channel knowledge (CCK) of all the sensors is available at the fusion center in that the fusion center has the exact gain and phase information of each sensor. In the second case, we assume that partial channel knowledge (PCK) is available at the fusion center where only the statistics of the channel gain and phase information are known at the fusion center.

We derive the Posterior Cramer-Rao lower bound (PCRLB) of the Bayesian estimate. For sensor selection, we extend the metrics based on mutual information and PCRLB taking the channel effects into account. In the previous chapter, we have shown



that the PCRLB-based sensor selection metric is better than the MI-based sensor selection metric in terms of computational complexity which is still valid when imperfect channels are considered. Simulation results show that under low channel SNR, MI-based sensor selection gives higher priority to the sensors with large channel gain while PCRLB-based sensor selection starts selecting the sensors which are close to the source location. So the estimation performance of the MI-based sensor selection becomes better than the PCRLB-based sensor selection.

The rest of the chapter is organized as follows. In Section 4.1, we provide our system assumptions and derive the Cramér-Rao lower bound (CRLB) of the maximum likelihood based source location estimate for  $M$ -bit sensor measurements transmitted over fading channels. In Section 4.2, first we derive the PCRLB of the source location estimator, we then describe the Bayesian estimate of the source location using a Monte Carlo-based method. We then extend the mutual information based and PCRLB based sensor selection methods under fading channels. In Section 4.3, we give some numerical examples to illustrate the estimation performance and finally we devote Section 4.4 to to discussion of the results.

## 4.1 System Model

### 4.1.1 WSN assumptions

In this chapter, we make the same assumptions as in Section 3.1. Each local decision  $D_k$  is mapped to  $M$  bit binary sequence  $\mathbf{B}_k$  as [6] ,

$$\mathbf{B}_k = [b_{k,1}, \dots, b_{k,M}], (b_{k,j} \in \{0, 1\})$$

and transmitted using B-PSK modulation as

$$\mathbf{Q}_k = [q_{k,1}, \dots, q_{k,M}]$$

where

$$q_{k,j} = 2b_{k,j} - 1, j \in \{1, 2, \dots, M\} \quad k \in \{1, 2, \dots, N\}$$

We consider a discrete-time flat fading channel with a stationary and ergodic complex gain of  $h_k e^{j\phi_k}$  between sensor  $k$  and the fusion center where  $h_k$  is the gain of the channel and  $\phi_k$  is the phase of the channel. We assume that channel remains constant during the transmission of  $\mathbf{Q}_k$ . The received symbols from  $s_k$  have the form,

$$\tilde{r}_{k,j} = \sqrt{\epsilon_b} h_k e^{j\phi_k} q_{k,j} + \tilde{n}_{k,j} \quad (4.1)$$

where  $\epsilon_b$  is the bit energy and  $\tilde{n}_{k,j}$  is a zero-mean complex Gaussian noise with independent real and imaginary parts having identical variance. Then  $\tilde{n}_{k,j} \sim \mathcal{CN}(0, 2\sigma^2)$ . Let

$$\mathbf{R}_k = [r_{k,1}, r_{k,2}, \dots, r_{k,M}]$$

be the soft-decoding symbols received from  $s_k$  after phase coherent reception [4]. Then  $r_{k,j}$  has the form,

$$r_{k,j} = \sqrt{\epsilon_b} h_k q_{k,j} + n_{k,j} \quad (4.2)$$

where  $n_{k,j} \sim \mathcal{N}(0, \sigma^2)$  is independent and identically distributed at each symbol.

In this chapter we assume two different cases for the channel. The first case assumes complete channel knowledge (CCK) where both channel gain and phase information are available at the fusion center. The second case assumes partial channel knowledge (PCK) where the phase information and the probability distribution of channel gain are known at the fusion center.

Given source location, the likelihood  $p(\mathbf{R}|\boldsymbol{\theta})$  becomes,

$$p(\mathbf{R}|\boldsymbol{\theta}) = \prod_{k=1}^N p(\mathbf{R}_k|\boldsymbol{\theta}) \quad (4.3)$$

In the above equation,

$$p(\mathbf{R}_k|\boldsymbol{\theta}) = \sum_{l=0}^{L-1} p(\mathbf{R}_k|D_k = l)p(D_k = l|\boldsymbol{\theta}) \quad (4.4)$$

where

$$p(\mathbf{R}_k | D_k = l) = \prod_{j=1}^M p(r_{k,j} | q_{k,j})$$

$$p(D_k = l | \boldsymbol{\theta}) = Q\left(\frac{\eta_l - a_k}{\sigma}\right) - Q\left(\frac{\eta_{l+1} - a_k}{\sigma}\right)$$

### The likelihood of the received symbols under CCK

Under independent channel noise assumption, the vector of symbols received from each sensor is also independent. We first assume that complete channel knowledge is available at the receiver. Conditioning on channel gain  $h_k$ , yields  $p(\mathbf{R}_k | D_k = l, h_k) = p(\mathbf{R}_k | \mathbf{Q}_k, h_k)$  as,

$$p(\mathbf{R}_k | D_k, h_k) = \frac{1}{(2\pi)^{M/2} \sigma^M} \exp\left(-\sum_{j=1}^M \frac{(r_{k,j} - \sqrt{\epsilon_b} h_k q_{k,j})^2}{2\sigma^2}\right) \quad (4.5)$$

### The likelihood of the received symbols under PCK

We next incorporate imperfect channel statistics. Assuming a Rayleigh fading channel with unit power (i.e.,  $E[h_k^2] = 1$ ), the pdf of  $h_k$  is expressed as,

$$p(h_k) = 2h_k \exp(-h_k^2), \quad h_k \geq 0 \quad (4.6)$$

and the distribution of  $p(r_{k,j} | h_k, q_{k,j})$  becomes,

$$p(r_{k,j} | q_{k,j}, h_k) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(r_{k,j} - \sqrt{\epsilon_b} h_k q_{k,j})^2}{2\sigma^2}\right) \quad (4.7)$$

Now using (4.5), the likelihood of the received symbols at the fusion center is described as,

$$\begin{aligned}
p(\mathbf{R}_k|D_k) &= \int_{h_k} p(\mathbf{R}_k|\mathbf{Q}_k, h_k)p(h_k)dh_k \\
&= \int_{h_k} \left( \prod_{j=1}^M p(r_{k,j}|h_k, q_{k,j}) \right) p(h_k)dh_k \\
&= \int_0^\infty \frac{2h_k \exp(-h_k^2)}{(2\pi)^{M/2}\sigma^M} \exp\left(-\sum_{j=1}^M \frac{(r_{k,j} - \sqrt{\epsilon_b} h_k q_{k,j})^2}{2\sigma^2}\right) dh_k
\end{aligned} \tag{4.8}$$

Taking average with respect to channel gain  $h_k$ , we have Lemma 4.1.

**Lemma 4.1** *The conditional pdf of  $\mathbf{R}_k$ , given local decision  $D_k$  is*

$$\begin{aligned}
p(\mathbf{R}_k|D_k) &= \\
&\frac{2}{(2\pi)^{M/2}\sigma^{M-2} (2\sigma^2 + \epsilon_b M)} \exp\left(-\frac{\sum_{j=1}^M r_{k,j}^2}{2\sigma^2}\right) \times \\
&\left[ 1 + \sqrt{2\pi}\beta \left( \sum_{j=1}^M r_{k,j} q_{k,j} \right) \right. \\
&\exp\left(\frac{\beta^2 \left( \sum_{j=1}^M r_{k,j} q_{k,j} \right)^2}{2}\right) \\
&\left. Q\left(-\beta \left( \sum_{j=1}^M r_{k,j} q_{k,j} \right)\right) \right]
\end{aligned} \tag{4.9}$$

where

$$\beta = \frac{\sqrt{\epsilon_b}}{\sigma\sqrt{2\sigma^2 + \epsilon_b M}}$$

The detailed proof of Lemma 4.1 is presented in Appendix A.1. Note that Lemma 1 presented in [4] is a special case of (4.9) with  $M = 1$ .

#### 4.1.2 CRLB of the source location estimate

Let  $\hat{\boldsymbol{\theta}}(\mathbf{R})$  be an estimate of  $\boldsymbol{\theta} = [x \ y]^T$  based on available data  $\mathbf{R}$ . The maximum likelihood estimate (MLE) of the source location  $\hat{\boldsymbol{\theta}}^{MLE}$  is the solution to the following

optimization problem:

$$\hat{\boldsymbol{\theta}}^{MLE} = \hat{\boldsymbol{\theta}}(\mathbf{R}) = \arg \max_{\boldsymbol{\theta}} \log p(\mathbf{R}|\boldsymbol{\theta}) \quad (4.10)$$

For a constant but unknown  $\boldsymbol{\theta}$ , the Fisher information matrix (FIM)  $\mathbf{J}_{CRLB}(\boldsymbol{\theta})$  can be obtained as,

$$\mathbf{J}_{CRLB}(\boldsymbol{\theta}) = -E \begin{bmatrix} \nabla_x^x \log p(\mathbf{R}|\boldsymbol{\theta}) & \nabla_x^y \log p(\mathbf{R}|\boldsymbol{\theta}) \\ \nabla_y^x \log p(\mathbf{R}|\boldsymbol{\theta}) & \nabla_y^y \log p(\mathbf{R}|\boldsymbol{\theta}) \end{bmatrix} \quad (4.11)$$

It can be shown that  $\hat{\boldsymbol{\theta}}(\mathbf{R})$  is an unbiased estimator which satisfies the regularity condition. Then,

$$E\{[\hat{\boldsymbol{\theta}}(\mathbf{R}) - \boldsymbol{\theta}][\hat{\boldsymbol{\theta}}(\mathbf{R}) - \boldsymbol{\theta}]^T\} \geq \mathbf{J}_{CRLB}^{-1}(\boldsymbol{\theta}) \quad (4.12)$$

Due to the independence assumption given in (4.3),

$$\mathbf{J}_{CRLB}(\boldsymbol{\theta}) = \sum_{k=1}^N \mathbf{J}_k(\boldsymbol{\theta}) \quad (4.13)$$

$\mathbf{J}_k(\boldsymbol{\theta})$  is the FIM of  $\mathbf{R}_k$  and the expectation of the second-order derivative of  $\log p(\mathbf{R}_k|\boldsymbol{\theta})$  yields,

$$\begin{aligned} \mathbf{J}_k(\boldsymbol{\theta}) &= -E \left[ \frac{\partial^2 \log p(\mathbf{R}_k|\boldsymbol{\theta})}{\partial x^2} \right] = \\ &\int_{\mathbf{R}_k} \frac{1}{p(\mathbf{R}_k|\boldsymbol{\theta})} \left( \frac{\partial p(\mathbf{R}_k|\boldsymbol{\theta})}{\partial x} \right)^2 d\mathbf{R}_k \end{aligned} \quad (4.14)$$

where

$$\frac{\partial p(\mathbf{R}_k|\boldsymbol{\theta})}{\partial x} = \sum_{l=0}^{L-1} p(\mathbf{R}_k|D_k = l) \frac{\partial p(D_k = l|\boldsymbol{\theta})}{\partial x} \quad (4.15)$$

and

$$\frac{\partial p(D_k = l|\boldsymbol{\theta})}{\partial x} = \frac{na_k(x - x_k)}{2\sqrt{2}\sigma d_k^2} \left[ e^{-\frac{(\eta_l - a_k)^2}{2\sigma^2}} - e^{-\frac{(\eta_{l+1} - a_k)^2}{2\sigma^2}} \right] \quad (4.16)$$

If CCK is available,  $p(\mathbf{R}_k|D_k = l, h_k)$  is calculated according to (4.5) otherwise

$p(\mathbf{R}_k|D_k = l)$  is calculated according to (4.9). The other terms in (4.11) can be derived in a similar manner.

## 4.2 Iterative Source Location Estimation under Channel Fading

### 4.2.1 PCRLB of the source location estimate

In this section, we assume that the source location  $\boldsymbol{\theta}$  follows a prior pdf  $p_0(\boldsymbol{\theta})$  which is  $\mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$  where  $\boldsymbol{\mu}_0$  is the center of the ROI and  $\boldsymbol{\Sigma}_0$  is the covariance matrix which is very coarse so that its confidence region covers the whole ROI. Let  $p(\mathbf{R}, \boldsymbol{\theta})$  be the joint probability density of the pair of  $(\mathbf{R}, \boldsymbol{\theta})$ . Then, the PCRLB of the estimation error has the form [13],

$$E\{[\hat{\boldsymbol{\theta}}(\mathbf{R}) - \boldsymbol{\theta}][\hat{\boldsymbol{\theta}}(\mathbf{R}) - \boldsymbol{\theta}]^T\} \geq \mathbf{J}_{PCRLB}^{-1} \quad (4.17)$$

where  $\mathbf{J}$  is the  $2 \times 2$  Fisher information matrix (FIM)

$$\mathbf{J}_{PCRLB} = -E \begin{bmatrix} \nabla_x \log p(\mathbf{R}, \boldsymbol{\theta}) & \nabla_y \log p(\mathbf{R}, \boldsymbol{\theta}) \\ \nabla_x \log p(\mathbf{R}, \boldsymbol{\theta}) & \nabla_y \log p(\mathbf{R}, \boldsymbol{\theta}) \end{bmatrix} \quad (4.18)$$

Using the equality  $p(\mathbf{R}, \boldsymbol{\theta}) = p(\mathbf{R}|\boldsymbol{\theta})p_0(\boldsymbol{\theta})$ , the Fisher information matrix (FIM) can be written as,

$$\begin{aligned} \mathbf{J}_{PCRLB} &= E[-\nabla_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p(\mathbf{R}|\boldsymbol{\theta})] + E[-\nabla_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p_0(\boldsymbol{\theta})] \\ &= \mathbf{J}_d + \mathbf{J}_p \end{aligned} \quad (4.19)$$

In (4.19),  $\mathbf{J}_p \triangleq E[-\nabla_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p_0(\boldsymbol{\theta})] = \boldsymbol{\Sigma}_0^{-1}$  represents the *a priori* information, and  $\mathbf{J}_d \triangleq E[-\nabla_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p(\mathbf{R}|\boldsymbol{\theta})]$  is the standard FIM which has been presented in (4.13) averaged over the prior pdf of the source location as,

$$\mathbf{J}_d = E[-\nabla_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p(\mathbf{R}|\boldsymbol{\theta})] = \int_{\boldsymbol{\theta}} \mathbf{J}_{CRLB}(\boldsymbol{\theta}) p_0(\boldsymbol{\theta}) d\boldsymbol{\theta} \quad (4.20)$$

### 4.2.2 Source localization using a Sequential Monte-Carlo method

As presented in Chapter 3, at the beginning of the algorithm the fusion center gathers  $M$ -bit information from each of  $K$  anchor sensors and at each subsequent iteration of the algorithm, the fusion center gathers the  $M$ -bit data from additional  $A$  non-anchor sensors. Let  $p(\boldsymbol{\theta}|\mathbf{W}_i)$  be the posterior pdf of the source location given the available data  $\mathbf{W}_i = [\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_{K+iA}]$  at iteration  $i$  ( $i \geq 0$ ). Instead of using the importance sampling based method presented in the previous chapter, to speed up the computations, we approximate  $p(\boldsymbol{\theta}|\mathbf{W}_i)$  using a sequential importance sampling Monte-Carlo method as follows.

$$p(\boldsymbol{\theta}|\mathbf{W}_i) = \sum_{m=1}^{N_s} w^{m,i} \delta(\boldsymbol{\theta} - \boldsymbol{\theta}^{m,0}) \quad (4.21)$$

where  $\boldsymbol{\theta}^{m,0}$  represents a particle at iteration  $i$ , which are drawn from the prior distribution  $p_0(\boldsymbol{\theta})$  with statistics  $\mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$  and  $w^{m,i}$  represents the weight of a particle. Let  $\tilde{w}^{m,i}$  be the weight of particle  $\boldsymbol{\theta}^{m,0}$ . For the first iteration,  $\tilde{w}^{m,1}$  is obtained as,

$$\tilde{w}^{m,1} \propto \prod_{k=1}^K p(\mathbf{R}_k|\boldsymbol{\theta}^{m,0}) w^{m,0} \quad (4.22)$$

where  $\prod_{k=1}^K p(\mathbf{R}_k|\boldsymbol{\theta}^{m,0})$  is the likelihood of the  $K$  anchor sensor data at the end of the first iteration. For the rest of the iterations ( $i > 1$ ),

$$\tilde{w}^{m,i} \propto \prod_{k=1}^A p(\mathbf{R}_k|\boldsymbol{\theta}^{m,0}) \tilde{w}^{m,i-1} \quad (4.23)$$

Note that  $\prod_{k=1}^A p(\mathbf{R}_k|\boldsymbol{\theta}^{m,0})$  is the likelihood of the  $A$  non-anchor sensors activated at each iteration. For all the iterations, the particle weights are further normalized as,

$$w^{m,i} = \frac{\tilde{w}^{m,i}}{\sum_{m=1}^{N_s} \tilde{w}^{m,i}} \quad (4.24)$$

Then the Bayesian estimate (BE) of the source location is found from,

$$\hat{\boldsymbol{\theta}}_i^{BE} = \sum_{m=1}^{N_s} w^{m,i} \boldsymbol{\theta}^{m,0} \quad (4.25)$$

### 4.2.3 Mutual Information Based Sensor Selection under Channel Fading

Let  $\mathcal{R}^{(i,A)} = \{\mathcal{R}_1^{(i,A)}, \mathcal{R}_2^{(i,A)}, \dots, \mathcal{R}_{C_A^{N-K-iA}}^{(i,A)}\}$  be the collection of all distinct  $A$ -element subsets of  $N - K - iA$  remaining non-anchor sensors for the iteration  $i$  where  $C_A^{N-K-iA}$  is the combination operation.

Let  $\mathcal{R}_A^{(i,\nu)}$  be the set of  $A$  non-anchor sensors to be activated at the  $i^{th}$  iteration according to the sensor selection strategy  $\nu$ . Then,  $\mathcal{R}_A^{(i,\nu)} = \{\mathbf{R}_1^{i,\nu}, \mathbf{R}_2^{i,\nu}, \dots, \mathbf{R}_A^{i,\nu}\}$  and  $\mathbf{R}_k^{i,\nu}$  ( $k \in \{1, \dots, A\}$ ) are the received symbols from the  $k^{th}$  activated non-anchor sensor according to  $\nu$  at iteration  $i$ . Then,  $\mathbf{R}_k^{i,\nu} = [r_{k,1}^{i,\nu}, \dots, r_{k,M}^{i,\nu}]$ . Now, the objective is to find the optimal sensor selection strategy  $\nu^*$  which activates the set  $\mathcal{R}_{\nu^*}^{(i,A)} = \{\mathbf{R}_1^{i,\nu^*}, \dots, \mathbf{R}_A^{i,\nu^*}\}$  which activates  $A$  sensors out of  $N - K - iA$  remaining non-anchor sensors whose data minimize the expected conditional entropy of the posterior source location distribution as,

$$\nu^* = \arg \min_{\nu} H(\boldsymbol{\theta} | \mathcal{R}_A^{(i,\nu)}) \quad (4.26)$$

Let  $I(\boldsymbol{\theta}, \mathcal{R}_A^{(i,\nu)}) = H(\boldsymbol{\theta}) - H(\boldsymbol{\theta} | \mathcal{R}_A^{(i,\nu)})$  be the mutual information between the source location  $\boldsymbol{\theta}$  and the measurements of the activated sensors according to the activation scheme  $\nu$ . The sensor selection problem now turns into,

$$\nu^* = \arg \max_{\nu} I(\boldsymbol{\theta}, \mathcal{R}_A^{(i,\nu)}) \quad (4.27)$$

$I(\boldsymbol{\theta}, \mathcal{R}_A^{(i,\nu)})$  can also be expanded as,

$$I(\boldsymbol{\theta}, \mathcal{R}_A^{(i,\nu)}) = H(\mathcal{R}_A^{(i,\nu)}) - H(\mathcal{R}_A^{(i,\nu)} | \boldsymbol{\theta}) \quad (4.28)$$



To compute (4.28) using Monte-Carlo approximation, we start with writing the entropy of  $\mathcal{R}_{\mathbf{A}}^{(i,\nu)}$ ,

$$H(\mathcal{R}_{\mathbf{A}}^{(i,\nu)}) = - \int p(\mathcal{R}_{\mathbf{A}}^{(i,\nu)}) \log p(\mathcal{R}_{\mathbf{A}}^{(i,\nu)}) \quad (4.29)$$

where given source location  $\boldsymbol{\theta}$ ,  $p(\mathcal{R}_{\mathbf{A}}^{(i,\nu)})$  can be decomposed as,

$$p(\mathcal{R}_{\mathbf{A}}^{(i,\nu)}) = \int_{\boldsymbol{\theta}} p(\mathbf{R}_1^{i,\nu}|\boldsymbol{\theta}) \dots p(\mathbf{R}_A^{i,\nu}|\boldsymbol{\theta}) p_i(\boldsymbol{\theta}) d\boldsymbol{\theta} \quad (4.30)$$

Note that for iteration  $i$ ,  $p(\boldsymbol{\theta}|\mathbf{W}_{i-1})$  serves as the prior pdf of the source location  $p_i(\boldsymbol{\theta})$ . Then  $p_i(\boldsymbol{\theta}) = p(\boldsymbol{\theta}|\mathbf{W}_{i-1})$  which is obtained by the previously presented importance sampling-based Monte Carlo method and  $p(\mathbf{R}_1^{i,\nu}|\boldsymbol{\theta})$ , ...,  $p(\mathbf{R}_A^{i,\nu}|\boldsymbol{\theta})$  are the likelihood functions. Using (4.21), (4.30) becomes,

$$p(\mathcal{R}_{\mathbf{A}}^{(i,\nu)}) = \sum_{m=1}^{N_s} w^{m,i} p(\mathbf{R}_1^{i,\nu}|\boldsymbol{\theta}^{m,0}) \dots p(\mathbf{R}_A^{i,\nu}|\boldsymbol{\theta}^{m,0}) \quad (4.31)$$

With (4.31),  $H(\mathcal{R}_{\mathbf{A}}^{(i,\nu)})$  defined in (4.29) is rewritten as follows,

$$\begin{aligned} H(\mathcal{R}_{\mathbf{A}}^{(i,\nu)}) = & \quad (4.32) \\ & - \int_{\mathbf{R}_1^{i,\nu}} \dots \int_{\mathbf{R}_A^{i,\nu}} \left\{ \left[ \sum_{m=1}^{N_s} w^{m,i} \prod_{k=1}^A p(\mathbf{R}_k^{i,\nu}|\boldsymbol{\theta}^{m,0}) \right] \right. \\ & \left. \times \log \left[ \sum_{m=1}^{N_s} w^{m,i} \left( \prod_{k=1}^A p(\mathbf{R}_k^{i,\nu}|\boldsymbol{\theta}^{m,0}) \right) \right] \right\} d\mathbf{R}_1^{i,\nu} \dots d\mathbf{R}_A^{i,\nu} \end{aligned}$$

Now let us compute the second term of (4.28). First we have,

$$\begin{aligned} H(\mathcal{R}_{\mathbf{A}}^{(i,\nu)}|\boldsymbol{\theta}) = & \quad (4.33) \\ & - \int_{\boldsymbol{\theta}} \int_{\mathbf{R}_1^{i,\nu} \dots \mathbf{R}_A^{i,\nu}} p(\mathcal{R}_{\mathbf{A}}^{(i,\nu)}, \boldsymbol{\theta}) \log p(\mathcal{R}_{\mathbf{A}}^{(i,\nu)}|\boldsymbol{\theta}) \\ & d\mathbf{R}_1^{i,\nu} \dots d\mathbf{R}_A^{i,\nu} d\boldsymbol{\theta} \end{aligned}$$

Since  $p(\mathcal{R}_{\mathbf{A}}^{(i,\nu)}, \boldsymbol{\theta}) = p(\mathcal{R}_{\mathbf{A}}^{(i,\nu)}|\boldsymbol{\theta}) p_i(\boldsymbol{\theta})$  and considering the Monte-Carlo approximation of

the prior source location pdf (4.21), (4.33) is expressed as,

$$H(\mathcal{R}_{\mathbf{A}}^{(i,\nu)}|\boldsymbol{\theta}) = - \sum_{k=1}^A \int_{\mathbf{R}_k^{i,\nu}} \left[ \sum_{m=1}^{N_s} w^{m,i} p(\mathbf{R}_k^{i,\nu}|\boldsymbol{\theta}^{m,0}) \log(p(\mathbf{R}_k^{i,\nu}|\boldsymbol{\theta}^{m,0})) d\mathbf{R}_k^{i,\nu} \right] \quad (4.34)$$

Finally using (4.32) and (4.34), the mutual information function  $I(\boldsymbol{\theta}, \mathcal{R}_{\mathbf{A}}^{(i,\nu)})$  expressed in (4.28) is calculated as follows.

### MI-based sensor selection metric under CCK

Under perfect channel knowledge, MI-based sensor selection metric can be defined as,

$$I(\mathcal{R}_{\mathbf{A}}^{(i,\nu)}, \boldsymbol{\theta}) = \quad (4.35)$$

$$- \int_{\mathbf{R}_1^{i,\nu}} \dots \int_{\mathbf{R}_A^{i,\nu}} \left[ \sum_{m=1}^{N_s} w^{m,i} \left( \prod_{k=1}^A p(\mathbf{R}_k^{i,\nu} | h_k, \boldsymbol{\theta}^{m,0}) \right) \right]$$

$$\log \left[ \sum_{m=1}^{N_s} w^{m,i} \left( \prod_{k=1}^A p(\mathbf{R}_k^{i,\nu} | h_k, \boldsymbol{\theta}^{m,0}) \right) \right] d\mathbf{R}_1^{i,\nu} \dots d\mathbf{R}_A^{i,\nu}$$

$$+ \sum_{k=1}^A \int_{\mathbf{R}_k^{i,\nu}} \left[ \sum_{m=1}^{N_s} w^{m,i} p(\mathbf{R}_k^{i,\nu} | h_k, \boldsymbol{\theta}^{m,0}) \right]$$

$$\log(p(\mathbf{R}_k^{i,\nu} | h_k, \boldsymbol{\theta}^{m,0})) d\mathbf{R}_k^{i,\nu} \left. \right]$$

### MI-based sensor selection metric under PCK

Using fading statistics of the channel, MI-based sensor selection metric can be defined as,

$$\begin{aligned}
I(\mathcal{R}_{\mathbf{A}}^{(i,\nu)}, \boldsymbol{\theta}) = & \tag{4.36} \\
& - \int_{\mathbf{R}_1^{i,\nu}} \dots \int_{\mathbf{R}_A^{i,\nu}} \left[ \sum_{m=1}^{N_s} w^{m,i} \left( \prod_{k=1}^A p(\mathbf{R}_k^{i,\nu} | \boldsymbol{\theta}^{m,0}) \right) \right] \\
& \log \left[ \sum_{m=1}^{N_s} w^{m,i} \left( \prod_{k=1}^A p(\mathbf{R}_k^{i,\nu} | \boldsymbol{\theta}^{m,0}) \right) \right] d\mathbf{R}_1^{i,\nu} \dots d\mathbf{R}_A^{i,\nu} \\
& + \sum_{k=1}^A \int_{\mathbf{R}_k^{i,\nu}} \left[ \sum_{m=1}^{N_s} w^{m,i} p(\mathbf{R}_k^{i,\nu} | \boldsymbol{\theta}^{m,0}) \right. \\
& \left. \log(p(\mathbf{R}_k^{i,\nu} | \boldsymbol{\theta}^{m,0})) d\mathbf{R}_k^{i,\nu} \right]
\end{aligned}$$

#### 4.2.4 PCRLB Based Sensor Selection under Channel Fading

After initialization via the use of  $K$  anchor sensors, during each iteration the fusion center requests data from  $A$  non-anchor sensors that minimize the PCRLB. At iteration  $i$ , given available data  $\mathbf{W}_{i-1}$ , the PCRLB of  $A$  non-anchor sensors is expressed as,

$$E\{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T | \mathbf{W}_{i-1}\} \geq F^{-1}(\mathcal{R}_{\mathbf{A}}^{(i,\nu)} | \mathbf{W}_{i-1}) \tag{4.37}$$

where  $\mathbf{J}_c(\nu) = F(\mathbf{D}_{\mathbf{A}}^{(i,\nu)} | \mathbf{W}_{i-1})$  is the Fisher information matrix (FIM) of the random variable  $\boldsymbol{\theta}$  contained in  $\mathbf{D}_{\mathbf{A}}^{(i,\nu)}$  given available data  $\mathbf{W}_{i-1}$ . Then  $\mathbf{J}_c(\nu)$  is expressed as,

$$\begin{aligned}
\mathbf{J}_c(\nu) = & F(\mathcal{R}_{\mathbf{A}}^{(i,\nu)} | \mathbf{W}_{i-1}) \tag{4.38} \\
\triangleq & E \left\{ \left[ -\nabla_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p(\boldsymbol{\theta}, \mathcal{R}_{\mathbf{A}}^{(i,\nu)} | \mathbf{W}_{i-1}) \right] | \mathbf{W}_{i-1} \right\} \\
= & - \int_{\boldsymbol{\theta}} \int_{\mathbf{R}_1^{i,\nu} \dots \mathbf{R}_A^{i,\nu}} \left[ \nabla_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p(\boldsymbol{\theta}, \mathcal{R}_{\mathbf{A}}^{(i,\nu)} | \mathbf{W}_{i-1}) \right] \\
& p(\boldsymbol{\theta}, \mathcal{R}_{\mathbf{A}}^{(i,\nu)} | \mathbf{W}_{i-1}) d\mathbf{R}_1^{i,\nu} \dots d\mathbf{R}_A^{i,\nu} d\boldsymbol{\theta}
\end{aligned}$$

where we take expectation over all possible source locations  $\boldsymbol{\theta}$  and all vector of received symbols from all activated sensors  $\{\mathbf{R}_1^{i,\nu}, \mathbf{R}_2^{i,\nu}, \dots, \mathbf{R}_A^{i,\nu}\}$ .

Using Bayesian decomposition, the joint probability density function  $p(\boldsymbol{\theta}, \mathcal{R}_{\mathbf{A}}^{(i,\nu)} | \mathbf{W}_{i-1})$  of source location  $\boldsymbol{\theta}$  and new quantized measurements  $\mathcal{R}_{\mathbf{A}}^{(i,\nu)}$  is written as,

$$p(\boldsymbol{\theta}, \mathcal{R}_{\mathbf{A}}^{(i,\nu)} | \mathbf{W}_{i-1}) = p(\mathcal{R}_{\mathbf{A}}^{(i,\nu)} | \boldsymbol{\theta}) p(\boldsymbol{\theta} | \mathbf{W}_{i-1}) \quad (4.39)$$

where the identity  $p(\mathcal{R}_{\mathbf{A}}^{(i,\nu)} | \mathbf{W}_{i-1}, \boldsymbol{\theta}) = p(\mathcal{R}_{\mathbf{A}}^{(i,\nu)} | \boldsymbol{\theta})$  has been used. Using the above decomposition, (4.38) can be written as,

$$\begin{aligned} \mathbf{J}_c(\nu) = & - \int_{\boldsymbol{\theta}} \int_{\mathbf{R}_1^{i,\nu} \dots \mathbf{R}_A^{i,\nu}} \nabla_{\boldsymbol{\theta}}^{\theta} \log \left\{ p(\mathcal{R}_{\mathbf{A}}^{(i,\nu)} | \boldsymbol{\theta}) p(\boldsymbol{\theta} | \mathbf{W}_{i-1}) \right\} \\ & p(\mathcal{R}_{\mathbf{A}}^{(i,\nu)} | \boldsymbol{\theta}) p(\boldsymbol{\theta} | \mathbf{W}_{i-1}) d\mathbf{R}_1^{i,\nu} \dots d\mathbf{R}_A^{i,\nu} d\boldsymbol{\theta} \end{aligned} \quad (4.40)$$

which can be decomposed as,

$$\begin{aligned} \mathbf{J}_c(\nu) = & - \left[ \int_{\boldsymbol{\theta}} \int_{\mathbf{R}_1^{i,\nu} \dots \mathbf{R}_A^{i,\nu}} \left\{ \nabla_{\boldsymbol{\theta}}^{\theta} \log p(\mathcal{R}_{\mathbf{A}}^{(i,\nu)} | \boldsymbol{\theta}) \right\} \right. \\ & p(\mathcal{R}_{\mathbf{A}}^{(i,\nu)} | \boldsymbol{\theta}) p(\boldsymbol{\theta} | \mathbf{W}_{i-1}) d\mathbf{R}_1^{i,\nu} \dots d\mathbf{R}_A^{i,\nu} d\boldsymbol{\theta} \\ & + \int_{\boldsymbol{\theta}} \left\{ \nabla_{\boldsymbol{\theta}}^{\theta} \log p(\boldsymbol{\theta} | \mathbf{W}_{i-1}) \right\} p(\boldsymbol{\theta} | \mathbf{W}_{i-1}) \\ & \left. \left\{ \int_{\mathbf{R}_1^{i,\nu} \dots \mathbf{R}_A^{i,\nu}} p(\mathcal{R}_{\mathbf{A}}^{(i,\nu)} | \boldsymbol{\theta}) \right\} d\boldsymbol{\theta} \right] \end{aligned} \quad (4.41)$$

Since  $p(\mathcal{R}_{\mathbf{A}}^{(i,\nu)} | \boldsymbol{\theta}) = \prod_{k=1}^A p(\mathbf{R}_k^{i,\nu} | \boldsymbol{\theta})$  and  $\int_{\mathbf{R}_1^{i,\nu} \dots \mathbf{R}_A^{i,\nu}} p(\mathcal{R}_{\mathbf{A}}^{(i,\nu)} | \boldsymbol{\theta}) = 1$ , (4.41) becomes,

$$\begin{aligned} \mathbf{J}_c(\nu) = & - \left[ \sum_{k=1}^A \int_{\boldsymbol{\theta}} \int_{\mathbf{R}_k^{i,\nu}} \left( \left\{ \nabla_{\boldsymbol{\theta}}^{\theta} \log p(\mathbf{R}_k^{(i,\nu)} | \boldsymbol{\theta}) \right\} \prod_{k=1}^A p(\mathbf{R}_k^{(i,\nu)} | \boldsymbol{\theta}) \right) \right. \\ & p(\boldsymbol{\theta} | \mathbf{W}_{i-1}) d\mathbf{R}_k^{i,\nu} d\boldsymbol{\theta} \end{aligned} \quad (4.42)$$

$$\left. + \int_{\boldsymbol{\theta}} \nabla_{\boldsymbol{\theta}}^{\theta} \log \{ p(\boldsymbol{\theta} | \mathbf{W}_{i-1}) \} p(\boldsymbol{\theta} | \mathbf{W}_{i-1}) d\boldsymbol{\theta} \right] \quad (4.43)$$

Finally (4.42) reduces to,

$$\begin{aligned} \mathbf{J}_c(\nu) = & - \left\{ \sum_{k=1}^A \int_{\boldsymbol{\theta}} \left[ \int_{\mathbf{R}_k^{i,\nu}} [\nabla_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p(\mathbf{R}_k^{(i,\nu)} | \boldsymbol{\theta})] \right. \right. \\ & \left. \left. p(\mathbf{R}_k^{(i,\nu)} | \boldsymbol{\theta}) d\mathbf{R}_k^{i,\nu} \right] p(\boldsymbol{\theta} | \mathbf{W}_{i-1}) d\boldsymbol{\theta} \right. \\ & \left. + \int_{\boldsymbol{\theta}} [\nabla_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p(\boldsymbol{\theta} | \mathbf{W}_{i-1})] p(\boldsymbol{\theta} | \mathbf{W}_{i-1}) d\boldsymbol{\theta} \right\} \end{aligned} \quad (4.44)$$

Note that the term in brackets  $[\cdot]$ ,

$$\int_{\mathbf{R}_k^{i,\nu}} [\nabla_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p(\mathbf{R}_k^{(i,\nu)} | \boldsymbol{\theta})] p(\mathbf{R}_k^{(i,\nu)} | \boldsymbol{\theta}) d\mathbf{R}_k^{i,\nu}$$

is the standard FIM and has been derived in (4.14). For the first term of (4.44), we use (4.21) to approximate  $p(\boldsymbol{\theta} | \mathbf{W}_{i-1})$ . The second term

$$\boldsymbol{\Psi}_i \triangleq \int_{\boldsymbol{\theta}} [\nabla_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p(\boldsymbol{\theta} | \mathbf{W}_{i-1})] p(\boldsymbol{\theta} | \mathbf{W}_{i-1}) d\boldsymbol{\theta}$$

can be computed using the approximations presented in Section 3.2.2.

### PCRLB based sensor selection metric under CCK

Under perfect channel knowledge, PCRLB-based sensor selection metric can be defined as,

$$\begin{aligned} \mathbf{J}_c(\nu) = & - \left\{ \sum_{k=1}^A \sum_{m=1}^{N_s} w^{m,i} \left[ \int_{\mathbf{R}_k^{i,\nu}} [\nabla_{\boldsymbol{\theta}}^{\boldsymbol{\theta}} \log p(\mathbf{R}_k^{(i,\nu)} | h_k, \boldsymbol{\theta}^{m,0})] \right. \right. \\ & \left. \left. p(\mathbf{R}_k^{(i,\nu)} | h_k, \boldsymbol{\theta}^{m,0}) d\mathbf{R}_k^{i,\nu} \right] + \boldsymbol{\Psi}_i \right\} \end{aligned} \quad (4.45)$$

## PCRLB Based Sensor Selection metric under PCK

Using channel fading statistics, PCRLB-based sensor selection metric can be defined as,

$$\mathbf{J}_c(\nu) = - \left\{ \sum_{k=1}^A \sum_{m=1}^{N_s} w^{m,i} \left[ \int_{\mathbf{R}_k^{i,\nu}} [\nabla_{\boldsymbol{\theta}} \log p(\mathbf{R}_k^{(i,\nu)} | \boldsymbol{\theta}^{m,0})] p(\mathbf{R}_k^{(i,\nu)} | \boldsymbol{\theta}^{m,0}) d\mathbf{R}_k^{i,\nu} \right] + \Psi_i \right\} \quad (4.46)$$

The fusion center gathers the soft-decoded  $M$ -bit sensor data from  $A$  non-anchor sensors which collectively have the maximum mutual information with the source location of maximum Fisher information. Once the selected sensors' data are received at the fusion center, the iterative source location algorithm continues with updating new weights  $w^{m,i+1}$ .

### 4.3 Simulation Results

In our examples, we consider the source energy and signal decay exponent as  $P_0 = 2500$  and  $n = 2$  respectively.  $N$  sensors are deployed in a  $20 \times 20m^2$  field in a grid topology and we use  $M = 3$  bits to quantize analog measurements. The decision thresholds of each sensor  $\boldsymbol{\eta}$  are selected according to the method described in [44]. We consider two different scenarios which are the low channel signal-to-noise ratio (SNR) and high channel SNR cases with  $\epsilon_b = 1$  and  $\epsilon_b = 5$  respectively. The mean squared error (MSE) matrix of the estimation algorithm is calculated according to,

$$\text{MSE} = \frac{1}{Z} \sum_{z=1}^Z (\hat{\boldsymbol{\theta}}_z - \boldsymbol{\theta}_z)(\hat{\boldsymbol{\theta}}_z - \boldsymbol{\theta}_z)^T \quad (4.47)$$

The integrations with respect to  $\mathbf{R}_k$  are performed by Monte Carlo integration [96]. The parameters of the prior probability distribution of the source location  $p_0(\cdot)$  are assumed to be  $\boldsymbol{\mu}_0 = [10 \text{ m. } 10 \text{ m.}]^T$  and  $\boldsymbol{\Sigma}_0 = \begin{bmatrix} \sigma_{x,0}^2 & 0 \\ 0 & \sigma_{y,0}^2 \end{bmatrix}$  is the covariance matrix

which is very coarse so that its 99% confidence region covers the whole ROI.

### 4.3.1 Performance of the one-shot location estimator

In Fig. 4.1, we find the Bayesian estimate (BE) of the source location using all sensor data where we tested our algorithm over  $Z = 1000$  different source locations. We plot the trace of the MSE matrix denoted as  $\text{trace}(\text{MSE})$ . We then compare the MSE performance of the BE with its PCRLB bound. Under low SNR, the MSE performance of PCK is near optimal as compared to the CCK case. As the bit energy increases, the MSE performances of CCK and PCK cases are almost indistinguishable. Moreover under high SNR and large number of sensors, the MSE performance of PCK case gets very close to its PCRLB bound.

In Table 4.1 and Table 4.2, we compare the estimation performance of the one-shot location estimator by varying  $M$  and  $N$  under PCK.  $N$  sensors are deployed in a grid topology and each sensor decision is assumed to be received at the fusion center under PCK. We calculate the estimation improvement under low channel SNR and high channel SNR where  $\epsilon_b = 1$  and  $\epsilon_b = 5$  respectively. Simulation results show that as  $N$  and  $M$  increase, the estimation performance improves. For the same number of sensors, as  $M$  increases, the estimation improvement of  $\epsilon_b = 5$  is larger than the estimation improvement of  $\epsilon_b = 1$ . This is due to the fact that as  $\epsilon_b$  increases, the SNR of the channel increases, the destructive effects of the channel are suppressed and sensor decisions become more informative about the source location.

Table 4.1: Estimation Improvement by increasing  $M$ ,  $\epsilon_b = 1$

$N$	Inter-sensor distance (m)	$\text{tr}(\text{MSE})$ $M = 1$	$\text{tr}(\text{MSE})$ $M = 5$	Improvement %
4	20	6.2861	5.5878	11.11
9	10	6.0088	4.1821	30.40
16	6.7	5.9362	2.0032	66.25
25	5	5.7104	1.1875	79.21
36	4	5.6142	0.5677	89.88
49	3.3	5.2849	0.3614	93.16

Table 4.2: Estimation Improvement by increasing  $M$ ,  $\epsilon_b = 5$

$N$	Inter-sensor distance (m)	$tr(\text{MSE})$ $M = 1$	$tr(\text{MSE})$ $M = 5$	Improvement %
4	20	6.0291	5.1080	15.28
9	10	5.5504	2.3538	57.59
16	6.7	5.2218	0.5709	89.07
25	5	4.9000	0.2749	94.39
36	4	4.4020	0.1557	96.46
49	3.3	4.002	0.1061	97.35

### 4.3.2 Performance of the iterative location estimation

For the proposed iterative source localization algorithm, we employ  $N = 7 \times 7 = 49$  sensors deployed in a grid topology. The algorithm is initialized with  $K = 3 \times 3 = 9$  anchor sensors. The MSE of BE at each iteration is averaged over  $Z = 100$  different trials. In our simulations, we activate  $A = 1$  sensor at a time after the initialization via anchor sensors.

As shown in Fig. 4.2, for MI-based sensor selection and for low channel SNR, CCK yields better performance than PCK as a result of using the complete channel information. Increasing the bit energy yields similar estimation performance for both CCK and PCK. For high channel SNR, iterative source localization converges to the MSE of all sensor data in about 20 iterations that is when the sensors close to source location have been selected. On the other hand, for low channel SNR, the sensors become less informative, hence about 30 sensors need to be selected to achieve the MSE of all sensor data. The mean channel gain and the mean distance to the source location of sensors selected at each iteration are shown in Fig. 4.3. Under low channel SNR, MI-based sensor selection activates the sensors with large channel gains where the sensors are within a radius of 7 meters of the source location. As the channel SNR increases, the channel gain becomes less important and sensors in a closer proximity of the source location are selected.

For PCRLB based sensor selection, we first evaluate  $\Psi_i = \Gamma_i$  according to (3.50), that is the numerical approach to calculate the FIM of the prior. Fig. 4.4 shows  $\Gamma_1(1, 1)$



and  $\Gamma_1(2, 2)$  as a function of  $\delta$ . Under channel fading, simulation results show us it is very hard to conclude a stable region for  $\delta$  which yields consistent  $\Gamma$  values.

Instead, we set  $\Psi_i = \Sigma_i^{-1}$  and approximate the posterior pdf with a Gaussian distribution. In Figure 4.5, we compare MI and PCRLB based sensor selection for source localization under CCK. Note that the MI-based sensor selection metric (4.35) depends on  $p(\mathbf{R}_k|\boldsymbol{\theta})$  which further depends on channel gain  $h_k$  and received signal strength  $a_k(d_k)$  which is a function of the distance ( $d_k$ ) between  $s_k$  and source location ( $\boldsymbol{\theta}$ ). On the other hand, the PCRLB-based sensor selection metric (4.45) depends not only on  $p(\mathbf{R}_k|\boldsymbol{\theta})$  but also its first order derivative. As shown in (4.16), the first order derivative of  $p(\mathbf{R}_k|\boldsymbol{\theta})$  not only depends on  $h_k$  and  $a_k(d_k)$ , but also depends directly on  $d_k^2$ . Since the PCRLB-based method, selects the sensors maximizing the Fisher information, a sensor with large gain but having  $d_k > 1$  would not likely be selected since its Fisher information is inversely related with  $d_k^4$ .

According to Fig. 4.6. Under low SNR, MI-based sensor selection gives priority to the sensors which have large channel gain where the sensors are selected within a radius of 7 meters of the source location. PCRLB-based sensor selection gives priority to the sensors which are close to the source location. For the first few iterations, the PCRLB-based method selects the sensors within a radius of 2 – 3 meters around the source location. Once the sensors close to the source location are selected, PCRLB-based sensor selection then selects the sensors according to their channel gain. Simulation results show that giving priority to the sensors with large channel gain yields better MSE performance as compared to giving priority to the sensors close to the source location. Therefore, the MSE performance of MI-based sensor selection is better than PCRLB-based sensor selection. Also note that, we can compute the MI-based sensor selection metric exactly. On the other hand, PCRLB-based sensor selection metric is computed by using the approximation (3.52). Moreover, the PCRLB metric is a lower bound on the MSE, due to the uncertainties (low SNR, channel fading), the MSE can highly deviate from its PCRLB. Even though MI-based sensor selection yields better estimation performance, it is computationally more expensive than the PCRLB-based sensor selection scheme.

## 4.4 Discussion

In this chapter, we have studied the iterative source localization problem where the multi-bit data of sensors have been transmitted over fading channels. We have first approximated the PCRLB of the Bayesian estimate of the source location using a Monte Carlo method. We have shown that having partial channel knowledge provides estimation performance very close to the case where complete channel information is available. We have also shown that, for high SNR and large number of sensors, the MSE gets very close to its PCRLB bound. For iterative source localization, we have derived the MI and PCRLB based sensor selection schemes and have compared their estimation performance. Simulation results show that under CCK, MI-based sensor selection performs better than PCRLB-based sensor selection under low SNR since MI-based sensor selection gives higher priority to the sensors with large channel gain. As SNR increases, the PCK assumption yields a similar iterative estimation performance as with the CCK assumption.

The performance of the PCRLB based sensor selection scheme can be further improved in a future work. In order to ease the calculations, at each iteration, we have assumed that the particle weights obey Gaussian distribution. Better models can be developed to obtain the Fisher information of the prior distribution. Extension of our methodology for a non-coherent reception employed at the fusion center for multi-bit sensor data can also be addressed.

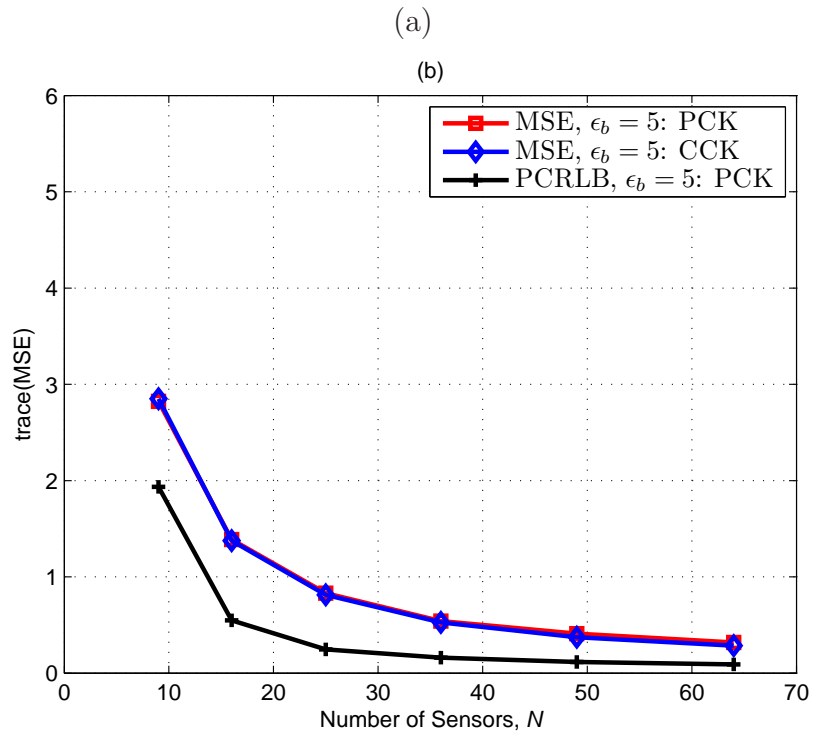
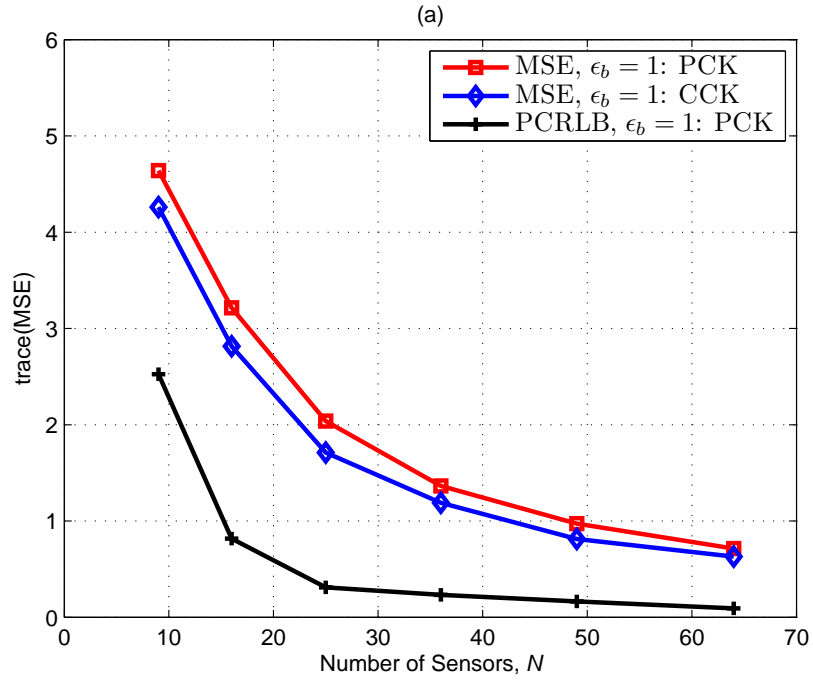
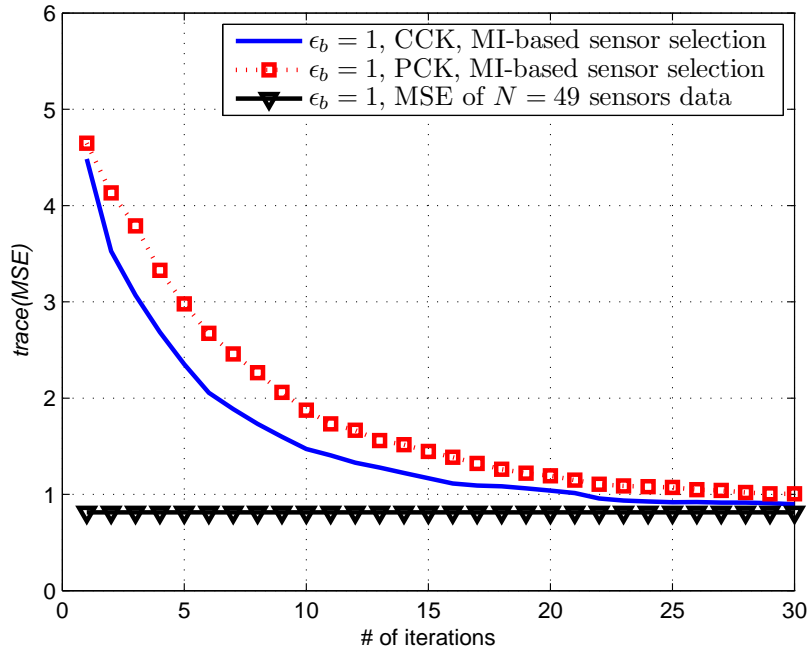
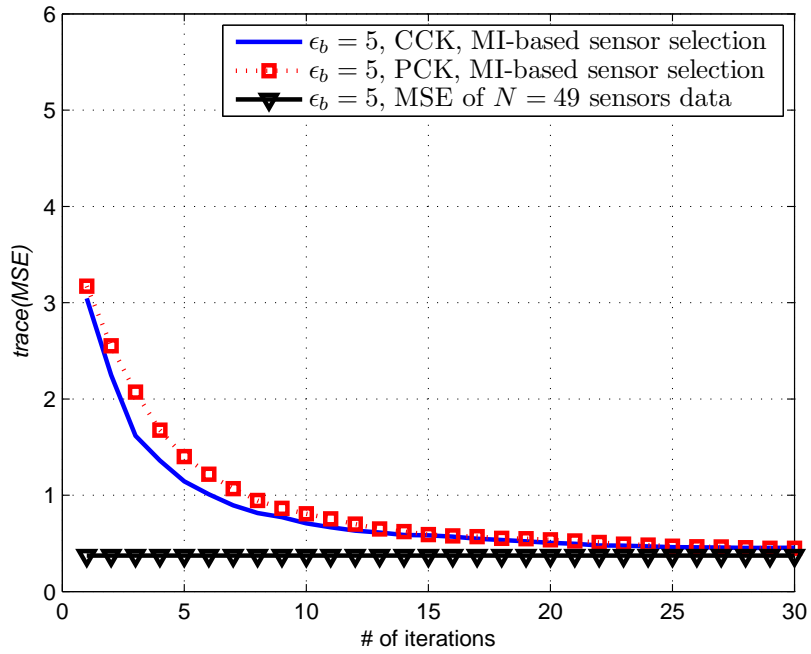


Figure 4.1: Trace of the MSE matrix of  $N$  sensor data, (a)  $\epsilon_b = 1$ , (b)  $\epsilon_b = 5$  ( $M = 3$ )



(a)



(b)

Figure 4.2: MSE performance of the iterative scheme using MI-based sensor selection (a)  $\epsilon_c = 1$  (b)  $\epsilon_c = 5$ , (The dashed line is the trace of the MSE matrix of  $N = 49$  sensor data.)

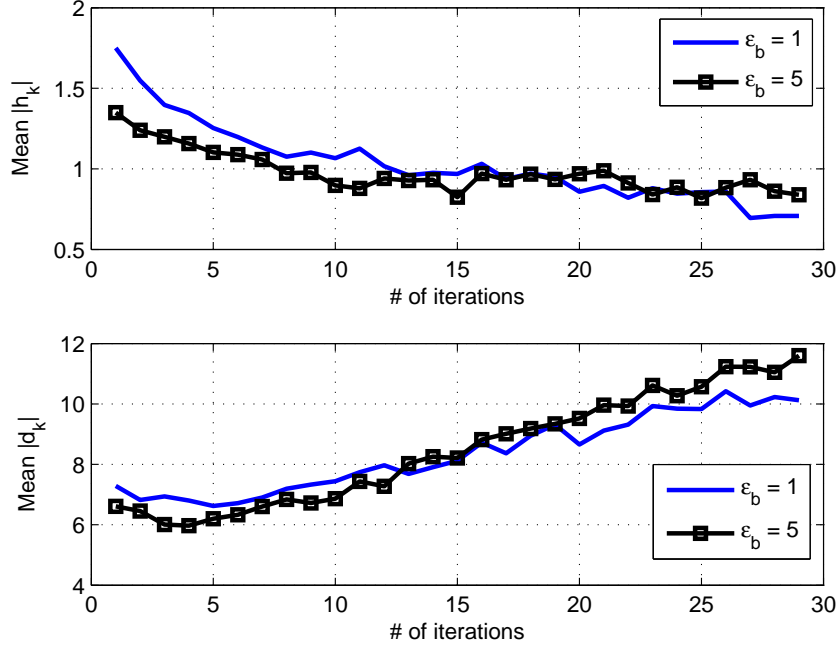


Figure 4.3: Mean channel gain  $|h_k|$ ; mean distance between the source location and the selected sensor  $|d_k|$  at each iteration (MI-based sensor selection).

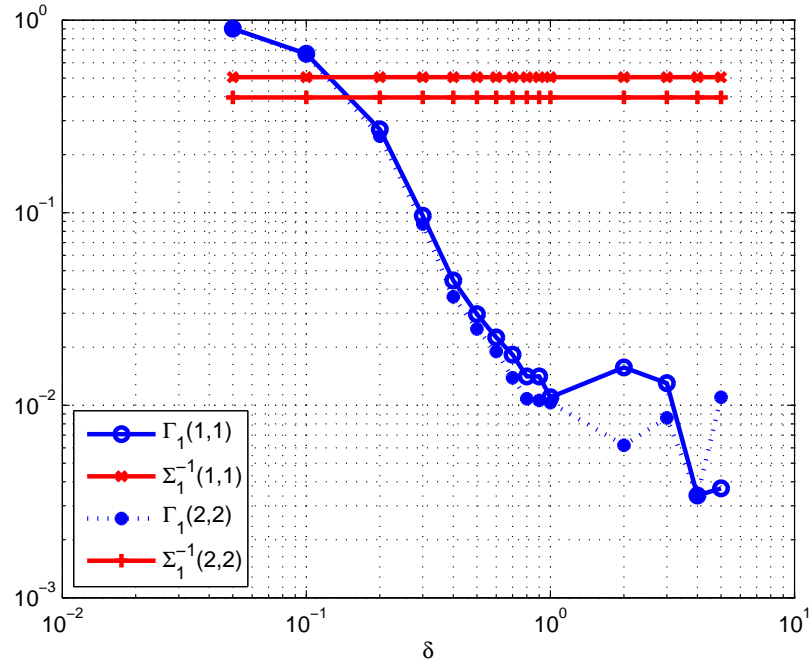
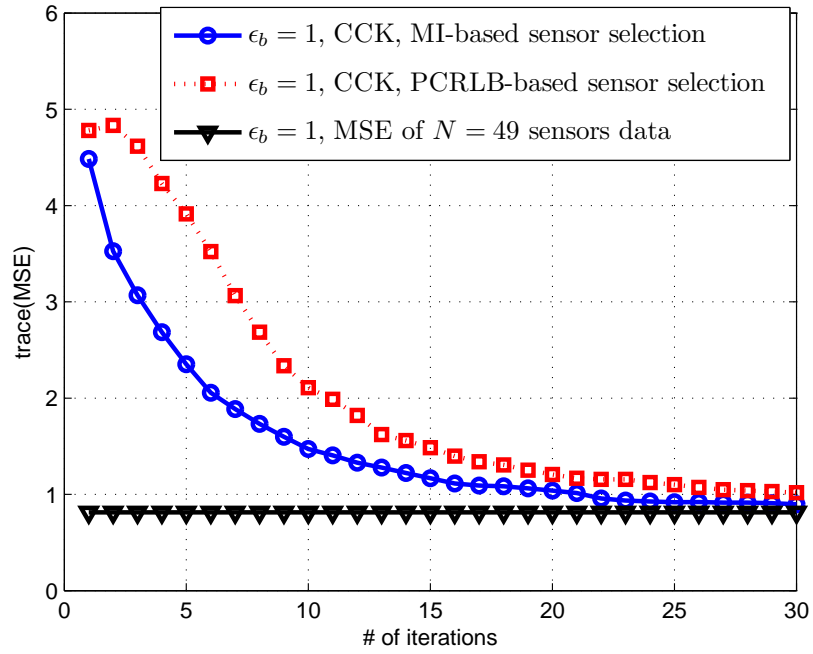
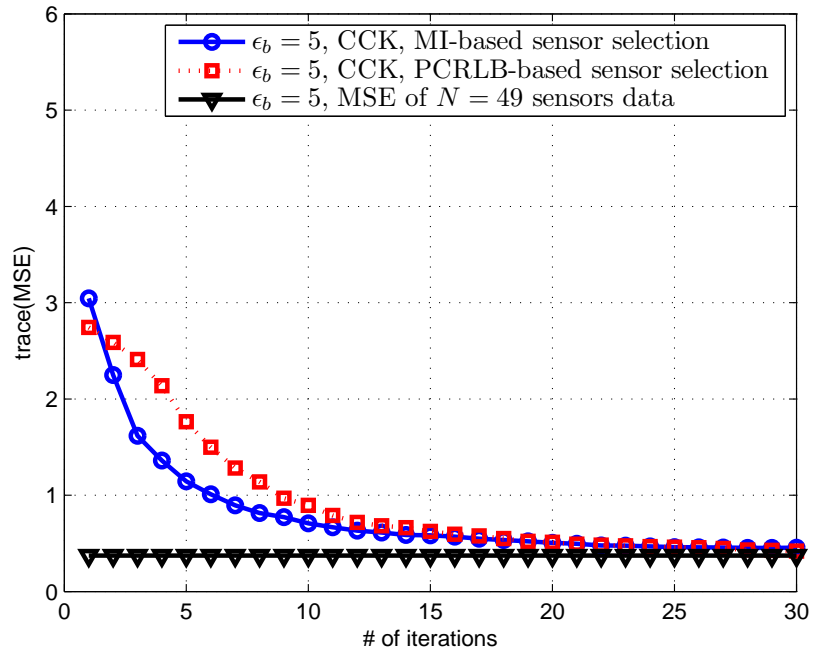


Figure 4.4: Evaluation of the prior FIM as a function of  $\delta$ , ( $M = 3$ , ROI =  $20 \times 20$ )

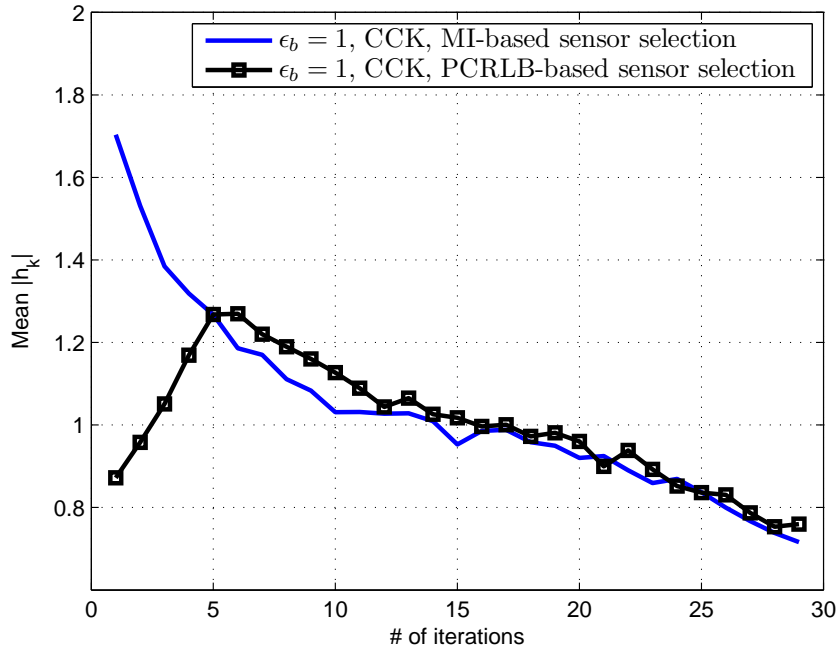


(a)

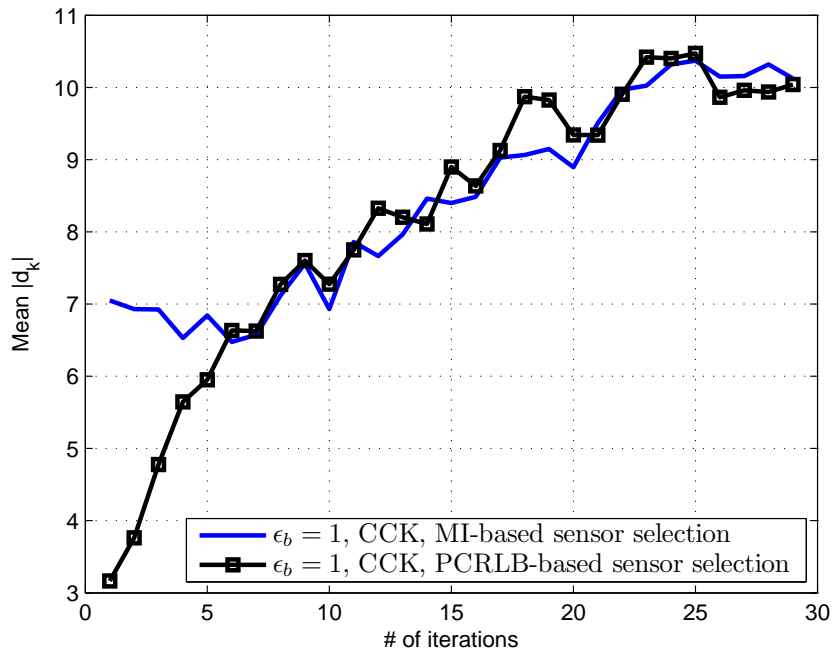


(b)

Figure 4.5: MSE performance of the iterative scheme using MI and PCRLB-based sensor selection (a)  $\epsilon_b = 1$  (b)  $\epsilon_b = 5$ , (The dashed line is the trace of the MSE matrix of  $N = 49$  sensor data.)



(a)



(b)

Figure 4.6: (a) Mean channel gain and (b) Mean distance to the source location of the sensor selected at each iteration (MI and PCRLB based sensor selection).

## Chapter 5

# A Probabilistic Rate Transmission Scheme for Distributed Estimation

In the previous chapters, we considered a homogenous WSN where the sensors experience independent and identically distributed observation noise and all the sensors send the same amount of data to the fusion center. In this chapter, we consider a heterogeneous WSN where the sensors experience independent but not identical observation noise and each sensor transmits data to the fusion center as a function of its observation SNR. We consider a distributed parameter estimation problem in a WSN where the sensors observe a common parameter  $\theta$  and the fusion center estimates  $\theta$  based on the quantized data transmitted from the sensors. Note that the framework presented in this chapter is suitable for other estimation problems such as the source localization problem discussed in the previous chapter. Here, we consider a fully distributed estimation scheme where the fusion center does not know the individual noise characteristics of each sensor and can not employ a dynamic rate allocation scheme based on the instantaneous conditions of each sensor as in [50], [56]. Different from [51], we assume that the dynamic range of the parameter to be estimated is not bounded but the estimation parameter follows a certain prior probability distribution function. As an example, a control system may be responsible to fix the temperature of an indoor facility like an office, greenhouse or a cold storage. The instantaneous temperature may be recorded in time and for instance the indoor temperature may obey a pdf such as a Gaussian distribution around the mean desired temperature. Using the statistics of the parameter, the WSN can be



accordingly deployed and designed in a more intelligent way to perform its task.

In a heterogenous WSN where all the sensors employ different quantization data rates, as we will show later in the chapter, the complexity to compute the average PCRLB is high. Therefore, we first introduce the inverse of the average Fisher information as a lower bound on the average PCRLB. We then assume that a sensor is able to measure its observation signal-to-noise ratio and its quantized measurement is transmitted at a data rate that maximizes the Fisher information per bit. In other words, a sensor with a higher observation SNR, quantizes its observation with a higher data rate. Recall from Chapter 4, the destructive effects of the channel can be suppressed by increasing the transmission energy per bit. Then, the multi-bit sensor decisions become more informative, which leads to a better estimation performance. In this chapter, we assume sensors transmit their data in multi-bits and we neglect the channel impairments in the estimation process. We assume that the data transmission between sensors and the fusion center is error free, which can be provided by using suitable error correction codes or sufficient transmit power at each sensor. At the same time, we consider that the available bandwidth of the channel is limited, which means the channel can reliably deliver at most  $B$  bits from sensors to the fusion center. If the constraint on available bandwidth is stringent, the transmission of all sensor data to the fusion center may result an outage. To alleviate this problem, a transmission probability is assigned to each particular quantization data rate. In other words, a sensor whose observation is quantized with rate  $R_k$  bits, sends its data to the fusion center with probability  $\rho_{R_k}$ . Our problem is then to find the optimal transmission probabilities corresponding to each particular quantization data rate so as to minimize the inverse of average Fisher information subject to total bandwidth and bandwidth utilization constraints.

Given the bandwidth constraint, a simple way to perform distributed estimation is to share the available bandwidth ( $B$ ) equally among all the sensors in the network ( $N$ ). We refer to this scheme as *Equal Rate Transmission* (ERT). As an example using time division multiple access (TDMA), ERT can be implemented in a Round-Robin fashion where each sensor sends  $B/N$  bits to the fusion center. The *Probabilistic Rate Transmission* (PRT) scheme can be implemented as follows. The fusion center may

first send a beacon signal, where sensors synchronize with the fusion center and the fusion center communicates the transmission probability of each quantization data rate  $\rho_{R_k}$  to all the sensors. Then a sensor whose observation is quantized with rate  $R_k$ , reports back to the fusion center with probability  $\rho_{R_k}$  and announcing that its data will be transmitted with rate  $R_k$  bits. The sensors that report back to the fusion center are then scheduled for data transmission at the fusion center. As an example, in a TDMA manner, a sensor with  $R_k$  bit data may be assigned  $R_k$  time slots to deliver its observation. Simulation results show that, PRT outperforms ERT significantly in terms of MSE. The optimal transmission probabilities are assigned in such a way that the sensors with higher observation SNR have priority to transmit their data, and the resulting mean squared estimation error is quite close to the case where all the sensors transmit at the maximum quantization data rate.

The rest of the chapter is organized as follows. In Section 5.1, we provide the WSN assumptions and describe the rate decision to be made at each sensor. In Section 5.2, we define the PCRLB of the estimate and describe the probabilistic rate transmission scheme. In Section 5.3, we introduce the Bayesian parameter estimator of the received sensor data and in Section 5.4, we present some numerical examples. Finally, we devote Section 5.5 to concluding remarks and discussion on future applications.

## 5.1 Problem Formulation

The WSN is composed of  $N$  distributed sensors and a fusion center for estimating the parameter  $\theta$ . We assume that the parameter  $\theta$  is a random variable which is generated from a Gaussian distribution  $p_0(\theta) \sim \mathcal{N}(\mu_\theta, \sigma_\theta^2)$ . Each sensor  $s_k$  receives  $z_k$  which is a noisy version of  $\theta$ ,

$$z_k = \theta + n_k \tag{5.1}$$

We consider a heterogeneous WSN where the noise of each sensor  $n_k$  follows a Gaussian distribution  $\mathcal{N}(0, \sigma_{n,k}^2)$ . Further, we assume that each sensor's noise variance  $\tau_k \triangleq \sigma_{n,k}^2$

follows a  $\chi^2$ -distribution with  $a = 1$  degree of freedom as modeled in [50].

$$p(\tau_k) = \frac{(\tau_k)^{\frac{a}{2}-1} e^{-(\tau_k/2)}}{2^{a/2} \Gamma(a/2)} \quad (5.2)$$

where  $\Gamma(\cdot)$  denotes the Gamma function. The observation signal-to-noise ratio (SNR) of each sensor is defined as,

$$\text{SNR}_k = 10 \log_{10} \frac{E[\theta^2]}{E[n_k^2]} = 10 \log_{10} \left( \frac{\mu_\theta^2 + \sigma_\theta^2}{\tau_k} \right) \text{ dB} \quad (5.3)$$

Let  $D_k$  be the  $R_k = j$ -bit ( $j \in \{1, 2, \dots, M\}$ ) quantized measurement of  $z_k$ ,  $M$  be the maximum data rate at which a sensor can communicate with the fusion center, and  $L_j = 2^j$  be the number of quantization levels. The set of quantization thresholds corresponding to rate  $R_k = j$  is represented as  $\boldsymbol{\eta}_j = [\eta_{j,0} \ \eta_{j,1} \ \dots \ \eta_{j,L_j}]^T$ . Using  $\boldsymbol{\eta}_j$ , we obtain the quantized data  $D_k$  from the analog measurement  $z_k$  according to,

$$D_k = \begin{cases} 0 & -\infty < z_k < \eta_{j,1} \\ 1 & \eta_{j,1} < z_k < \eta_{j,2} \\ \vdots & \\ L_j - 1 & \eta_{j,L_j-1} < z_k < \infty \end{cases} \quad (5.4)$$

where  $\eta_{j,0} = -\infty$  and  $\eta_{j,L_j} = \infty$ .

### 5.1.1 Data rate decision of each sensor

We consider a one-shot estimation problem. Let the vector of sensor data rates be  $\mathbf{R} = [R_1 \ R_2 \ \dots \ R_N]^T$  and the collected data from all the  $N$  sensors be  $\mathbf{D} = [D_1 \ D_2 \ \dots \ D_N]^T$ . Given  $\mathbf{R}$ , the PCRLB has the form,

$$E \left\{ \left[ \hat{\theta}(\mathbf{D}) - \theta \right] \left[ \hat{\theta}(\mathbf{D}) - \theta \right]^T \middle| \mathbf{R} \right\} \geq J^{-1} \quad (5.5)$$

where  $J$  is the Fisher information (FI) defined as,

$$\begin{aligned} J &= E \left[ -\nabla_{\theta}^{\theta} \log p(\mathbf{D}, \theta) \mid \mathbf{R} \right] \\ &= \sum_{k=1}^N E \left[ -\nabla_{\theta}^{\theta} \log p(D_k | \theta, R_k) \right] + E \left[ -\nabla_{\theta}^{\theta} \log p_0(\theta) \right] \end{aligned} \quad (5.6)$$

The first term in the second line of (5.6) is valid as long as  $\tau_k$  are independent across the sensors. The second term is the Fisher information of the Gaussian distribution which is equal to,

$$E \left[ -\nabla_{\theta}^{\theta} \log p_0(\theta) \right] = \frac{1}{\sigma_{\theta}^2}$$

Other prior pdf's for the parameter to be estimated can be used. As an example, the Fisher information of the Weibull distribution is shown in [97].

Let  $J_{d,k}(R_k = j) = E \left[ -\nabla_{\theta}^{\theta} \log p(D_k | \theta, R_k = j) \right]$  be the standard FI of a single sensor averaged over the prior distribution  $p_0(\theta)$  when  $R_k = j$  bits are used to communicate with the fusion center. Note that (5.6) is maximized by maximizing  $J_{d,k}(R_k = j)$ . Fig. 5.1 shows  $J_{d,k}(R_k = j)$  of a sensor as a function of  $j$  for various  $\text{SNR}_k$ . We have assumed that the sensors can measure  $\text{SNR}_k$  and hence can determine  $\tau_k$ . The optimal quantization thresholds maximize  $J_{d,k}(R_k = j | \tau_k)$  according to,

$$\max_{\boldsymbol{\eta}_j} J_{d,k}(R_k = j | \tau_k) \quad (5.7)$$

subject to the constraint  $\eta_{j,1}(\tau_k) \leq \dots \leq \eta_{j,L-1}(\tau_k)$ .

In Fig. 5.1,  $J_{d,k}(R_k = j | \tau_k)$  increases with the data rate and finally converges to the FI corresponding to the analog measurements. As an example, for  $\text{SNR}_k = 0 \text{ dB}$  (or  $\text{SNR}_k = 30 \text{ dB}$ ),  $R_k = 3$  (or  $R_k = 6$ ) bits is sufficient to achieve the FI of the analog measurement. Note that, assigning more than  $R_k = 3$  (or  $R_k = 6$ ) bits is unnecessary because increasing  $R_k$  does not improve  $J_{d,k}$  significantly. Since a sensor knows its  $\text{SNR}_k$ , it is capable of calculating the  $J_{d,k}(R_k = j | \tau_k)$  for different  $R_k = j$ . Then the data rate  $R_k = j^*$  bits is selected which has the maximum Fisher information per bit

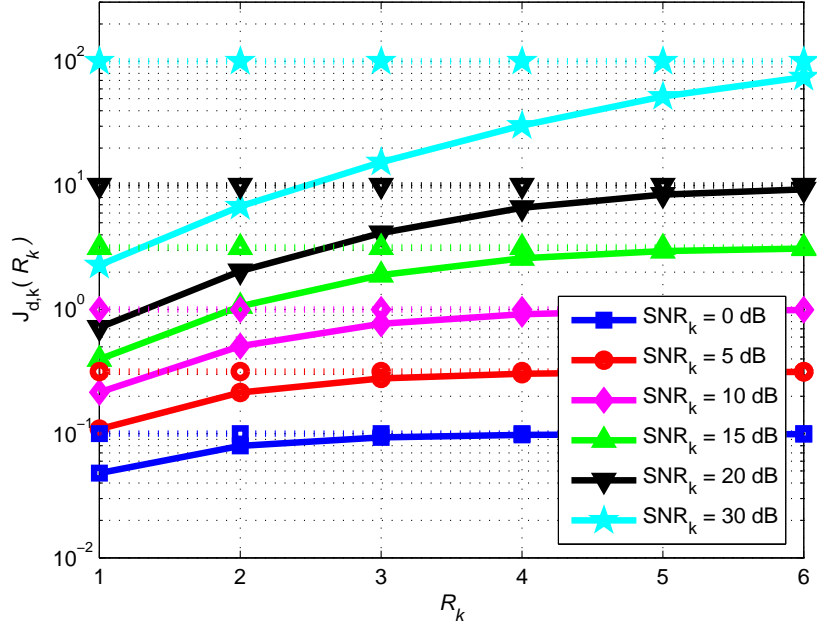


Figure 5.1: Fisher information of a single sensor at various  $SNR_k$ . Dashed lines represent the FI for the case where a sensor transmits analog data and  $M = 6$ .

which is defined as,

$$j^* = \arg \max_j \left( \frac{J_{d,k}(R_k = j | \tau_k)}{j} \right) \quad (5.8)$$

For different  $SNR_k$ , (or  $\tau_k$ ), the optimal rate  $j^*$  is presented in Fig. 5.2 by using the criterion (5.8). Then,  $P(R_k = j)$  is the probability of each sensor deciding on quantizing its data in  $R_k = j$  bits.  $P(R_k = j)$  is found from,

$$P(R_k = j) = F_{\chi^2}(\zeta_{M-j+1}) - F_{\chi^2}(\zeta_{M-j}) \quad (5.9)$$

where  $F_{\chi^2}(\cdot)$  is the cumulative distribution function of the  $\chi^2$  random variable with one degree of freedom. The vector of noise variance thresholds  $\zeta = [\zeta_0 \zeta_1 \dots \zeta_M]$  is used by the sensors to choose an appropriate communication rate. Each element of  $\zeta$  corresponds to the noise variance where the optimal transmission rate changes by one in Fig. 5.2. For this example, using  $M = 6$ ,  $\zeta$  is obtained as  $\zeta = [0, 0.06, 0.13, 0.38, 1.07, 2.95, \infty]$ . As an example, if  $\tau_k$  falls between 0.06 and 0.13, a sensor measurement is quantized with data rate  $R_k = 5$  bits.

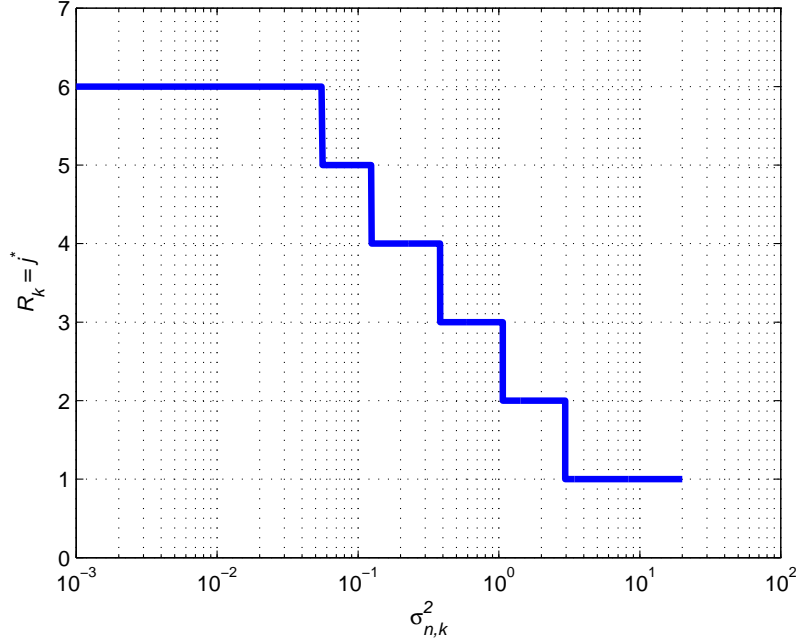


Figure 5.2: Optimal data rate which maximizes the FI per bit vs.  $\tau_k$ .  $M = 6$ ,  $\mu_\theta = 0$  and  $\sigma_\theta^2 = 10$ .

### 5.1.2 Determination of the quantization thresholds

The quantization thresholds corresponding to each data rate optimized based on  $\tau_k$  can not be obtained in practice, since we assume that  $\tau_k$ 's are not known at the fusion center. Instead, we assume that the vector of noise variance thresholds  $\zeta$  are known both at the sensors and the fusion center. We first find the mean noise variance ( $\bar{\tau}_j$ ) of the interval  $[\zeta_{M-j}, \zeta_{M-j+1}]$  where  $R_k = j$ ,

$$\bar{\tau}_j = \int_{\zeta_{M-j}}^{\zeta_{M-j+1}} \tau_k p(\tau_k | R_k = j) d\tau_k \quad (5.10)$$

where

$$p(\tau_k | R_k = j) = \begin{cases} p(\tau_k) / P(R_k = j) & \zeta_{M-j} \leq \tau_k < \zeta_{M-j+1} \\ 0 & \text{otherwise} \end{cases} \quad (5.11)$$

For  $\tau_k \in [\zeta_{M-j}, \zeta_{M-j+1}]$ , we approximate  $J_{d,k}(R_k = j|\tau_k)$  by  $J_{d,k}(R_k = j|\bar{\tau}_j)$ . Then, we obtain the set of optimal quantization thresholds  $\boldsymbol{\eta}_j^*$  for data rate  $R_k = j$  based on  $\bar{\tau}_j$ . Then, the objective function presented in (5.7) is replaced with,

$$\boldsymbol{\eta}_j^* = \arg \max_{\boldsymbol{\eta}_j} J_{d,k}(R_k = j|\bar{\tau}_j) \quad (5.12)$$

subject to the constraint  $\eta_{j,1}(\bar{\tau}_j) \leq \dots \leq \eta_{j,L-1}(\bar{\tau}_j)$ . In summary, after  $\boldsymbol{\zeta}$  is determined, the sensors and the fusion center employ the decision thresholds  $\boldsymbol{\eta}_j^*$  for rate  $R_k = j$ .

## 5.2 Probabilistic Rate Transmission

In order to determine the optimal transmission probabilities ( $\rho_j$ ) of each possible rate decision  $R_k = j$ , we minimize a lower bound on the average PCRLB (A-PCRLB) of the estimate subject to the total rate and utilization constraints.

### 5.2.1 The average FI of the estimate

From (5.6), the average PCRLB (APCRLB) of the estimate can be written as follows,

$$APCRLB = \sum_{\text{all } \mathbf{R}} \frac{1}{E[-\nabla_{\theta}^2 \log p(\mathbf{D}, \theta) | \mathbf{R}]} P(\mathbf{R}) \quad (5.13)$$

and  $P(\mathbf{R})$  is calculated from,

$$P(\mathbf{R}) = (P(R_1 = j_1)\rho_{j_1}) \dots (P(R_N = j_N)\rho_{j_N}) \quad (5.14)$$

where  $P(R_k = j)$  is the probability that a sensor quantizes its decision in  $j$  bits and  $\rho_j$  is the probability of sending  $j$  bits to the fusion center for a sensor. Therefore,  $P(R_k = j)\rho_j$  is the probability that the fusion center receives  $j$ -bit information from sensor  $s_k$ . The APCRLB can now be written as,

$$APCRLB = \sum_{j_1=1}^M \dots \sum_{j_N=1}^M \frac{P(R_1 = j_1), \dots, P(R_N = j_N)}{E[-\nabla_{\theta}^2 \log p(D_1, \dots, D_N, \theta) | R_1 = j_1, \dots, R_N = j_N]} \quad (5.15)$$

Note that the complexity to compute the APCRLB is very large since it requires an N-fold summation. To alleviate this problem, we first show in the following lemma that the inverse of the average Fisher Information is a lower bound on the APCRLB and we use the inverse of the average Fisher information as the objective function to be minimized for the optimization problem to find the optimal transmission probabilities.

**Lemma 5.1** *Let  $MSE_A = E \left\{ E \left[ (\theta - \hat{\theta})^2 | \mathbf{R} \right] \right\}$  be the MSE of the estimate averaged over all possible rate vectors  $\mathbf{R}$ . The  $MSE_A$  is lower bounded by the inverse of average Fisher information as,*

$$MSE_A \geq \frac{1}{J_a} \quad (5.16)$$

where

$$J_a = \sum_{R_1 \dots R_N} E \left[ -\nabla_{\theta}^{\theta} \log p(\mathbf{D}|\theta) | \mathbf{R} \right] p(\mathbf{R}) + 1/\sigma_{\theta}^2$$

is the average Fisher information and the expectation is taken over  $\mathbf{D}$ ,  $\theta$  and  $\mathbf{R}$  respectively.

The detailed proof of Lemma 5.1 is presented in Appendix A.2. The average FI for all the sensor data,  $J_a$ , is calculated from,

$$J_a = - \sum_{R_1 \dots R_N} \sum_{D_1 \dots D_N} \left[ \int_{\theta} (\nabla_{\theta}^{\theta} \log p(\mathbf{D}|\theta, \mathbf{R})) p(\mathbf{D}|\theta, \mathbf{R}) p(\theta) d\theta \right] p(\mathbf{R}) + \frac{1}{\sigma_{\theta}^2} \quad (5.17)$$

In the above definition, since the noise variance of each sensor is independent, the likelihood  $p(\mathbf{D}|\theta, \mathbf{R})$  has the form,

$$p(\mathbf{D}|\theta, \mathbf{R}) = \prod_{k=1}^N p(D_k|\theta, R_k) \quad (5.18)$$



Using (5.18) and (5.14) in (5.17) yields,

$$J_a = -N \left\{ \sum_{j=1}^M \left[ \sum_{l=0}^{L_j-1} \int_{\theta} \nabla_{\theta}^{\theta} \log p(D_k = l | \theta, R_k = j) p(D_k = l | \theta, R_k = j) d\theta \right] P(R_k = j) \rho_j \right\} + \frac{1}{\sigma_{\theta}^2} \quad (5.19)$$

We define  $\Upsilon(R_k = j) \triangleq E[-\nabla_{\theta}^{\theta} \log p(D_k | \theta, R_k = j)]$  as the average Fisher information when the quantized data are transmitted with rate  $R_k = j$ ,

$$\begin{aligned} \Upsilon(R_k = j) = & \quad (5.20) \\ & \sum_{l=0}^{L_j-1} \int_{\theta} \int_{\omega_k} -(\nabla_{\theta}^{\theta} \log p(D_k = l | R_k = j, \omega_k, \theta)) \times \\ & p(D_k = l | R_k = j, \omega_k, \theta) p(\omega_k | R_k = j) p(\theta) d\omega_k d\theta \end{aligned}$$

Note that  $\nabla_{\theta}^{\theta} \log p(D_k = l | R_k = j, \omega_k, \theta)$  can be calculated using an approach similar to that presented in [44]. Since  $\tau_k$  is  $\chi^2$ -distributed,  $\omega_k = \sqrt{\tau_k}$  becomes  $\chi$ -distributed as,

$$p(\omega_k) = \frac{\omega_k^{a-1} e^{-(\omega_k^2/2)}}{2^{a/2-1} \Gamma(a/2)} \quad (5.21)$$

with  $a = 1$  and

$$p(\omega_k | R_k = j) = \begin{cases} p(\omega_k)/p(R_k = j) & \sqrt{\zeta_{M-j}} \leq \omega_k < \sqrt{\zeta_{M-j+1}} \\ 0 & \text{otherwise} \end{cases} \quad (5.22)$$

The objective is to find the optimal transmission probabilities  $[\rho_1, \rho_2, \dots, \rho_M]$  that minimize the lower bound on the A-PCRLB as defined in Proposition 5.1 and (5.19),

$$J_a^{-1}(\rho_1, \rho_2, \dots, \rho_M) = \frac{1}{N \left( \sum_{j=1}^M \Upsilon(R_k = j) P(R_k = j) \rho_j \right) + \frac{1}{\sigma_{\theta}^2}} \quad (5.23)$$

### 5.2.2 Total Rate Constraint

Let  $B_k$  be the number of bits that the fusion center receives from sensor  $s_k$ . Then,  $B_k = R_k I(R_k)$  where  $I(\cdot)$  is the indicator function defined as,

$$I(R_k = j) = \begin{cases} 1 & \text{with probability } \rho_j \\ 0 & \text{with probability } 1 - \rho_j \end{cases} \quad (5.24)$$

The case  $\sum_{i=1}^N B_k > B$  results in an outage. Then, we would like to ensure that the outage rate  $C_{out}$  falls under a certain probability  $\epsilon$ .

$$C_{out} = p \left( \sum_{i=1}^N B_k > B \right) = Q \left( \frac{B - \nu}{\sigma} \right) \leq \epsilon \quad (5.25)$$

where the Central Limit Theorem has been used by assuming a large number of sensors, to approximate the distribution of the sum rate with a Gaussian distribution.  $Q(\cdot)$  is the complementary distribution function of a standard Gaussian distribution with zero mean and unit variance, and

$$\begin{aligned} \nu &= N \sum_{j=1}^M j P(R_k = j) \rho_j \\ \sigma &= \sqrt{\left( N \sum_{j=1}^M j^2 P(R_k = j) \rho_j \right) - \frac{\nu^2}{N}} \end{aligned} \quad (5.26)$$

### 5.2.3 Bandwidth utilization constraint

Under stringent bandwidth availability, minimizing the lower bound on MSE using only the total rate constraint may assign  $\rho_M \approx 1$ , since under high SNR, the average Fisher information of transmitting in  $M$ -bits  $\Upsilon(R_k = M)$  is much larger than the average Fisher information of other rates  $\Upsilon(R_k = j)$  where  $j < M$ . On the other hand, since  $P(R_k = M)$  is small as compared to all the other possible rates, for some instances there may be no sensors in the network that would employ rate  $R_k = M$ . Therefore, under stringent bandwidth availability, giving priority for the transmission of  $R_k = M$ -bits may result in bandwidth under-utilization. To alleviate this problem, we define another

constraint which ensures that on the average 100 $\delta\%$  of the bandwidth is utilized,

$$E \left[ \sum_{k=1}^N B_k \right] \geq \delta B \quad (5.27)$$

where  $\delta$  is the bandwidth utilization factor and  $\delta \in [0, 1]$ .

Thus, we solve the following constrained optimization problem to find the optimum transmission probabilities of each possible data rate.

$$\begin{aligned} \min_{\rho_1, \dots, \rho_M} f(\rho_1, \dots, \rho_M) &= \frac{1}{N \left( \sum_{j=1}^M \Upsilon(R_k = j) P(R_k = j) \rho_j \right) + \frac{1}{\sigma_\theta^2}} \\ \text{subject to } g_1(\rho_1, \dots, \rho_M) &= Q \left( \frac{B - \nu}{\sigma} \right) \leq \epsilon \\ g_2(\rho_1, \dots, \rho_M) &= E \left[ \sum_{k=1}^N B_k \right] \geq \delta B \end{aligned} \quad (5.28)$$

Note that the above optimization problem can also be formulated as follows,

$$\begin{aligned} \min_{\rho_1, \dots, \rho_M} f(\rho_1, \dots, \rho_M) &= \frac{1}{N \left( \sum_{j=1}^M \Upsilon(R_k = j) P(R_k = j) \rho_j \right) + \frac{1}{\sigma_\theta^2}} \\ \text{subject to } g_1(\rho_1, \dots, \rho_M) &= Q^{-1}(\epsilon) \sigma(\rho_1, \dots, \rho_M) + \nu(\rho_1, \dots, \rho_M) \leq b_1 \\ g_2(\rho_1, \dots, \rho_M) &= - \sum_{j=1}^M j P(R_k = j) \rho_j \leq b_2 \end{aligned} \quad (5.29)$$

where we redefine the constraints as  $g_1 = Q^{-1}(\epsilon) \sigma + \nu$  and  $g_2 = -\nu/N$  with  $b_1 = B$  and  $b_2 = -\delta B/N$ . Note that  $\nu$  and  $\sigma$  are functions of the decision variables  $(\rho_1, \dots, \rho_M)$  as defined in (5.26).

**Theorem 3 *Karush-Kuhn-Tucker (KKT) Conditions:*** *Since (5.29) is a minimization problem, let  $\bar{\rho} = [\bar{\rho}_1, \dots, \bar{\rho}_M]$  be an optimal solution to (5.29), then  $\bar{\rho} = [\bar{\rho}_1, \dots, \bar{\rho}_M]$  must satisfy the constraints in (5.30) and there exists multipliers  $\lambda_1, \lambda_2$*

satisfying [98]

$$\begin{aligned}
\frac{\partial f(\bar{\rho})}{\partial \rho_j} + \sum_{i=1}^m \lambda_i \frac{\partial g_i(\bar{\rho})}{\partial \rho_j} &= 0 \quad (j = 1, \dots, M) \\
\lambda_i [b_i - g_i(\bar{\rho})] &= 0 \quad (i = 1, 2) \\
\lambda_i &\geq 0 \quad (i = 1, 2)
\end{aligned} \tag{5.30}$$

**Theorem 4** *Since (5.29) is a minimization problem, if  $f(\rho_1, \dots, \rho_M)$  is a convex function and  $g_1(\rho_1, \dots, \rho_M)$  and  $g_2(\rho_1, \dots, \rho_M)$  are convex functions, then any point  $\bar{\rho} = [\bar{\rho}_1, \dots, \bar{\rho}_M]$  satisfying the constraints given in Theorem 3 is an optimal solution to (5.29).*

The reason that the hypothesis of Theorem 4 requires that each  $g_1(\rho_1, \dots, \rho_M)$  and  $g_2(\rho_1, \dots, \rho_M)$  be convex is that this ensures that the feasible region for (5.29) is a convex set [98].

The Hessian Matrix can be used to determine whether  $f(\rho_1, \dots, \rho_M)$ ,  $g_1(\rho_1, \dots, \rho_M)$  and  $g_2(\rho_1, \dots, \rho_M)$  are convex or concave functions. The Hessian of  $f(\rho_1, \dots, \rho_M)$  is the  $M \times M$  matrix whose  $\{i, j\}^{th}$  entry is

$$\frac{\partial^2 f}{\partial \rho_i \partial \rho_j}$$

Then an  $i^{th}$  principal minor of an  $M \times M$  matrix is the determinant of any  $i \times i$  matrix obtained by deleting  $M - i$  rows and  $M - i$  columns of the matrix.

**Definition 1** *A function  $f(\rho)$  is assumed to have continuous second-order partial derivatives for each point  $\rho = (\rho_1, \dots, \rho_M)$ . If all principal minors of the Hessian are non-negative,  $f(\rho)$  is a convex function for each  $\rho$ . For  $k = 1, \dots, M$ , if all non-zero principal minors of the Hessian have the same sign as  $(-1)^k$ , then  $f(\rho)$  is a concave function of  $\rho$ .*

The objective function  $f(\rho_1, \dots, \rho_M) = 1/\left[N \left(\sum_{j=1}^M \Upsilon(R_k = j)P(R_k = j)\rho_j\right) + \frac{1}{\sigma_\theta^2}\right]$  has continuous second-order partial derivatives for each point  $\rho = (\rho_1, \dots, \rho_M)$ . Then, we can easily show that  $f(\rho_1, \dots, \rho_M)$  is a convex function of  $\rho$ , since all princi-

pal minors of the Hessian are non-negative. The first constraint  $g_1(\rho_1, \dots, \rho_M) = Q^{-1}(\epsilon) \sigma(\rho_1, \dots, \rho_M) + \nu(\rho_1, \dots, \rho_M)$  is composed of two parts  $\sigma(\rho_1, \dots, \rho_M)$  and  $\nu(\rho_1, \dots, \rho_M)$  which both have continuous second-order partial derivatives. For  $\sigma(\rho_1, \dots, \rho_M)$ , due to the  $\sqrt{\cdot}$  operation, the first principal minors of the Hessian are negative. We can easily show that for  $k = 2, \dots, M$ , all non-zero principal minors of the Hessian have the same sign as  $(-1)^k$ . Then,  $\sigma(\rho_1, \dots, \rho_M)$  becomes a concave function of  $\rho$ . The function  $\nu(\rho_1, \dots, \rho_M)$  and the second constraint  $g_2(\rho_1, \dots, \rho_M) = -\sum_{j=1}^M jP(R_k = j)\rho_j$  are linear functions of the decision variables  $\rho = (\rho_1, \dots, \rho_M)$ . Linearity implies that  $g_2(\rho_1, \dots, \rho_M)$  is both convex and concave. Since  $\sigma(\rho_1, \dots, \rho_M)$  is concave and  $\nu(\rho_1, \dots, \rho_M)$  is convex, their sum  $g_1(\rho_1, \dots, \rho_M)$  is not a convex function ( $\partial^2 g_1(\rho) / \partial \rho_i \rho_j = \partial^2 \sigma(\rho) / \partial \rho_i \rho_j$ ). Therefore, as defined in Theorem 3 the feasible set is not convex. Hence, in the absence of convexity, a KKT point can be a global minimum, a local minimum, a saddlepoint, or even a local or global maximum. One way to obtain the optimal transmission coefficients is solve the KKT conditions given in Theorem 4 and pick the solution that minimizes the objective function from many local optimal solutions. On the other hand, although there are  $M+2$  equations and  $M+2$  unknowns, solving (5.29) using (5.30) is very hard, because when we solve it for  $\rho_j$ ,  $\rho_j$  becomes a non-linear function of high powers of other transmission probabilities  $\rho_l^n$  ( $l \neq j, n \gg 1$ ) and the multipliers  $\lambda_1, \lambda_2$ .

Heuristic techniques have been widely used to solve difficult optimization problems. As in the problem presented in this section, when optimal solutions are difficult to obtain, heuristic techniques tend to exploit the structure of the problem and arrive at a good solution. Therefore, we solve the optimization problems presented in (5.29) numerically by using a genetic search whose algorithm is provided by MATLAB.

### 5.3 Parameter Estimation based on Received Sensor Data

Let us assume that the fusion center receives data  $\mathbf{D}$  with rates  $\mathbf{R}$  from the sensors and let  $\hat{\theta}$  be the Bayesian estimate of the parameter  $\theta$  based on  $\mathbf{D}$  and  $\mathbf{R}$ . The fusion

center calculates  $\hat{\theta}$  according to,

$$\hat{\theta}(\mathbf{D}, \mathbf{R}) = \int_{\theta} \theta p(\theta | \mathbf{D}, \mathbf{R}) d\theta = \frac{\int_{\theta} \theta p(\mathbf{D} | \theta, \mathbf{R}) p(\theta) d\theta}{\int_{\theta} p(\mathbf{D} | \theta, \mathbf{R}) p(\theta) d\theta} \quad (5.31)$$

where  $p(\mathbf{D} | \theta, \mathbf{R})$  is computed using (5.18). The likelihood of a sensor deciding on  $l$ ,  $p(D_k = l | \theta, R_k)$ , is further computed as follows.

### 5.3.1 The Likelihood under probabilistic rate transmission

For the probabilistic rate transmission case, the fusion center computes the likelihood of an individual sensor  $p(D_k = l | \theta, R_k)$  according to,

$$p(D_k = l | \theta, R_k) = \int_{\sqrt{\zeta_{M-j}}}^{\sqrt{\zeta_{M-j+1}}} p(D_k = l | \omega_k, \theta, R_k) p(\omega_k | R_k) d\omega_k \quad (5.32)$$

where

$$p(D_k = l | \omega_k, \theta, R_k) = Q\left(\frac{\eta_{R_k, l} - \theta}{\omega_k}\right) - Q\left(\frac{\eta_{R_k, (l+1)} - \theta}{\omega_k}\right)$$

### 5.3.2 The Likelihood under equal rate transmission

For performance comparison, we consider a simple equal rate transmission scheme where the total bandwidth is evenly distributed among all the sensors in the network and  $\rho_1 = \dots = \rho_N = 1$ . In other words, all the sensors quantize their measurements with rate  $R_k = B/N$  bits and transmit to the fusion center. The likelihood function of an individual sensor decision can then be calculated as,

$$p(D_k = l | \theta, R_k) = \int_0^{\infty} p\left(D_k = l | \omega_k, \theta, R_k = \frac{B}{N}\right) p(\omega_k) d\omega_k \quad (5.33)$$

We can either use (5.32) or (5.33) in (5.31) to compute  $\hat{\theta}$ . In the next section, we present some illustrative examples.

#### 5.4 Simulation Results

We use  $N = 20$  sensors in the WSN. For the probabilistic rate transmission scheme, we consider two different cases. The first case is the conservative transmission case where  $\epsilon = 0.01$  and  $\delta = 0.5$  and the second case is the liberal transmission case where  $\epsilon = 0.1$  and  $\delta = 0.7$ .

We present the optimal transmission probabilities for conservative and liberal cases for rate  $R_k$  in Table 5.2. In order to meet the average bandwidth utilization constraint  $\delta$ , when  $B$  is small, to utilize the bandwidth, the sensors with smaller rates are given priority for transmission. As the total allowable bandwidth ( $B$ ) increases, the solution of the optimization problem gives priority to the transmission of  $R_k = 6$  bit information because of the better  $\text{SNR}_k$  and high precision of the quantized measurements. Also note that, as  $B$  increases, the sensors with rates  $R_k = 2$  and  $R_k = 3$  bits transmit with less probability due to their poorer  $\text{SNR}_k$ . Hence, the probabilistic bit allocation scheme saves energy by reducing the probability that a sensor with low SNR transmits its data. The optimal transmission probabilities of the conservative case are comparatively smaller than the liberal case to ensure less outage probability.

Fig. 5.3 compares the mean squared error (MSE) of the proposed algorithm with a simple scheme where the total bandwidth is equally distributed among the sensors. For example, if  $N = 20$  and  $B = 20$ , in the equal bit allocation scheme, 20 sensors send their data in  $R_k = 1$  bit. Also, if  $N = 20$  and  $B = 30$ , in the equal bit allocation scheme, 15 sensors send their information in  $R_k = 2$  bits and so on. The MSE of the estimation is obtained by averaging over 1000 different trials. The probabilistic bit allocation scheme outperforms the equal bit allocation scheme significantly when the total bandwidth is small. As  $B$  increases, all the sensors are able to quantize their data to a larger number of bits and the MSE performance of the equal bit allocation scheme gets closer to the MSE performance of probabilistic bit allocation scheme and finally converges to the case where all the sensors transmit with the maximum rate.

Table 5.1: Optimal transmission probabilities of each possible transmission rate  $R_k = j$  bits. (Conservative case:  $\epsilon = 0.01$ ,  $\delta = 0.5$ )

	$R_k = 1$	$R_k = 2$	$R_k = 3$	$R_k = 4$	$R_k = 5$	$R_k = 6$
$B = 20$	0.2667	0.7644	0.1432	0.0588	0.0001	0.0017
$B = 30$	0.9066	0.5252	0.1192	0.2005	0.0225	0.1790
$B = 40$	0.9775	0.3593	0.0925	0.0000	0.1560	0.5580
$B = 50$	0.4374	0.0001	0.0000	0.0175	0.1968	1.0000
$B = 60$	0.1995	0.0001	0.0001	0.1129	0.9999	1.0000
$B = 70$	0.0001	0.0000	0.0000	0.7835	1.0000	1.0000
$B = 80$	0.9700	0.6869	0.8409	0.6631	0.7426	1.0000

Table 5.2: Optimal transmission probabilities of each possible transmission rate  $R_k = j$  bits. (Liberal case:  $\epsilon = 0.1$ ,  $\delta = 0.7$ )

	$R_k = 1$	$R_k = 2$	$R_k = 3$	$R_k = 4$	$R_k = 5$	$R_k = 6$
$B = 20$	0.6605	0.9283	0.3442	0.0001	0	0.0001
$B = 30$	0.0646	0.9996	0.2720	0.0000	0.9003	0.0172
$B = 40$	0.6802	0.8592	0.2640	0.1105	0.0874	0.6151
$B = 50$	0.4720	0.0001	0.0001	0.2122	1.0000	1.0000
$B = 60$	0.9712	0.6508	0.3228	0.2484	0.9548	1.0000
$B = 70$	0.9988	0.6529	0.8429	0.4860	1.0000	1.0000
$B = 80$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

## 5.5 Discussion

In this chapter, we considered the distributed estimation problem and presented the probabilistic rate transmission scheme, which minimizes the lower bound on the average PCRLB to find the optimal transmission probabilities corresponding to possible data rates under total bandwidth and network utilization constraints. Simulation results show that under stringent constraint on available bandwidth, the probabilistic rate transmission scheme outperforms the equal rate transmission scheme. By allowing transmissions only from the high SNR sensors, the MSE of estimation is very close to the case when all the sensors transmit using  $M$ -bits. Therefore, the probabilistic rate transmission scheme saves bandwidth significantly yet achieving fairly accurate estimation.



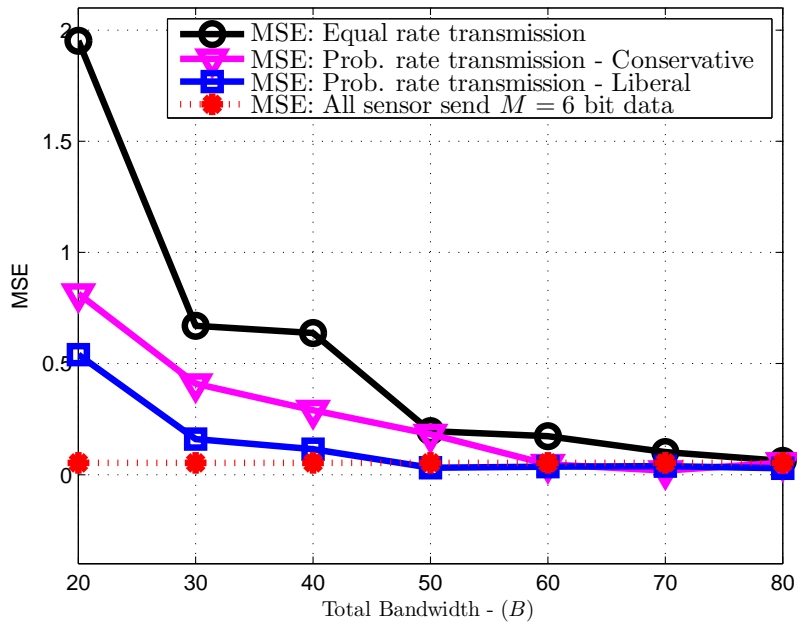


Figure 5.3: MSE of probabilistic vs equal rate transmission schemes,  $\mu_\theta = 0$  and  $\sigma_\theta^2 = 10$ .

## Chapter 6

# Concluding Remarks and Suggestions for Future Work

In this dissertation, resource aware distributed detection and estimation of random events in WSNs has been investigated. The focus of the dissertation was to develop novel methods which provide significant savings in resource consumption such as energy and bandwidth while keeping relatively high detection/estimation performance at the cost of slight potential increase on decision error probability, estimation delay or outage probability.

We first proposed a multi-objective optimization approach for providing different alternatives to WSN design for the detection task. Rather than only minimizing the global probability of error of the network, an MOP approach finds a number of trade-off solutions between the two objectives namely probability of decision error and the total energy consumption. A number of alternative solutions provide significant energy savings as compared to the minimum error solution at the cost of slightly increasing the best achievable global probability of error. The proposed MOP can be easily extended to multi-objective problems with more than two objectives as well as under specified constraints. Future work will include adapting the proposed framework to larger number of sensors, and more than two objectives as well as development of computationally efficient approaches. Extension of our methodology for a general network topology with multiple events occurring at the same time will also be addressed. For this purpose, earlier work of Alhakeem and Varshney [99] could be used.

For static source localization, simultaneously requesting complete sensor data may incur resource challenges in practice. Hence, we have presented an iterative source localization scheme which arrives at a coarse location estimate first using anchor sensor data and the location estimate is refined by activating a number of non-anchor sensors in an iterative manner. By only selecting the most informative sensors about the source location, the iterative localization method provides significant energy and communication savings as compared to one-shot location estimation, at the cost of tolerating some estimation latency. We have developed a Monte Carlo-based method to approximate the posterior probability distribution function of the source location and its Bayesian estimate. Using this approximate posterior probability distribution function of the source location, we have developed and compared two sensor selection metrics which are based on mutual information between source location and sensor locations and posterior Cramer-Rao Lower bound of the estimate. We first considered the problem under the assumption of perfect communication channels where both sensor selection schemes achieve similar estimation performance. PCRLB-based sensor selection is found to be computationally more efficient than the MI-based sensor selection scheme. We next considered the case where the channels between sensors and the fusion center are subject to fading. Under phase coherent reception, we derived the Bayesian estimate of the source location for multi-bit data and showed that using channel gain statistics only yields near optimal performance. We also extended the sensor selection metrics which include the channel impairments. The case where the fusion center employs non-coherent reception for multi-bit sensor data can be considered as a future work. Suggested future work also includes defining the communication costs in terms of more specific path loss models. A theoretical framework can be developed to study the trade-off between estimation performance in source localization and energy costs. An extension of our methodology for localization of multiple sources can also be addressed.

In a wireless video sensor network which employs multiple camera sensors, the problem of selecting the most appropriate camera or a set of cameras to perform target localization is an important task to balance the trade-off between the localization accuracy and the energy consumption. The mutual information-based camera selection

method has been recently introduced for wireless video sensor networks [100], [101] The camera selection based on the PCRLB-based sensor selection is also suitable for wireless video sensor networks and can be considered in a future work.

Bit allocation is an important problem in the rate-constrained distributed estimation problems. We proposed a fully distributed scheme for heterogenous WSNs, where the quantization data rate of a sensor has been determined as a function of its observation SNR. In order not to exceed the allowed bandwidth, after deciding on a data rate, the quantized observation of each sensor is transmitted to the fusion center with a certain probability. Under stringent constraint on available bandwidth, the probabilistic approach favors the use of sensors with high observation SNR, and provides a better estimation performance as compared to the case where the available bandwidth is equally distributed among the sensors. In this work, we assumed that the wireless channels between sensors and fusion center are error free, which can be realized by sufficient transmission energy per each sensor. As a future work, the proposed framework can be further extended while considering energy limitations as well as including the channel impairments. Notice that the estimation performance improves with the multibit data coming from the sensors having good quality of observations with higher reliability. On the other hand, for a fixed transmission energy, mapping the multibit data into an M-ary symbol increases the probability of symbol error at the fusion center. Such a trade-off can be examined by solving a multiobjective optimization problem. A MOP can be defined to find the trade-off solutions between the two objectives, MSE of the estimation and the energy consumption of the network.

# Appendix A

## Proof of Lemmas

### A.1 Proof of Lemma 4.1

**Proof.** If the channel noise is independent and identically distributed among the the received symbols of each sensor, the likelihood function of the received symbols has the form,

$$p(\mathbf{R}_k | D_k = l) = p([r_{k,1}, r_{k,2}, \dots, r_{k,M}] | [q_{k,1}, q_{k,2}, \dots, q_{k,M}]) = \quad (\text{A.1})$$

$$\int_{h_k} \left( \prod_{j=1}^M p(r_{k,j} | h_k, q_{k,j}) \right) p(h_k) dh_k$$

In the above equation we assume that entire symbols transmitted from a sensor experience the same channel. Assuming a Rayleigh fading channel with unit power (i.e.,  $E[h_k^2] = 1$ ), the pdf of  $h_k$  is expressed as,

$$p(h_k) = 2h_k \exp(-h_k^2), \quad h_k \geq 0 \quad (\text{A.2})$$

In Equation (A.1), the distribution  $p(r_{k,j} | h_k, q_{k,j})$  is expressed as,

$$p(r_{k,j} | q_{k,j}, h_k) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(r_{k,j} - \sqrt{\epsilon_c} h_k q_{k,j})^2}{2\sigma^2}\right) \quad (\text{A.3})$$

Than (A.1) can be written as,

$$p(\mathbf{R}_k | D_k = l) = \int_{h_k} \frac{2h_k \exp(-h_k^2)}{(2\pi)^{M/2} \sigma^M} \exp\left(-\sum_{j=1}^M \frac{(r_{k,j} - \sqrt{\epsilon_c} h_k q_{k,j})^2}{2\sigma^2}\right) dh_k \quad (\text{A.4})$$

$$= \int_{h_k} \frac{2h_k \exp(-h_k^2)}{(2\pi)^{M/2} \sigma^M} \exp\left(-\frac{1}{2\sigma^2} \sum_{j=1}^M (r_{k,j}^2 - 2h_k \sqrt{\epsilon_c} r_{k,j} q_{k,j} + \epsilon_c h_k^2 q_{k,j}^2)\right) dh_k \quad (\text{A.5})$$

$$= \int_{h_k} \frac{2h_k}{(2\pi)^{M/2} \times \sigma^M} \exp\left(-\frac{1}{2\sigma^2} \left\{ \left( \sum_{j=1}^M r_{k,j}^2 \right) - 2h_k \sqrt{\epsilon_c} \left( \sum_{j=1}^M r_{k,j} q_{k,j} \right) + \epsilon_c h_k^2 \sum_{j=1}^M q_{k,j}^2 + h_k^2 2\sigma^2 \right\}\right) dh_k \quad (\text{A.6})$$

Note that  $\sum_{j=1}^M q_{k,j}^2 = M$ , then,

$$= \int_{h_k} \frac{2h_k}{(2\pi)^{M/2} \sigma^M} \times \exp\left(-\frac{1}{2\sigma^2} \left\{ \left( \sum_{j=1}^M r_{k,j}^2 \right) - 2h_k \sqrt{\epsilon_c} \left( \sum_{j=1}^M r_{k,j} q_{k,j} \right) + h_k^2 (2\sigma^2 + \epsilon_c M) \right\}\right) dh_k \quad (\text{A.7})$$

$$= \int_{h_k} \frac{2h_k}{(2\pi)^{M/2} \sigma^M} \exp\left(-\frac{\left\{ \frac{(\sum_{j=1}^M r_{k,j}^2)}{(2\sigma^2 + \epsilon_c M)} - \frac{2h_k \sqrt{\epsilon_c} (\sum_{j=1}^M r_{k,j} q_{k,j})}{(2\sigma^2 + \epsilon_c M)} + h_k^2 \right\}}{\frac{2\sigma^2}{(2\sigma^2 + \epsilon_c M)}}\right) dh_k \quad (\text{A.8})$$

$$= \frac{2}{(2\pi)^{M/2} \sigma^M} \int_{h_k} h_k \times \quad (\text{A.9})$$

$$\exp\left(-\frac{\left\{ \frac{(\sum_{j=1}^M r_{k,j}^2)}{(2\sigma^2 + \epsilon_c M)} - \frac{\epsilon_c (\sum_{j=1}^M r_{k,j} q_{k,j})^2}{(2\sigma^2 + \epsilon_c M)^2} + \frac{\epsilon_c (\sum_{j=1}^M r_{k,j} q_{k,j})^2}{(2\sigma^2 + \epsilon_c M)^2} - \frac{2h_k \sqrt{\epsilon_c} (\sum_{j=1}^M r_{k,j} q_{k,j})}{(2\sigma^2 + \epsilon_c M)} + h_k^2 \right\}}{\frac{2\sigma^2}{(2\sigma^2 + \epsilon_c M)}}\right) dh_k$$

$$= \frac{2}{(2\pi)^{M/2} \sigma^M} \exp\left(-\frac{\frac{\sum_{j=1}^M r_{k,j}^2}{2\sigma^2 + \epsilon_c M} - \frac{\epsilon_c (\sum_{j=1}^M r_{k,j} q_{k,j})^2}{(2\sigma^2 + \epsilon_c M)^2}}{\frac{2\sigma^2}{2\sigma^2 + \epsilon_c M}}\right) \times \quad (\text{A.10})$$

$$\int_{h_k} h_k \exp\left(-\frac{\left\{ \frac{\epsilon_c (\sum_{j=1}^M r_{k,j} q_{k,j})^2}{(2\sigma^2 + \epsilon_c M)^2} - \frac{2h_k \sqrt{\epsilon_c} (\sum_{j=1}^M r_{k,j} q_{k,j})}{(2\sigma^2 + \epsilon_c M)} + h_k^2 \right\}}{\frac{2\sigma^2}{(2\sigma^2 + \epsilon_c M)}}\right) dh_k$$

$$\begin{aligned}
&= \frac{2}{(2\pi)^{M/2} \sigma^M} \exp \left( -\frac{\frac{\sum_{j=1}^M r_{k,j}^2}{2\sigma^2 + \epsilon_c M} - \frac{\epsilon_c \left(\sum_{j=1}^M r_{k,j} q_{k,j}\right)^2}{(2\sigma^2 + \epsilon_c M)^2}}{\frac{2\sigma^2}{2\sigma^2 + \epsilon_c M}} \right) \times \\
&\int_{h_k} h_k \exp \left( -\frac{\left( h_k - \frac{\sqrt{\epsilon_c} \left(\sum_{j=1}^M r_{k,j} q_{k,j}\right)}{(2\sigma^2 + \epsilon_c M)} \right)^2}{\frac{2\sigma^2}{(2\sigma^2 + \epsilon_c M)}} \right) dh_k
\end{aligned} \tag{A.11}$$

Let  $t = h_k - \frac{\sqrt{\epsilon_c} \left(\sum_{j=1}^M r_{k,j} q_{k,j}\right)}{(2\sigma^2 + \epsilon_c M)}$ . Then,

$$\begin{aligned}
&= \frac{2}{(2\pi)^{M/2} \sigma^M} \exp \left( -\frac{\frac{\sum_{j=1}^M r_{k,j}^2}{2\sigma^2 + \epsilon_c M} - \frac{\epsilon_c \left(\sum_{j=1}^M r_{k,j} q_{k,j}\right)^2}{(2\sigma^2 + \epsilon_c M)^2}}{\frac{2\sigma^2}{2\sigma^2 + \epsilon_c M}} \right) \\
&\int_t \left( t + \frac{\sqrt{\epsilon_c} \left(\sum_{j=1}^M r_{k,j} q_{k,j}\right)}{(2\sigma^2 + \epsilon_c M)} \right) \exp \left( -\frac{t^2}{\frac{2\sigma^2}{(2\sigma^2 + \epsilon_c M)}} \right) dt
\end{aligned} \tag{A.12}$$

Let us evaluate the first summation term inside the integral

$$\begin{aligned}
&\int_t t \exp \left( -\frac{t^2 (2\sigma^2 + \epsilon_c M)}{2\sigma^2} \right) dt \\
&= -\frac{2\sigma^2}{2(2\sigma^2 + \epsilon_c M)} \left( \int_t -\frac{2t (2\sigma^2 + \epsilon_c M)}{2\sigma^2} \exp \left( -\frac{t^2 (2\sigma^2 + \epsilon_c M)}{2\sigma^2} \right) dt \right) \\
&= -\frac{2\sigma^2}{2(2\sigma^2 + \epsilon_c M)} \left[ \exp \left( -\frac{t^2 (2\sigma^2 + \epsilon_c M)}{2\sigma^2} \right) \right]_{-\frac{\sqrt{\epsilon_c} \left(\sum_{j=1}^M r_{k,j} q_{k,j}\right)}{(2\sigma^2 + \epsilon_c M)}}^\infty \\
&= \frac{2\sigma^2}{2(2\sigma^2 + \epsilon_c M)} \exp \left( -\frac{\epsilon_c \left(\sum_{j=1}^M r_{k,j} q_{k,j}\right)^2}{2\sigma^2 (2\sigma^2 + \epsilon_c M)} \right)
\end{aligned} \tag{A.13}$$

Now let's compute the second term inside the summation,

$$= \left( \frac{\sqrt{\epsilon_c} \left(\sum_{j=1}^M r_{k,j} q_{k,j}\right)}{(2\sigma^2 + \epsilon_c M)} \right) \frac{\sqrt{2\pi} \sigma}{\sqrt{2\sigma^2 + \epsilon_c M}} \times \tag{A.14}$$

$$\begin{aligned}
& \int_{-\frac{\sqrt{\epsilon_c} \left( \sum_{j=1}^M r_{k,j} q_{k,j} \right)}{(2\sigma^2 + \epsilon_c M)}}^{\infty} \frac{1}{\sqrt{2\pi} \frac{\sigma}{\sqrt{2\sigma^2 + \epsilon_c M}}} \exp \left( -\frac{t^2}{(2\sigma^2 + \epsilon_c M)} \right) dt \\
&= \left( \frac{\sqrt{\epsilon_c} \left( \sum_{j=1}^M r_{k,j} q_{k,j} \right)}{(2\sigma^2 + \epsilon_c M)} \right) \frac{\sqrt{2\pi}\sigma}{\sqrt{2\sigma^2 + \epsilon_c M}} Q \left( -\frac{\sqrt{\epsilon_c} \left( \sum_{j=1}^M r_{k,j} q_{k,j} \right)}{\sigma \sqrt{2\sigma^2 + \epsilon_c M}} \right)
\end{aligned}$$

Using (A.13) and (A.14) in (A.12) results,

$$\begin{aligned}
&= \frac{2}{(2\pi)^{M/2} \sigma^M} \exp \left( -\frac{\sum_{j=1}^M r_{k,j}^2}{2\sigma^2} + \frac{\epsilon_c \left( \sum_{j=1}^M r_{k,j} q_{k,j} \right)^2}{2\sigma^2 (2\sigma^2 + \epsilon_c M)} \right) \times \tag{A.15} \\
&\quad \left[ \frac{2\sigma^2}{2(2\sigma^2 + \epsilon_c M)} \exp \left( -\frac{\epsilon_c \left( \sum_{j=1}^M r_{k,j} q_{k,j} \right)^2}{2\sigma^2 (2\sigma^2 + \epsilon_c M)} \right) + \right. \\
&\quad \left. \left( \frac{\sqrt{\epsilon_c} \left( \sum_{j=1}^M r_{k,j} q_{k,j} \right)}{(2\sigma^2 + \epsilon_c M)} \right) \frac{\sqrt{2\pi}\sigma}{\sqrt{2\sigma^2 + \epsilon_c M}} Q \left( -\frac{\sqrt{\epsilon_c} \left( \sum_{j=1}^M r_{k,j} q_{k,j} \right)}{\sigma \sqrt{2\sigma^2 + \epsilon_c M}} \right) \right] \\
&= \frac{2}{(2\pi)^{M/2} \sigma^M} \frac{2\sigma^2}{2(2\sigma^2 + \epsilon_c M)} \times \\
&\quad \exp \left( -\frac{\sum_{j=1}^M r_{k,j}^2}{2\sigma^2} + \frac{\epsilon_c \left( \sum_{j=1}^M r_{k,j} q_{k,j} \right)^2}{2\sigma^2 (2\sigma^2 + \epsilon_c M)} - \frac{\epsilon_c \left( \sum_{j=1}^M r_{k,j} q_{k,j} \right)^2}{2\sigma^2 (2\sigma^2 + \epsilon_c M)} \right) \times \\
&\quad \left[ 1 + \left\{ \frac{(2\sigma^2 + \epsilon_c M)}{\sigma^2} \left( \frac{\sqrt{\epsilon_c} \left( \sum_{j=1}^M r_{k,j} q_{k,j} \right)}{(2\sigma^2 + \epsilon_c M)} \right) \frac{\sqrt{2\pi}\sigma}{\sqrt{2\sigma^2 + \epsilon_c M}} \times \right. \right. \\
&\quad \left. \left. \exp \left( \frac{\epsilon_c \left( \sum_{j=1}^M r_{k,j} q_{k,j} \right)^2}{2\sigma^2 (2\sigma^2 + \epsilon_c M)} \right) Q \left( -\frac{\sqrt{\epsilon_c} \left( \sum_{j=1}^M r_{k,j} q_{k,j} \right)}{\sigma \sqrt{2\sigma^2 + \epsilon_c M}} \right) \right\} \right]
\end{aligned}$$

Let,

$$\beta = \frac{\sqrt{\epsilon_c}}{\sigma \sqrt{2\sigma^2 + \epsilon_c M}} \tag{A.16}$$



Then (A.15) can be simplified as,

$$p(R_k|D_k = l) = \frac{2}{(2\pi)^{M/2}\sigma^{M-2}(2\sigma^2 + \epsilon_c M)} \exp\left(-\frac{\sum_{j=1}^M r_{k,j}^2}{2\sigma^2}\right) \times \quad (\text{A.17})$$

$$\left[ 1 + \sqrt{2\pi}\beta \left( \sum_{j=1}^M r_{k,j} q_{k,j} \right) \exp\left(\frac{\beta^2 \left( \sum_{j=1}^M r_{k,j} q_{k,j} \right)^2}{2}\right) Q\left(-\beta \left( \sum_{j=1}^M r_{k,j} q_{k,j} \right)\right) \right]$$

■

## A.2 Proof of Lemma 5.1

**Proof.** Given the rate vector  $\mathbf{R}$ , the MSE is lower bounded by its PCRLB( $\mathbf{R}$ ),

$$E \left[ (\theta - \hat{\theta})^2 | \mathbf{R} \right] \geq \text{PCRLB}(\mathbf{R}) \quad (\text{A.18})$$

where

$$\text{PCRLB}(\mathbf{R}) = \frac{1}{E \left[ -\nabla_{\theta}^2 \log p(\mathbf{D}|\theta) | \mathbf{R} \right] + 1/\sigma_{\theta}^2}$$

Let  $\xi \triangleq E \left[ -\nabla_{\theta}^2 \log p(\mathbf{D}|\theta) | \mathbf{R} \right]$  and calculated from,

$$\xi = \int_{\theta} \sum_{D_1 \dots D_N} -\nabla_{\theta}^2 \log p(\mathbf{D}|\theta, \mathbf{R}) p(\mathbf{D}, \theta | \mathbf{R}) d\theta$$

where  $\xi > 0$ . Further, let

$$f(\xi) \triangleq \frac{1}{\xi + 1/\sigma_{\theta}^2}$$

Averaging both sides of the inequality (A.18) over  $\mathbf{R}$  yields,

$$\text{MSE}_A \geq E \{ f(\xi) \} \quad (\text{A.19})$$

Since  $f''(\xi) > 0$ ,  $f(\xi)$  is a convex function of  $\xi$  and  $f(\xi)$  satisfies Jensen's inequality  $E \{ f(\xi) \} \geq f [ E (\xi) ]$ . Hence,

$$\text{MSE}_A \geq \frac{1}{E \{ E \left[ -\nabla_{\theta}^2 \log p(\mathbf{D}|\theta) | \mathbf{R} \right] \} + 1/\sigma_{\theta}^2} \quad (\text{A.20})$$

where the expectation is first taken over  $\mathbf{D}$  and  $\theta$  and then over  $\mathbf{R}$ . ■

# Bibliography

- [1] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, “A survey on sensor networks,” *IEEE Communications Magazine*, vol. 40, no. 8, pp. 102–114, Aug 2002.
- [2] P. K. Dutta, *On Random Event Detection with Wireless Sensor Networks*. M. Sc. Thesis, Ohio State University, 2004.
- [3] P. K. Varshney, *Distributed Detection and Data Fusion*. Springer-Verlag, 1996.
- [4] 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.
- [5] 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.
- [6] I. Bahceci, *Multiple-Input Multiple-Output Wireless Systems: Coding, Distributed Detection and Antenna Selection*. PhD. Thesis, Georgia Institute of Technology, 2005.
- [7] I. Schizas, A. Ribeiro, and G. Giannakis, “Consensus in ad hoc WSNs with noisy links; Part I: Distributed estimation of deterministic signals,” *IEEE Transactions on Signal Processing*, vol. 56, no. 1, pp. 350–364, Jan. 2008.
- [8] A. Dogandzic and B. Zhang, “Distributed estimation and detection for sensor networks using hidden Markov random field models,” *IEEE Transactions on Signal Processing*, vol. 54, no. 8, pp. 3200–3215, Aug. 2006.

- [9] 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, Aug. 2004.
- [10] N. Patwari and A. Hero, "Hierarchical censoring for distributed detection in wireless sensor networks," vol. 4, April 2003, pp. IV – 848–51 vol.4.
- [11] 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, Jan 1997.
- [12] R. Jiang and B. Chen, "Fusion of censored decisions in wireless sensor networks," *IEEE Transactions on Wireless Communications*, vol. 4, no. 6, pp. 2668–2673, Nov. 2005.
- [13] H. L. Van-Trees, *Detection. Estimation and Modulation Theory*. New York: Wiley, 1968.
- [14] F. Zhao, J. Shin, and J. Reich, "Information-driven dynamic sensor collaboration," *IEEE Signal Processing Magazine*, vol. 19, no. 2, pp. 61–72, Mar 2002.
- [15] F. Zhao, J. Liu, J. Liu, L. Guibas, and J. Reich, "Collaborative signal and information processing: an information-directed approach," *Proceedings of the IEEE*, vol. 91, no. 8, pp. 1199–1209, Aug. 2003.
- [16] C. M. Kreucher, K. D. Kastella, and A. O. Hero, "Sensor management using an active sensing approach," *ELSEVIER Signal Processing*, vol. 85, no. 3, pp. 607–624, March 2005.
- [17] C. Kreucher, A. Hero, K. Kastella, and M. Morelande, "An information-based approach to sensor management in large dynamic networks," *Proceedings of the IEEE*, vol. 95, no. 5, pp. 978–999, May 2007.
- [18] K. Punithakumar, T. Kirubarajan, and M. Hernandez, "Multisensor deployment using pcrlbs, incorporating sensor deployment and motion uncertainties," *IEEE*

- Transactions on Aerospace and Electronic Systems*, vol. 42, no. 4, pp. 1474–1485, October 2006.
- [19] R. Tharmarasa, T. Kirubarajan, and M. Hernandez, “Large-scale optimal sensor array management for multitarget tracking,” *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 37, no. 5, pp. 803–814, Sept. 2007.
- [20] H. V. Poor, *An Introduction to Signal Detection and Estimation*. Springer, 1994.
- [21] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, 1993.
- [22] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, 1991.
- [23] G. Kramer, “Topics in multi-user information theory,” *Foundations and Trends in Communications and Information Theory*, vol. 4, no. 4-5, pp. 265–444, 2007.
- [24] A. Doucet and X. Wang, “Monte Carlo methods for signal processing: a review in the statistical signal processing context,” *IEEE Signal Processing Magazine*, vol. 22, no. 6, pp. 152–170, Nov. 2005.
- [25] I. Das and J. Dennis, “Normal-boundary interaction: A new method for generating the pareto surface in nonlinear multicriteria optimization problems,” *SIAM Journal of Optimization*, vol. 8, pp. 631–657, 1998.
- [26] K. Deb, *Multi-Objective Optimization using Evolutionary Algorithms*. John Wiley and Sons, 2001.
- [27] E. Zitzler, *Evolutionary Algorithms for Multiobjective Optimization: Methods and Applications*. PhD. Dissertation, Swiss Federal Institute of Technology, 1999.
- [28] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, “A fast and elitist multiobjective genetic algorithm: NSGA-II,” *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, Apr 2002.

- [29] R. Rajagopalan, C. K. Mohan, K. G. Mehrotra, and P. K. Varshney, “Emoca: An evolutionary multi-objective crowding algorithm,” *Journal of Intelligent Systems*, vol. 17, no. 1/3, pp. 107–123, 2008.
- [30] P. K. Shukla and K. Deb, “On finding multiple pareto-optimal solutions using classical and evolutionary generating methods.” [Online]. Available: <http://www.iitk.ac.in/kangal/reports.shtml>
- [31] L. T. Bui and S. Alam, “Multi-objective optimization in computational intelligence: Theory and practice,” *Idea Group Inc*, pp. 208–238, 2008.
- [32] E. L. Lehmann, *Testing statistical hypothesis*. New York: Wiley, 1986.
- [33] R. Ahlswede and I. Csiszar, “Hypothesis testing with communication constraints,” *IEEE Transactions on Information Theory*, vol. 32, no. 4, pp. 533–542, Jul 1986.
- [34] H. Shalaby and A. Papamarcou, “Multiterminal detection with zero-rate data compression,” *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 254–267, Mar 1992.
- [35] T. Han, “Hypothesis testing with multiterminal data compression,” *IEEE Transactions on Information Theory*, vol. 33, no. 6, pp. 759–772, Nov 1987.
- [36] J. N. Tsitsiklis, “Decentralized detection by a large number of sensors,” *Math. Control, Signals, Syst.*, vol. 1, no. 2, pp. 167–182, 1988.
- [37] 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.
- [38] Z.-B. Tang, K. Pattipati, and D. Kleinman, “An algorithm for determining the decision thresholds in a distributed detection problem,” *IEEE Transactions on Systems, Man and Cybernetics*, vol. 21, no. 1, pp. 231–237, Jan/Feb 1991.

- [39] B. Chen, L. Tong, and P. K. Varshney, “Channel-aware distributed detection in wireless sensor networks,” *IEEE Signal Processing Magazine*, vol. 23, no. 4, pp. 16–26, July 2006.
- [40] X. Zhang, H. V. Poor, and M. Chiang, “Optimal power allocation for distributed detection over MIMO channels in wireless sensor networks,” *IEEE Transactions on Signal Processing*, vol. 56, no. 9, pp. 4124–4140, Sept. 2008.
- [41] V. Aalo and R. Viswanathou, “On distributed detection with correlated sensors: two examples,” *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 25, no. 3, pp. 414–421, May 1989.
- [42] J.-F. Chamberland and 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, Nov. 2006.
- [43] P. Willett, P. Swaszek, and R. 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, Dec 2000.
- [44] 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, Dec. 2006.
- [45] —, “Performance analysis of distributed detection in a random sensor field,” *IEEE Transactions on Signal Processing*, vol. 56, no. 1, pp. 339–349, Jan. 2008.
- [46] S. Appadwedula, V. Veeravalli, and D. Jones, “Decentralized detection with censoring sensors,” *IEEE Transactions on Signal Processing*, vol. 56, no. 4, pp. 1362–1373, April 2008.
- [47] R. Jiang, Y. Lin, B. Chen, and B. Suter, “Distributed sensor censoring for detection in sensor networks under communication constraints,” in *Thirty-Ninth Asilomar Conference on Signals, Systems and Computers*, 28 - November 1, 2005, pp. 946–950.

- [48] X. Sheng and Y.-H. Hu, “Maximum likelihood multiple-source localization using acoustic energy measurements with wireless sensor networks,” *IEEE Transactions on Signal Processing*, vol. 53, no. 1, pp. 44–53, Jan. 2005.
- [49] R. Niu and P. K. Varshney, “Joint detection and localization in sensor networks based on local decisions,” in *Fortieth Asilomar Conference on Signals, Systems and Computers, 2006. ACSSC '06*, 29 2006–Nov. 1 2006, pp. 525–529.
- [50] J. Li and G. AlRegib, “Rate-constrained distributed estimation in wireless sensor networks,” *IEEE Transactions on Signal Processing*, vol. 55, no. 5, pp. 1634–1643, May 2007.
- [51] 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, oct. 2005.
- [52] J.-J. Xiao, A. Ribeiro, Z.-Q. Luo, and G. Giannakis, “Distributed compression-estimation using wireless sensor networks,” *IEEE Signal Processing Magazine*, vol. 23, no. 4, pp. 27–41, july 2006.
- [53] Z.-Q. Luo, G. Giannakis, and S. Zhang, “Optimal linear decentralized estimation in a bandwidth constrained sensor network,” sept. 2005, pp. 1441–1445.
- [54] H. Chen and P. K. Varshney, “Performance limit for distributed estimation systems with identical one-bit quantizers,” *IEEE Transactions on Signal Processing*, vol. 58, no. 1, pp. 466–471, Jan. 2010.
- [55] —, “Nonparametric quantizers for distributed estimation,” *To appear IEEE Transactions on Signal Processing*.
- [56] Y.-W. Hong, K.-U. Lei, and C.-Y. Chi, “Channel-aware random access control for distributed estimation in sensor networks,” *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 2967–2980, July 2008.



- [57] A. Anandkumar, L. Tong, and A. Swami, “Distributed estimation via random access,” *IEEE Transactions on Information Theory*, vol. 54, no. 7, pp. 3175–3181, July 2008.
- [58] E. Masazade, R. Rajagopalan, P. K. Varshney, G. Sendur, and M. Keskinöz, “Evaluation of local decision thresholds for distributed detection in wireless sensor networks using multiobjective optimization,” in *42nd Asilomar Conference on Signals, Systems and Computers*, Oct. 2008, pp. 1958–1962.
- [59] E. Masazade, R. Niu, P. K. Varshney, and M. Keskinöz, “A Monte Carlo based energy efficient source localization method for wireless sensor networks,” in *CAM-SAP’09, IEEE Proc. 3<sup>rd</sup> Workshop on Computational Advances in Multi-Sensor Adaptive Processing*, December 13-16 2009.
- [60] —, “A probabilistic communication scheme for distributed estimation in wireless sensor networks,” in *44th Annual Conference on Information Sciences and Systems, CISS 2009.*, March 2010.
- [61] —, “Channel aware iterative source localization for wireless sensor networks,” in *Accepted to be presented at 13th International Conference on Information Fusion*, July 2010.
- [62] E. Masazade, R. Rajagopalan, P. K. Varshney, C. K. Mohan, G. K. Sendur, and M. Keskinöz, “A multiobjective optimization approach to obtain decision thresholds for distributed detection in wireless sensor networks,” *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 40, no. 2, pp. 444–457, april 2010.
- [63] E. Masazade, R. Niu, P. K. Varshney, and M. Keskinöz, “Energy aware iterative source localization schemes for wireless sensor networks,” *IEEE Transactions on Signal Processing*, vol. Accepted for future publication, 2010.

- [64] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in *Proc. 33rd International Conference on System Sciences*, January 2000.
- [65] R. Niu and P. K. Varshney, "Distributed detection and fusion in a large wireless sensor network of random size," *EURASIP Journal on Wireless Communications and Networking*, vol. 5, no. 4, pp. 462–472, September 2005.
- [66] D. Li, K. Wong, Y. H. Hu, and A. Sayeed, "Detection, classification, and tracking of targets," *IEEE Signal Processing Magazine*, vol. 19, no. 2, pp. 17–29, Mar 2002.
- [67] A. Nasipuri and K. Li, "Multi-sensor collaboration in wireless sensor networks for detection of spatially correlated signals," *International Journal of Mobile Network Design and Innovation*, vol. 1, no. 3/4, pp. 215–223, 2006.
- [68] J.-J. Xiao, S. Cui, Z.-Q. Luo, and A. Goldsmith, "Power scheduling of universal decentralized estimation in sensor networks," *IEEE Transactions on Signal Processing*, vol. 54, no. 2, pp. 413–422, Feb. 2006.
- [69] R. Rajagopalan, C. Mohan, P. Varshney, and K. Mehrotra, "Multi-objective mobile agent routing in wireless sensor networks," in *The IEEE Congress on Evolutionary Computation*, vol. 2, 2005, pp. 1730 – 1737.
- [70] K. Deb and R. B. Agrawal, "Simulated binary crossover for continuous search space," in *Complex Syst.*, vol. 9, pp. 115–148, April 1995.
- [71] M. M. Raghuwanshi and O. G. Kakde, "Survey on multiobjective evolutionary and real coded genetic algorithms," *Complexity International*, vol. 11, pp. 150–161, 2005.
- [72] J. R. Schott, "Fault tolerant design using single and multi-criteria genetic algorithm optimization," May 1995.
- [73] A. Seshadri, "Kanpur genetic algorithms laboratory." [Online]. Available: <http://www.iitk.ac.in/kangal/codes.shtml>

- [74] I. Das, “The normal-boundary intersection home page.” [Online]. Available: <http://softlib.rice.edu/NBI.html>
- [75] P. K. Shukla, K. Deb, and S. Tiwari, “Comparing classical generating methods with an evolutionary multi-objective optimization method,” *Lecture Notes in Computer Science*, vol. 3410/2005, pp. 311–325, 2005.
- [76] M. Nicoli, C. Morelli, and V. Rampa, “A jump markov particle filter for localization of moving terminals in multipath indoor scenarios,” *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 3801–3809, Aug. 2008.
- [77] A. Ihler, I. Fisher, J.W., R. Moses, and A. Willsky, “Nonparametric belief propagation for self-localization of sensor networks,” *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 4, pp. 809–819, April 2005.
- [78] H. Wang, K. Yao, G. Pottie, and D. Estrin, “Entropy-based sensor selection heuristic for target localization,” in *IPSN’04, 3rd International Symposium on Information Processing in Sensor Networks*, April 2004.
- [79] K. Hintz and E. McVey, “Multi-process constrained estimation,” *IEEE Transactions on Systems, Man and Cybernetics*, vol. 21, no. 1, pp. 237–244, Jan/Feb 1991.
- [80] J. Williams, J. Fisher, and A. Willsky, “Approximate dynamic programming for communication-constrained sensor network management,” *IEEE Transactions on Signal Processing*, vol. 55, no. 8, pp. 4300–4311, Aug. 2007.
- [81] G. M. Hoffmann and C. J. Tomlin, “Mobile sensor network control using mutual information methods and particle filters,” *IEEE Transactions on Automatic Control*, vol. 55, no. 1, pp. 32–47, Jan. 2010.
- [82] P. Tichavsky, C. Muravchik, and A. Nehorai, “Posterior Cramér-Rao bounds for discrete-time nonlinear filtering,” *IEEE Transactions on Signal Processing*, vol. 46, no. 5, pp. 1386–1396, May 1998.

- [83] M. Hernandez, T. Kirubarajan, and Y. Bar-Shalom, "Multisensor resource deployment using posterior Cramér-Rao bounds," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 40, no. 2, pp. 399–416, April 2004.
- [84] L. Zuo, R. Niu, and P. Varshney, "Posterior crlb based sensor selection for target tracking in sensor networks," in *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, vol. 2, 15-20 2007, pp. II–1041–II–1044.
- [85] —, "A sensor selection approach for target tracking in sensor networks with quantized measurements," in *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, march 2008, pp. 2521–2524.
- [86] R. Tharmarasa, T. Kirubarajan, M. Hernandez, and A. Sinha, "PCRLB-based multisensor array management for multitarget tracking," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 43, no. 2, pp. 539–555, April 2007.
- [87] M. Hurtado, T. Zhao, and A. Nehorai, "Adaptive polarized waveform design for target tracking based on sequential bayesian inference," *IEEE Transactions on Signal Processing*, vol. 56, no. 3, pp. 1120–1133, March 2008.
- [88] P. Djuric, M. Vemula, and M. Bugallo, "Target tracking by particle filtering in binary sensor networks," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2229–2238, June 2008.
- [89] J. Chou, D. Petrovic, and K. Ramachandran, "A distributed and adaptive signal processing approach to reducing energy consumption in sensor networks," in *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies*, vol. 2, March-3 April 2003, pp. 1054–1062 vol.2.
- [90] Z. Xiong, A. Liveris, and S. Cheng, "Distributed source coding for sensor networks," *IEEE Signal Processing Magazine*, vol. 21, no. 5, pp. 80–94, Sept. 2004.

- [91] M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, “A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking,” *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–188, Feb 2002.
- [92] N. Brunel and J. P. Nadal, “Mutual information, fisher information, and population coding,” *Neural Computation*, MIT Press, 1998.
- [93] E. L. Lehmann, *Theory of point estimation*. John Wiley & Sons, 1983.
- [94] D. Slepian and J. K. Wolf, “Noiseless encoding of correlated information sources,” *IEEE Transactions on Information Theory*, vol. 19, pp. 471–480, July 1973.
- [95] 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.
- [96] R. E. Caflisch, “Monte carlo and quasi-monte carlo methods,” *Acta Numerica*, Cambridge University Press, vol. 7, pp. 1–49, 1988.
- [97] R. D. Gupta and D. Kundu, “On the comparison of Fisher information of the Weibull and GE distributions,” *Journal of Statistical Planning and Inference*, vol. 136, no. 9, pp. 3130 – 3144, 2006.
- [98] W. L. Winston and M. Venkataramanan, *Operations Research Vol 1: Introduction to Mathematical Programming*. CA,USA: Thomson, Brooks/Cole, 2003.
- [99] S. Alhakeem and P. K. Varshney, “A unified approach to the design of decentralized detection systems,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 31.
- [100] L. Liu, X. Zhang, and H. Ma, “Optimal nodes selection for target localization in wireless camera sensor networks,” *IEEE Transactions on Vehicular Technology*, vol. Accepted for future publication, 2009.
- [101] —, “Dynamic node collaboration for mobile target tracking in wireless camera sensor networks,” in *INFOCOM 2009, IEEE*, April 2009, pp. 1188–1196.

# Curriculum Vitae of Engin Maşazade

## Research Interests

- Distributed detection, estimation and localization for wireless sensor networks, multi-objective optimization, statistical signal processing.
- Bit-error rate estimation and cross-layer design for Multiband OFDM based UWB systems,

## Education

- **PhD, Sabancı University, İstanbul, Turkey, 2010,**

Department of Electronics Engineering

Thesis Topic: Resource Aware Distributed Detection and Estimation of Random Events in Wireless Sensor Networks

Advisor: Dr. Mehmet Keskinöz, Sabancı University

Co-advisor: Dr. Pramod K. Varshney, Syracuse University

- **M.Sc, Sabancı University, İstanbul, Turkey, 2006,**

Department of Electronics Engineering

Thesis Topic: Instantaneous bit error rate estimation based link adaptation and scheduling in Multiband OFDM based UWB WPANs

Advisor: Dr. Mehmet Keskinöz

- **B.Sc, İstanbul Technical University, İstanbul, Turkey, 2003,**

Electronics and Communication Engineering (September 1999 - June 2003)

### **Publications**

- *Journal Papers*

- E. Maşazade, R. Niu, P. K. Varshney, M. Keskinöz, “Energy Aware Iterative Source Localization Schemes for Wireless Sensor Networks”, In Press, *IEEE Transactions on Signal Processing*, 2010.
- E. Maşazade, R. Rajagopalan, P. K. Varshney, C. K. Mohan, G. K. Sendur, M. Keskinöz, “A Multiobjective Optimization Approach to Obtain Decision Thresholds for Distributed Detection in Wireless Sensor Networks”, *IEEE Transactions on Systems, Man, and Cybernetics–Part B*, vol.40, no.2, pp.444-457, April 2010.
- E. Maşazade, M. Keskinöz, “An Efficient Realization Bit Error Rate Estimation Method for Multiband OFDM based UWB Systems”, *European Transactions on Telecommunications*, vol. 20, issue 6, pp. 617624, October 2009.
- M. Keskinöz, O. Gurbuz, E. Maşazade, “Cross-Layer Enhanced Time Scheduling for Multi-Band OFDM UWB Networks”, *Wireless Networks*, vol. 16 , issue 3, pp. 863 - 873, April 2010.

- *Conference Proceedings*

- E. Maşazade, R. Niu, P. K. Varshney, M. Keskinöz, “Channel aware iterative source localization for wireless sensor networks” submitted to 13<sup>th</sup> *International Conference on Information Fusion*, Edinburgh, UK, July, 2010.

- E. Maşazade, R. Niu, P. K. Varshney, M. Keskinöz, “A Probabilistic Transmission Scheme for Distributed Estimation in Wireless Sensor Networks”, *IEEE Proc. 44<sup>rd</sup> Conference on Information Sciences and Systems (CISS’10)*, Princeton, NJ, March 2010.
- E. Maşazade, R. Niu, P. K. Varshney, M. Keskinöz, “A Monte Carlo Based Energy Efficient Source Localization Method for Wireless Sensor Networks”, *IEEE Proc. 3<sup>rd</sup> International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, CAMSAP’09*, Aruba, Dutch Antilles, December 13-16 2009.
- E. Maşazade, R. Niu, P. K. Varshney, M. Keskinöz, “An Energy Efficient Iterative Method for Source Localization in Wireless Sensor Networks”, *IEEE Proc. 43<sup>rd</sup> Conference on Information Sciences and Systems (CISS’09)*, Baltimore, MD, March 2009.
- E. Maşazade, R. Rajagopalan, P. K. Varshney, G. K. Sendur, and M. Keskinöz, “Evaluation of local decision thresholds for distributed detection in wireless sensor networks using multi-objective optimization”, *IEEE Proc. 42<sup>nd</sup> Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, October 2008.
- K. Eritmen, E. Maşazade, M. Keskinöz, “A Simplified Parallel Distributed Detection Method for Wireless Sensor Networks under Fading Channels” *IEEE Proc. 15<sup>th</sup> Signal Processing and Communications Applications, SIU*, 11-13 June 2007 Page(s):1 - 4, Eskisehir, Turkey (in Turkish).
- E. Maşazade, M. Keskinöz, O. Gurbuz, “Proportional Time Sharing with Frame Size Adaptation for MB-OFDM based UWB WPANs”, *IEEE Proc.*



*Long Island Systems, Applications and Technology Conference, 2006 (LISAT),*  
5 May 2006 Page(s):1 - 7.

- E. Maşazade, M. Keskinöz, O. Gurbuz; “A Proportional Time Scheduling Scheme with Frame Size Adaptation for MB-OFDM UWB WPANs”, *IEEE Proc. 14<sup>th</sup> Signal Processing and Communications Applications (SIU)*, 2006 17-19 April 2006 Page(s):1 - 4, Antalya, Turkey (in Turkish).

### **Professional Experience**

- Visiting Researcher, Syracuse University, Department of Electrical Engineering and Computer Science, 2008 - 2010.
- Research Assistant, Sabanci University, Department of Electronics Engineering, 2003 - 2008.
- Teaching Assistant, Sabanci University, Department of Electronics Engineering, 2003 - 2008.