

INFORMATION-THEORETIC OPTIMIZATION OF WIRELESS  
SENSOR NETWORKS AND RADAR SYSTEMS

by

Hyoung-soo Kim

---

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

In Partial Fulfillment of the Requirements  
For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2010

THE UNIVERSITY OF ARIZONA  
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Hyoung-soo Kim entitled Information-Theoretic Optimization of Wireless Sensor Networks and Radar Systems and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

Date: 10 September 2010

---

Nathan A. Goodman

Date: 10 September 2010

---

Ali Bilgin

Date: 10 September 2010

---

Ivan Djordjevic

---

---

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

Date: 10 September 2010

---

Dissertation Director: Nathan A. Goodman

### STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: Hyoung-soo Kim

## ACKNOWLEDGEMENTS

In order to achieve a successful PhD, the most important thing is to find a good research adviser and promising research topic. Even though a student is very smart, it is still almost impossible to achieve a successful PhD without a good advisor. On the other hand, although a student is not very smart, it is still very likely for him/her to enjoy a successful PhD with a good research advisor. In this point of view, I must be a very lucky student since I have been researching promising and interesting topics under the guidance of Dr. Nathan A. Goodman. I really appreciate his numerous valuable suggestions and instructions. If it were not for his constant guidance and support, I could not finish my current research works. I wish to express my special thanks to Dr. Shuguang Cui who introduced very interesting research topics and gave essential instructions on how to do research. I wish to give thanks to my fellow students Shikhar Uttam and Junhyeong Bae, who always spare their valuable time for many discussions without hesitation. I also deeply appreciate Dr. Ivan Djordjevic and Dr. Ali Bilgin who have taken their precious time in order to being on my committee during their busy schedule.

## DEDICATION

*To my family, my wife, Mikyong Shim, two daughters, Elizabeth Kim, Hannah Kim, one son who will be born in this winter and my parents, Samyong Kim and Sooyang Jung.*

## TABLE OF CONTENTS

LIST OF FIGURES . . . . .	8
LIST OF TABLES . . . . .	10
ABSTRACT . . . . .	11
CHAPTER 1 INTRODUCTION . . . . .	13
1.1 Power Control Strategy for Distributed Wireless Sensor Network and Its Analysis . . . . .	16
1.2 Waveform Design for Radar System . . . . .	19
1.3 Dissertation Outline . . . . .	21
CHAPTER 2 DETECTION OUTAGE AND DETECTION DIVERSITY IN DIS- TRIBUTED WIRELESS SENSOR NETWORK . . . . .	23
2.1 System Model . . . . .	23
2.2 Uniform Transmission Strategy: Asymptotic $J(\cdot)$ , Detection Outage, and Detection Diversity . . . . .	27
2.2.1 Asymptotic J-divergence $J_\infty$ for large-scale networks . . . . .	29
2.2.2 Detection Outage and Detection diversity . . . . .	31
2.3 Optimal-power Transmission Strategy: Adaptive Power Gain and Full Diversity Gain . . . . .	34
2.3.1 Optimal-power transmission with a total power constraint . . . . .	34
2.3.2 Optimal-power transmission with both individual and total power constraints . . . . .	36
CHAPTER 3 POWER CONTROL STRATEGY FOR DISTRIBUTED MULTIPLE HYPOTHESIS DETECTION . . . . .	38
3.1 System Model . . . . .	38
3.2 Power Control Strategy . . . . .	42
3.2.1 Local Power Control . . . . .	42
3.2.2 System Performance Measure, Total J-divergence & Channel Power Control . . . . .	43
3.3 Classification Outage of a Uniform Transmit Strategy with Local Power Optimization . . . . .	47
3.3.1 Uniform Transmit Strategy . . . . .	47
3.3.2 Asymptotic total J-divergence $J_\infty$ and Classification Outage . . . . .	49

TABLE OF CONTENTS – *Continued*

3.3.3	Long-Term System Performance Measure . . . . .	52
3.3.4	Total J-divergence with $\mathcal{D}$ -dimensional Local Modulation . . . . .	55
3.4	Local-Channel Power Transmission Strategy . . . . .	59
3.4.1	Optimization with a Total Power Constraint . . . . .	59
3.4.2	Optimization with Individual and Total Power Constraints . . . . .	63
3.4.3	Power distribution across sensors . . . . .	64
3.5	Conclusion of Power Optimization for Sensor Network . . . . .	65
 CHAPTER 4 WAVEFORM DESIGN WITH TASK-SPECIFIC INFORMATION		 67
4.1	System Model . . . . .	67
4.2	Waveform design with Task-Specific Information for Monostatic Radar . . . . .	69
4.2.1	$\mathcal{L}$ -dimensional spherical coordinate . . . . .	71
4.2.2	Global Optimality of TSI Waveform Design . . . . .	73
4.3	Waveform design with Task-Specific Information for MIMO Radar . . . . .	76
4.4	Simulation Result . . . . .	81
4.4.1	Energy Distribution in Waveform . . . . .	85
4.5	Conclusion of Radar Waveform Design . . . . .	86
 CHAPTER 5 FINAL CONCLUSIONS AND FUTURE WORKS		 89
 APPENDIX A DERIVATIONS AND PROOFS FOR CHAPTER 2		 92
A.1	Derivation of (2.3) . . . . .	92
A.2	Derivation of (2.6) . . . . .	93
A.3	Derivation of the two conditions in Theorem 3.1 . . . . .	93
A.4	Proof of Theorem 3.2 . . . . .	94
 APPENDIX B DERIVATIONS AND PROOFS FOR CHAPTER 3		 96
B.1	The Relationship between Bound Probability of Error and Total J-divergence . . . . .	96
B.2	Derivation of (3.6) . . . . .	97
B.3	Derivation of Case 1, (3.9) . . . . .	100
B.4	Derivation of Case 2, (3.11) . . . . .	101
B.5	Derivation of Nondecreasing Characteristic of Total J-Divergence . . . . .	102
B.6	Concave Condition of Total J-Divergence . . . . .	103
 APPENDIX C DERIVATIONS FOR CHAPTER 4		 104
C.1	Waveform design for monostatic radar with TSI . . . . .	104
C.2	Waveform design for SIMO radar with TSI . . . . .	111
 REFERENCES		 118

## LIST OF FIGURES

1.1	Outage Probability( $P_{J_0}$ ) vs. Average measure( $E\{J(\cdot)\}$ ) . . . . .	18
2.1	Distributed detection with a fusion center. . . . .	23
2.2	Outage probability versus $P_{tot}$ with the uniform transmission strategy. . .	33
2.3	Average J-divergence versus $P_{tot}$ with the uniform transmission strategy. .	34
2.4	Outage probabilities for the power-optimal transmission strategy and the uniform transmission strategy. . . . .	35
2.5	Outage probability versus $P_{tot}$ for different scenarios. . . . .	37
3.1	Distributed Multiple Hypothesis Detection with a Fusion Center. . . . .	39
3.2	Outage probability versus $P_{tot}$ with the uniform channel power transmission strategy in heterogeneous sensor configuration. "Local" in parenthesis means that Local power control is performed . . . . .	56
3.3	Outage probability versus $P_{tot}$ with the uniform channel power transmission, 2D-local modulation, and heterogeneous sensor configuration . . . .	58
3.4	Outage probability versus $P_{tot}$ . local and channel power control strategies are applied in a six-node system. . . . .	62
3.5	Outage probability of an optimal system with individual power constraints and a total power constraint in a nine-node system. . . . .	64
4.1	Plot of TSI on the feasible energy region for two tap waveform . . . . .	75
4.2	Plot of TSI on the feasible energy region for three tap waveform . . . . .	75
4.3	Performance comparison of various waveform design methods in maximum likelihood detection in case of 5 tap waveform and non-colored target sets . . . . .	82
4.4	Performance comparison of various waveform design methods in maximum likelihood detection in case of 10 tap waveform and non-colored target sets . . . . .	82
4.5	Performance comparison of various waveform design methods in maximum likelihood detection in case of 10 tap waveform with colored target set I . . . . .	83
4.6	Performance comparison of various waveform design methods in maximum likelihood detection in case of 10 tap waveform with colored target set II . . . . .	83
4.7	Performance comparison of various waveform design methods in maximum likelihood detection in case of 10 tap waveform with colored target set I for MIMO radar system of three radars . . . . .	84

LIST OF FIGURES – *Continued*

4.8	Performance comparison of various waveform design methods in maximum likelihood detection in case of 10 tap waveform with colored target set II for MIMO radar system of three radars . . . . .	84
4.9	Diversity gain by MIMO radar configuration using TSI waveform design in case of 10 tap waveform with colored target set II . . . . .	85
4.10	Target Magnitude Spectrums used for the Comparison of Waveform Energy Distribution . . . . .	87

## LIST OF TABLES

3.1	Power distribution by local power control in percentage of total power allocated to each sensor . . . . .	65
4.1	Comparison of Waveform Energy Distribution . . . . .	88

## ABSTRACT

Three information measures are discussed and used as objective functions for optimization of wireless sensor networks (WSNs) and radar systems. In addition, a long-term system performance measure is developed for evaluating the performance of slow-fading WSNs. Three system applications are considered: a distributed detection system, a distributed multiple hypothesis system, and a radar target recognition system.

First, we consider sensor power optimization for distributed binary detection systems. The system communicates over slow-fading orthogonal multiple access channels. In earlier work, it was demonstrated that system performance could be improved by adjusting transmit power to maximize the J-divergence measure of a binary detection system [1]. We define outage probability for slow-fading system as a long-term performance measure, and analytically develop the detection outage with the given system model. Based on the analytical result of the outage probability, diversity gain is derived and shown to be proportional to the number of the sensor nodes. Then, we extend the optimized power control strategy to a distributed multiple hypothesis system, and enhance the power optimization by exploiting a priori probabilities and local sensor statistics. We also extend outage probability to the distributed multiple-hypotheses problem. The third application is radar waveform design with a new performance measure: Task-Specific Information (TSI). TSI is an information-theoretic measure formulated for one or more specific sensor tasks by encoding the task(s) directly into the signal model via source variables. For example, we consider the problem of correctly classifying a linear system from a set of known alternatives, and the source variable takes the form of an indicator vector that

selects the transfer function of the true hypothesis. We then compare the performance of TSI with conventional waveforms and other information-theoretic waveform designs via simulation. We apply radar-specific constraints and signal models to the waveform optimization.

## CHAPTER 1

### INTRODUCTION

In this dissertation, we show three system optimization problems with three different information measures. Two applications are about transmit power optimization for wireless sensor networks (WSNs) and the other is radar waveform optimization. These systems are characterized by various system statistics, and design parameters are optimized based on the system statistics. For WSNs, the design parameters are the transmit sensor powers, and for the radar system, the design is over the spectral content of the transmit waveform.

After briefly introducing the practical systems and explaining the motivation of each system optimization, we will discuss what types of performance measures can be used as objective functions for system optimization, why a performance measure based on an information measure is preferred among the various kinds of performance measures, and how we can select/design an information measure specialized for the detection/classification problem.

Strategies to improve the lifetime of battery-powered WSNs have been an intensively studied topic due to the difficulty of replacing batteries in geographically deployed sensors. We consider distributed detection/classification systems where a network of sensors each observe the event status of a source, make their own local decision, and forward the decision to a fusion center through a non-ideal fading channel. Based on the local decisions, the fusion center makes a final decision according to a fusion rule. However, the focus here is not on the local decision or the fusion rules, but on efficient power control strategies for conserving the power used by WSNs. Based on the given WSN system model and the proper information measure, the transmit power of sensors can be

optimized.

In a radar system, target information is obtained by transmitting a waveform and analyzing the received observation. Linear FM is traditionally used as a radar waveform since it provides a good range resolution, high pulse-compression ratio, and a simple hardware implementation. The maximum detection range is improved by using a long pulse width carrying high energy without sacrificing bandwidth. Recently, transmit waveforms can be easily and digitally manipulated due to the advent of digital radar systems with a fast processor. This means that software-defined radar systems can show remarkable flexibility and versatility in capturing the information of targets. In this case, the waveform does not have to be a Linear FM. Instead, we can use any optimized waveform based on a priori information on the given system model including target signatures in order to obtain more reliable information about the target. Therefore, optimal waveform designs are recently getting more attention [2][3][4][5][6][7][8]. Based on the given radar system model and the proper information measure, the transmit waveform can be optimized.

In order to find a proper optimization scheme for these applications, we need to investigate a reasonable objective function characterizing the given system models. Traditionally, signal-to-noise ratio (SNR), and mean-square error (MSE) have been successfully used since these performance measures are intuitive and work well. However, SNR and MSE are based on only the second moment of a statistical process while information measures are based on a complete probability distribution function. Therefore, more researches based on information-theoretic optimization method are active now in various science and engineering fields.

Note that it is almost impossible to confirm that one specific performance measure is the best measure in all cases since the system performance depends highly on the statistics of the input data. However, in many cases, information measures are bet-

ter than traditional measures and have been adopted for many system optimizations [1][2][8][9][10][11][12]. Therefore, after selecting and designing information measures based on system behaviors, we perform system optimizations and analyze the performances expecting that the information measures are statistically more suited for characterizing the given system.

We use three different information measures: J-divergence, total J-divergence, and Task-Specific Information (TSI). All of these information measures depend on the statistics of the observations and target classes (or target hypotheses). For example, J-divergence is a symmetric Kullback-Leibler distance measure between two conditional distributions,  $p(\mathbf{y}|H_0)$  and  $p(\mathbf{y}|H_1)$ , for binary detection problem. Each distribution of the observation  $\mathbf{y}$  is conditioned by each target class,  $H_0$  or  $H_1$ . Therefore, J-divergence is related with the observations and target classes. For multiple-hypothesis detection problems, we use total J-divergence and TSI, both of which exploit a priori probabilities of the target or source. Total J-divergence is a pair-wise weighted sum of the individual J-divergences while TSI is a mutual information between the observation and a virtual source. The virtual source (or source variable) is a direct encoding of the classification task into the signal model via a discrete random variable. The number of potential realizations of the discrete source variable is the same as the number of hypotheses and the a priori probabilities of the source variable are the same as for the classification task. Therefore, identifying the source variable is equivalent to making a classification decision. Total J-divergence has the limitation that only the pair-wise distances among the multiple-target classes are considered while TSI has the advantage that the overall divergence or dispersion among the multiple-target classes is fully considered by directly estimating the target classification task or virtual source. By maximizing these three information measures, we can capture enhanced source information (target classes or hypotheses) from

the observations. We will discuss the detailed definitions of these information measures and give some explanation/simulation results on the relationship between the information measures and probability of error later.

Now, we briefly introduce the previous researches related with WSNs and optimized radar waveform design. First, Section 1.1 gives a brief introduction to the previous researches on WSNs. Then, Section 1.2 describes previous research on radar waveform design.

### 1.1 Power Control Strategy for Distributed Wireless Sensor Network and Its Analysis

Research on local decision and fusion rules for distributed sensor networks can be traced back to the early 1980s ([13], [14], [15]) and are still in process ([16], [17], [18], [19]), but studies of optimal power control strategies are only recently being explored ([1], [20], [21], [22]). Thomopoulos [23] et al showed that the local likelihood ratio test (LRT) is optimal in the presence of a fusion center under the Neyman-Pearson criterion where the decision scheme maximizes the probability of detection for fixed probability of false alarm. These decision rules were then evaluated and improved in the presence of network delay and channel errors in [24]. On the other hand, Hoballah and Varshney investigated the optimal local rule under the Bayesian criterion in [15], and the rule was expanded to consider the presence of non-ideal channels in [18].

Sadjadi developed local decision logic and fusion rules for generalizing the distributed detection problem of Tenney to distributed multiple hypothesis detection[25]. He obtained optimal detectors by minimizing an average cost function and finding the optimal decision regions. Chao showed optimum partitionings of decision regions based on multi-bit decisions to avoid the information loss by local hardlimiting process in [26]. In [27], [28], Wang studied local sensor decision logic and sensor fusion with a double-

bound testing method for both serial distributed sensor networks and parallel decision networks. Oh introduced a multiple-target tracking algorithm with a dynamic model of multiple targets for sensor networks [29]. However, power optimization strategies unique to multiple hypothesis scenarios have not been explored.

Recently, the emerging optimal power allocation issue has been considered in the context of estimation of an unknown parameter or detection of an unknown source. Xiao et al [21] introduced optimal power scheduling for the joint estimation of a Gaussian source in an inhomogeneous Gaussian sensor network by minimizing total power consumption while satisfying a mean-squared distortion constraint. Zhang [22] introduced an optimal power allocation scheme over a multiple access channel by maximizing J-divergence under a fixed total communication power constraint on the sensors of a distributed binary detection system. We extend the optimal power allocation for distributed binary hypothesis system to distributed multiple-hypothesis system and apply an additional local power optimization based on local sensor statistics to obtain an extra local power gain. In addition, we define and develop outage probability for slow-fading WSNs. Outage probability is a long term system performance measure to evaluate the performance of binary and multiple hypothesis sensor networks.

In Chapter 2, J-divergence is used as an objective function to optimize the power of the individual sensor. Here, the objective function is an instantaneous performance measure since it quantifies the system performance with J-divergence calculated based on instantaneous system states. In general, system performance can be enhanced if more system resources are applied. For a simple communication problem, if transmission power is increased, better estimation/detection can be achieved at a receiver since the system SNR increases. Next, two new concepts called detection outage and detection diversity are introduced to quantify the long-term system performance. Detection outage is defined as

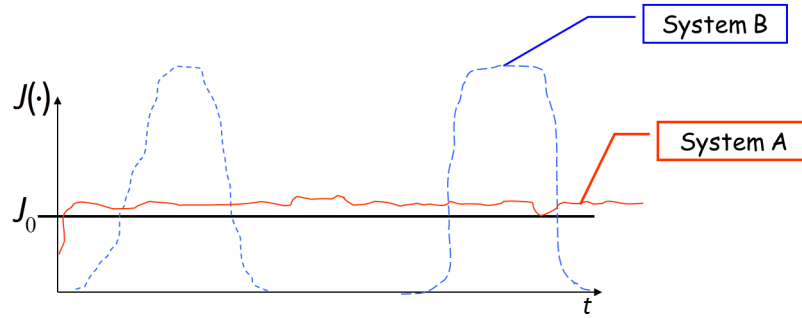


Figure 1.1: Outage Probability( $P_{J_0}$ ) vs. Average measure( $E\{J(\cdot)\}$ )

the probability that instantaneous J-divergence is smaller than a certain threshold. For a slow-fading system, outage probability is a more useful measure than average measure. The reason can be explained by Figure 1.1. Systems A and B have the same average J-divergence since areas under the J-divergence curves of systems A and B are same. However, in the viewpoint of outage probability, system A is almost everywhere above the threshold( $J_0$ ) while system B is below the threshold for more than 50 percent of the time. Therefore, system A is usually active while system B is in outage for more than 50 percent of time. System A is practically stable and manages efficiently the given system resource while system B is wasting the system resource. Therefore, outage probability can be a more meaningful performance measure than average J-divergence.

Based on the detection outage probability, detection diversity order is defined as the slope of the outage probability curve (versus system power consumption) when outage probability curves are plotted in logarithmic scale. We show that even with a simple uniform transmission strategy, full detection diversity can be achieved and is proportional to the total number of sensors in the network. However, if the transmission power is optimized across sensors, an adaptive power gain can be achieved in addition to the full diversity gain.

In Chapter 3, we introduce the concept of *local power control*, which is performed by

minimizing the average power of the local  $\mathcal{D}$ -dimensional transmit symbol constellation for distributed multiple hypothesis detection. Then *channel power control* is performed to maximize the total J-divergence, which is an instantaneous performance measure based on the instantaneous channel realization in distributed multiple-hypothesis detection system. The total J-divergence optimization results in optimized amplifying factors prior to transmission over the channel. We analyze the combined *local-channel power control scheme* using classification outage, which is redefined for the multiple hypothesis system.

## 1.2 Waveform Design for Radar System

The problem of radar waveform design for improved detection, parameter estimation, and/or classification has been an intensively studied topic. One result that has been explored recently is the one by Bell, who developed information-theoretic principles whereby waveform optimization is achieved by maximizing the mutual information between a Gaussian ensemble of extended targets and the received signal [2]. This approach yields a spectral waterfilling (WF) strategy for obtaining the waveform's energy spectrum. In other work, the WF strategy by Bell was applied to multiple-target parameter estimation by defining a weighted linear sum of mutual information terms in [7]. In [30], the Kullback-Leibler(KL) distance measure is used for waveform optimization. In [31], waveforms for target classification were optimized in the time domain for multiple-input, multiple-output (MIMO) radars by minimizing MMSE or maximizing the mutual information between the observation and the target impulse response. In [3][4][5][6][8], waveforms for target classification were designed according to SNR considerations, which led to an eigensolution based on an overall target difference correlation matrix[32]. Waveform design based directly on detection considerations has been studied in [33]. In [8], the approach in [2] was applied to the multiple target classification problem by defining a

probability-weighted spectral variance across the transfer functions of the various target hypotheses and using this spectral variance in the waterfilling operation. To the extent that accurately estimating the target transfer function yields good classification performance, these approaches yield improved performance. However, we are motivated to find a more effective and direct information measure optimized for the specific task of target classification. This measure should be directly related to target class identification, not necessarily to estimating target profiles that are then used in a classification algorithm.

The concept of task-specific information (TSI) was first proposed as an analysis tool in [34][35] where it was used to optimize features that would be extracted for the final task of recognition in imaging systems. We now develop and apply TSI for active waveform design with the goal of discriminating between known target hypotheses. We introduce a virtual source variable  $\mathbf{x}$ , which is in the form of an indicator vector that selects the true target transfer function from among the ensemble of target hypotheses. The TSI formulation maximizes the mutual information between the received measurements and the source variable  $\mathbf{x}$  rather than the measurements and the transfer functions. If the frequency-domain vector of observations is  $\mathbf{y}$ , then TSI is defined as  $I(\mathbf{y}; \mathbf{x}) = h(\mathbf{x}) - h(\mathbf{x}|\mathbf{y})$  where  $h(\cdot)$  is entropy. For example, if the goal is to correctly identify the target from among four equally likely hypotheses, then the maximum task-specific information that can be extracted is  $h(\mathbf{x}) = 2$  bits, which is consistent with the four-class recognition problem. Waveform design with TSI does not assume the frequency components of an observation to be independent, which is a necessary assumption for using the waterfilling solution from [2]. The fact that multiple observations are correlated when parameterized by an underlying discrete source variable is accommodated by the TSI design framework and leads to additional performance improvement.

### 1.3 Dissertation Outline

The rest of this dissertation has three main chapters. Chapters 2 and 3 treat the problem of power control for WSNs, and Chapter 4 treats waveform design for radar systems performing target recognition. The overall structure of each chapter is organized as follows. After introducing a system model, we apply a given information measure to the system model and used the measure as an objective function for optimization problem. Then, the system is optimized by a numerical approach and the optimization performance is compared with the performances of uniform power/energy scheme or a conventional optimization scheme. In order to compare the system performances, outage probability is defined/developed for slow-fading WSN while probability of error is numerically generated for radar system.

In Chapter 2, the system model of the detection problem for distributed wireless sensor network is introduced, and we develop the J-divergence measure based on the given system model in Section 2.1. In Section 2.2, the concepts of detection outage and detection diversity are introduced with the assumption of a uniform transmission strategy across different sensors. In Section 2.3, we optimize the power transmission strategy subject to total and/or individual sensor power constraints.

In Chapter 3, we extend the detection problem of Chapter 2 to a multiple-hypotheses detection problem by adopting the pair-wise sum of J-divergence measure. At first, the system model of distributed classification in WSNs is explained in Section 3.1. In Section 3.2, the local power optimization is explained and the total J-divergence measure is derived. In Section 3.3, after briefly explaining a uniform power transmission scheme with local power optimization, we derive the asymptotic total J-divergence that we use to generalize detection outage probability for the multiple hypothesis problem. The relationship between asymptotic total J-divergence and classification outage(or multiple hypothesis

detection outage) is also demonstrated. Using the extended outage probability, we analyze the power gain and classification diversity achieved through local power control. In Section 3.4, channel power optimization (symbol amplification) is combined with the local power optimization procedure of Section 3.3. Classification outage performance is evaluated under a) a total power constraint, and b) total and individual power constraints. Section 3.5 summarizes the results of Chapters 2 and 3, and gives some concluding remarks on the power optimization strategies for WSNs.

Chapter 4 begins by describing the system model for radar target recognition, and the concept of TSI is explained by defining a virtual source that encodes the target classification task into the signal model. Then, the optimization method with TSI is introduced. In Section 4.2, we develop the gradient of TSI with respect to waveform frequency coefficients for a monostatic radar configuration and briefly evaluate the global optimality of the gradient-based optimization process through Monte Carlo simulations. In Section 4.3, we introduce TSI for MIMO radar configurations. In Section 4.4, the performances of the proposed TSI-based waveform and other waveforms are compared through computer simulation results. Conclusions for the waveform design problem are made in Section 4.5.

Final conclusions and some suggestions for future work are provided in Chapter 5.

## CHAPTER 2

DETECTION OUTAGE AND DETECTION DIVERSITY IN DISTRIBUTED  
WIRELESS SENSOR NETWORK

In this chapter, we discuss power control strategy and detection outage for binary hypothesis system by using J-divergence which is a symmetric distance measure defined by adding two Kullback-Leibler distance measure.

## 2.1 System Model

We adopt a similar model to that in [22], which is described as follows. At each observation period, the detection target can be abstracted as two hypotheses:  $H_0$  for the absence of the target and  $H_1$  for the existence of the target, where we have  $\mathcal{K}$  sensors that make local observations and send local decisions to the fusion center. For simplicity, here we focus on orthogonal multiple access channels between the sensors and the fusion center.

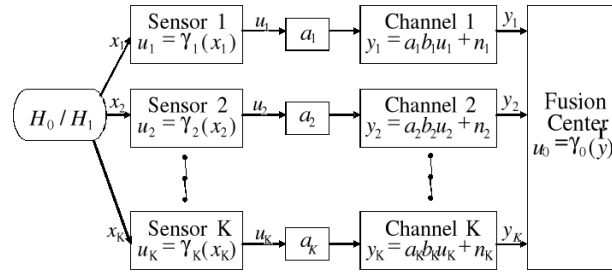


Figure 2.1: Distributed detection with a fusion center.

Specifically as shown in Fig. 2.1, sensor  $k$  collects a local observation  $x_k$  that may be

corrupted by an observation noise; then makes its local decision according to the rule:

$$u_k = \gamma_k(x_k) = \begin{cases} 0 & \text{under } H_0 \\ 1 & \text{under } H_1 \end{cases}$$

with  $k = 1, \dots, \mathcal{K}$ . Afterwards, node  $k$  amplifies the local decision result by a factor of  $a_k$  and transmits the signal over the  $k$ th channel, which is modeled by a random amplitude gain  $b_k$  (assuming a coherent reception at the fusion center) and an additive white Gaussian noise (AWGN)  $n_k$ . We assume that both  $b_k$  and  $n_k$  are independently and identically distributed (i.i.d.) over  $k$ , but remain constant within the  $k$ th observation period. Note that Fig. 2.1 only shows one time snapshot (one observation period) of the system. Since we assume that the source and the channel state are also i.i.d. over time, we could first focus on just one snapshot and later apply the result to analyzing long-term performance (for which we will discuss average and outage performances, respectively).

For sensor  $k$ , the false alarm probability  $P_F(k)$  and the detection probability  $P_D(k)$  are given as

$$P_F(k) = p(u_k = 1|H_0),$$

$$P_D(k) = p(u_k = 1|H_1).$$

We further assume that  $x_k$ 's are independent over  $k$  conditional on a particular hypothesis. As such, the joint conditional probability mass functions of the local decisions are as

follows [22],

$$P(\mathbf{u}|H_0) = \prod_{k=1}^{\mathcal{K}} P_F(k)^{u_k} (1 - P_F(k))^{(1-u_k)},$$

$$P(\mathbf{u}|H_1) = \prod_{k=1}^{\mathcal{K}} P_D(k)^{u_k} (1 - P_D(k))^{(1-u_k)},$$

where  $\mathbf{u} = [u_1, \dots, u_{\mathcal{K}}]^T$ .

At the fusion center, the received signals are represented as

$$\mathbf{y} = \mathbf{B}\mathbf{A}\mathbf{u} + \mathbf{n}$$

where  $\mathbf{y} = [y_1, \dots, y_{\mathcal{K}}]^T$ ,  $\mathbf{A} = \text{diag}\{a_1, a_2, \dots, a_{\mathcal{K}}\}$ ,  $\mathbf{B} = \text{diag}\{b_1, b_2, \dots, b_{\mathcal{K}}\}$ , and  $\mathbf{n} = \{n_1, \dots, n_{\mathcal{K}}\}$ . As such, the conditional probability density function of the received signals given the true hypotheses are [22]

$$p(\mathbf{y}|H_i) = \sum_{\mathbf{u}} p(\mathbf{y}|\mathbf{u})p(\mathbf{u}|H_i), \quad i = 0, 1. \quad (2.1)$$

The final decision at the fusion center,  $u_0$ , is determined by  $u_0 = \gamma_0(\mathbf{y})$  with  $\gamma_0(\cdot)$  the fusion rule.

As in [22], here we are not focusing on the design of the optimal fusion rule. Instead, we argue that the detection performance is optimized if the J-divergence  $J(p(\mathbf{y}|H_0), p(\mathbf{y}|H_1))$  is maximized [22]. Particularly,

$$J(p(\mathbf{y}|H_0), p(\mathbf{y}|H_1)) = D(p(\mathbf{y}|H_1)||p(\mathbf{y}|H_0))$$

$$+ D(p(\mathbf{y}|H_0)||p(\mathbf{y}|H_1))$$

where  $D(p_1||p_2)$  is Kullback-Leibler distance measure between two probability density

functions [36]. For the detection performance measure, there are usually three different criteria to adopt: the detection error probability (or probabilities of detection and false alarm for Nearman-Pearson detectors when *a priori* information is unknown), distance related bounds (such as J-divergence [37]), and the asymptotic relative efficiency (a.k.a. ARE [38]). Although the J-divergence is not being popularly used as the other two criteria, as pointed out in [22], the J-divergence usually leads to tractable analytical frameworks for distributed detection systems and has been adopted by many researchers, e.g., in [37, 39, 40, 38]. In addition, it has been established in [37] that the detection error probability  $p_e$  is related to the J-divergence via the following inequality:

$$p_e > p(H_0)p(H_1)e^{-J/2}.$$

As such, the J-divergence provides us a lower bound to the detection error probability and this bound monotonically decreases as  $J$  increases. In addition, the author of [38] has proved that there are fundamental connections between the J-divergence and the ARE. Furthermore, as established in [40], the J-divergence and Bhattacharyya bound both belong to a more general class of distance measures: the Ali-Silvey class of distance measures; meanwhile, the J-divergence is a symmetric version of the Kullback-Leibler distance, which is asymptotically the error exponent of the Chernoff bound from the Stein's lemma [36].

As discussed in [22], with AWGN channel noises, the system conditional probabilities in (2.1) can be approximated by Gaussian densities and the resulting J-divergence is given as

$$\begin{aligned} J(\cdot) &= \frac{1}{2}Tr \left[ [\mathbf{R} + \mathbf{BA}(\mathbf{C}_0 + \boldsymbol{\xi}\boldsymbol{\xi}^T)\mathbf{A}^T\mathbf{B}^T] \cdot [\mathbf{R} + \mathbf{BAC}_1\mathbf{A}^T\mathbf{B}^T]^{-1} \right] \\ &+ \frac{1}{2}Tr \left[ [\mathbf{R} + \mathbf{BA}(\mathbf{C}_1 + \boldsymbol{\xi}\boldsymbol{\xi}^T)\mathbf{A}^T\mathbf{B}^T] \cdot [\mathbf{R} + \mathbf{BAC}_0\mathbf{A}^T\mathbf{B}^T]^{-1} \right] - \mathcal{K}, \quad (2.2) \end{aligned}$$

where

$$\begin{aligned}\boldsymbol{\xi} &= \boldsymbol{\xi}_0 - \boldsymbol{\xi}_1, \\ \boldsymbol{\xi}_0 &= [P_D(1), \dots, P_D(\mathcal{K})]^T, \\ \boldsymbol{\xi}_1 &= [P_F(1), \dots, P_F(\mathcal{K})]^T, \\ \mathbf{C}_1 &= \text{diag}\{P_D(1) - P_D^2(1), \dots, P_D(\mathcal{K}) - P_D^2(\mathcal{K})\}, \\ \mathbf{C}_0 &= \text{diag}\{P_F(1) - P_F^2(1), \dots, P_F(\mathcal{K}) - P_F^2(\mathcal{K})\}.\end{aligned}$$

$Tr[\cdot]$  stands for the trace operation and  $\mathbf{R}$  is covariance matrix of the channel noise vector  $\mathbf{n}$ .

## 2.2 Uniform Transmission Strategy: Asymptotic $J(\cdot)$ , Detection Outage, and Detection Diversity

In this section, we focus on the simplest transmission scheme: uniform transmissions, where  $a_k$ 's are the same. For the convenience of notation, we define  $\sum_{k=1}^{\mathcal{K}} a_k^2 = P_{tot}$  such that  $a_k^2 = \frac{P_{tot}}{\mathcal{K}}, \forall k$ . Note that we keep  $P_{tot}$  fixed as we are investigating how the system performance scales with  $\mathcal{K}$ , for the sake of comparison fairness. In addition, we set  $\mathbf{R} = \text{diag}\{\sigma_1^2, \dots, \sigma_{\mathcal{K}}^2\}$  and we denote the channel SNR as  $s_k = b_k^2/\sigma_k^2, \forall k$ . As such,

the J-divergence in (2.2) becomes

$$\begin{aligned}
J(\cdot) &= \frac{1}{2} \sum_{k=1}^{\mathcal{K}} \frac{\sigma_k^2 + a_k^2 b_k^2 [P_F(k) + P_D(k)^2 - 2P_D(k)P_F(k)]}{\sigma_k^2 + a_k^2 b_k^2 [P_D(k) - P_D(k)^2]} \\
&+ \frac{1}{2} \sum_{k=1}^{\mathcal{K}} \frac{\sigma_k^2 + a_k^2 b_k^2 [P_D(k) + P_F(k)^2 - 2P_D(k)P_F(k)]}{\sigma_k^2 + a_k^2 b_k^2 [P_F(k) - P_F(k)^2]} - \mathcal{K} \\
&= \frac{1}{2} \sum_{k=1}^{\mathcal{K}} \frac{\sigma_k^2 + \frac{P_{tot}}{\mathcal{K}} b_k^2 [P_F(k) + P_D(k)^2 - 2P_D(k)P_F(k)]}{\sigma_k^2 + \frac{P_{tot}}{\mathcal{K}} b_k^2 [P_D(k) - P_D(k)^2]} \\
&+ \frac{1}{2} \sum_{k=1}^{\mathcal{K}} \frac{\sigma_k^2 + \frac{P_{tot}}{\mathcal{K}} b_k^2 [P_D(k) + P_F(k)^2 - 2P_D(k)P_F(k)]}{\sigma_k^2 + \frac{P_{tot}}{\mathcal{K}} b_k^2 [P_F(k) - P_F(k)^2]} - \mathcal{K} \\
&= \frac{1}{2} \left[ \sum_{k=1}^{\mathcal{K}} \left( \frac{\mathcal{K} + P_{tot} s_k \alpha_F(k)}{\mathcal{K} + P_{tot} s_k \beta_F(k)} - 1 \right) \right] + \frac{1}{2} \left[ \sum_{k=1}^{\mathcal{K}} \left( \frac{\mathcal{K} + P_{tot} s_k \alpha_D(k)}{\mathcal{K} + P_{tot} s_k \beta_D(k)} - 1 \right) \right] \\
&= \frac{1}{2} \left[ \sum_{k=1}^{\mathcal{K}} \frac{P_{tot} s_k [\alpha_F(k) - \beta_F(k)]}{\mathcal{K} + P_{tot} s_k \beta_F(k)} \right] + \frac{1}{2} \left[ \sum_{k=1}^{\mathcal{K}} \frac{P_{tot} s_k [\alpha_D(k) - \beta_D(k)]}{\mathcal{K} + P_{tot} s_k \beta_D(k)} \right] \\
&= \frac{1}{2} (J_1(\cdot) + J_2(\cdot))
\end{aligned}$$

where the system constants (which are assumed to be known)

$$\alpha_F(k) = P_F(k) + P_D(k)^2 - 2P_D(k)P_F(k),$$

$$\beta_F(k) = P_D(k) - P_D(k)^2,$$

$$\alpha_D(k) = P_D(k) + P_F(k)^2 - 2P_D(k)P_F(k),$$

$$\beta_D(k) = P_F(k) - P_F(k)^2,$$

and

$$J_1(\cdot) = \sum_{k=1}^{\mathcal{K}} \frac{P_{tot} s_k [\alpha_F(k) - \beta_F(k)]}{\mathcal{K} + P_{tot} s_k \beta_F(k)}, \quad J_2(\cdot) = \sum_{k=1}^{\mathcal{K}} \frac{P_{tot} s_k [\alpha_D(k) - \beta_D(k)]}{\mathcal{K} + P_{tot} s_k \beta_D(k)}.$$

### 2.2.1 Asymptotic J-divergence $J_\infty$ for large-scale networks

To gain some intuitions, let us first investigate the asymptotic performance when the number of sensors gets large. We first work on  $J_1(\cdot)$ , i.e., the first part of  $J(\cdot)$ .

**Case 1:**  $0 < \alpha_F(k) - \beta_F(k) < 1$ . It can be shown in Appendix A.1 that the following inequalities hold:

$$\begin{aligned} \sum_{k=1}^{\mathcal{K}} \frac{P_{tot} s_k [\alpha_F(k) - \beta_F(k)]}{\mathcal{K}} - \sum_{k=1}^{\mathcal{K}} \frac{P_{tot}^2 s_k^2 \beta_F(k) [\alpha_F(k) - \beta_F(k)]}{\mathcal{K}^2} \\ \leq J_1(\cdot) \leq \sum_{k=1}^{\mathcal{K}} \frac{P_{tot} s_k [\alpha_F(k) - \beta_F(k)]}{\mathcal{K}}. \end{aligned} \quad (2.3)$$

In the case of a homogeneous sensor network, i.e.,  $\alpha_F(k)$ ,  $\alpha_D(k)$ ,  $\beta_F(k)$ , and  $\beta_D(k)$  are the same over  $k$  while  $s_k$  being i.i.d. over  $k$ , according to the strong Law of Large Numbers (LLN), we obtain the following limits as  $\mathcal{K} \rightarrow \infty$ :

$$\begin{aligned} \sum_{k=1}^{\mathcal{K}} \frac{P_{tot} s_k [\alpha_F(k) - \beta_F(k)]}{\mathcal{K}} &\rightarrow P_{tot} E[s_k (\alpha_F(k) - \beta_F(k))], \\ \sum_{k=1}^{\mathcal{K}} \frac{P_{tot}^2 s_k^2 \beta_F(k) [\alpha_F(k) - \beta_F(k)]}{\mathcal{K}} &\rightarrow P_{tot}^2 E[s_k^2 \beta_F(k) (\alpha_F(k) - \beta_F(k))] \\ \text{and } \sum_{k=1}^{\mathcal{K}} \frac{P_{tot}^2 s_k^2 \beta_F(k) [\alpha_F(k) - \beta_F(k)]}{\mathcal{K}^2} &\rightarrow 0, \end{aligned} \quad (2.4)$$

since  $E[s_k (\alpha_F(k) - \beta_F(k))]$  and  $E[s_k^2 \beta_F(k) (\alpha_F(k) - \beta_F(k))]$  are finite in practical cases.

Therefore, with  $0 < \alpha_F(k) - \beta_F(k) < 1$ , we have

$$\lim_{\mathcal{K} \rightarrow \infty} J_1(\cdot) = P_{tot} E[s_k (\alpha_F(k) - \beta_F(k))]. \quad (2.5)$$

**Case 2:**  $-1 < \alpha_F(k) - \beta_F(k) < 0$ . We can show in Appendix A.2 that the following

inequality holds:

$$\begin{aligned} \sum_{k=1}^{\mathcal{K}} \frac{P_{tot} s_k [\alpha_F(k) - \beta_F(k)]}{\mathcal{K}} - \sum_{k=1}^{\mathcal{K}} \frac{P_{tot}^2 s_k^2 \beta_F(k) [\alpha_F(k) - \beta_F(k)]}{\mathcal{K}^2} \\ \geq J_1(\cdot) \geq \sum_{k=1}^{\mathcal{K}} \frac{P_{tot} s_k [\alpha_F(k) - \beta_F(k)]}{\mathcal{K}}. \end{aligned} \quad (2.6)$$

Similar to Case 1, it can be proved that  $\lim_{\mathcal{K} \rightarrow \infty} J_1(\cdot) = P_{tot} E[s_k(\alpha_F(k) - \beta_F(k))]$  for Case 2.

In summary, we have the following result for  $J_1(\cdot)$ :

$$\lim_{\mathcal{K} \rightarrow \infty} J_1(\cdot) = P_{tot} E[s_k(\alpha_F(k) - \beta_F(k))] \quad (2.7)$$

for  $|\alpha_F(k) - \beta_F(k)| < 1$ . Following similar procedures, the following result can be proved for  $J_2(\cdot)$ :

$$\lim_{\mathcal{K} \rightarrow \infty} J_2(\cdot) = P_{tot} E[s_k(\alpha_D(k) - \beta_D(k))] \quad (2.8)$$

when  $|\alpha_D(k) - \beta_D(k)| < 1$ . Combining (2.7) and (2.8), we obtain the following theorem.

**Theorem 2.2.1** *In a homogeneous detection system under two conditions:  $|\alpha_F(k) - \beta_F(k)| < 1$ ,  $|\alpha_D(k) - \beta_D(k)| < 1$ , the J-divergence for a large  $\mathcal{K}$  is as follows,*

$$J_\infty := \lim_{\mathcal{K} \rightarrow \infty} J(\cdot) = P_{tot} E[\eta_k]$$

with  $\eta_k = \frac{s_k(\alpha_F(k) + \alpha_D(k) - \beta_F(k) - \beta_D(k))}{2}$ .

It can be proved in Appendix A.3 that the two conditions in Theorem 2.2.1 always hold in the distributed detection system that we proposed. From Theorem 2.2.1, we see

that when the number of sensors increases, the J-divergence becomes saturated eventually unless the total transmission power keeps increasing (i.e.,  $P_{tot}$  keeps increasing). A natural question to ask: Do we gain any benefits by deploying more sensors if  $P_{tot}$  is fixed? The answer is positive if we are concerned with the long-term system performance in a fading environment, which will be discussed in the following section.

### 2.2.2 Detection Outage and Detection diversity

In this section, we investigate the long-term system performance that is a statistical quantity, which can be calculated over a large number of observation periods given our assumption that the system parameters are i.i.d. over time. For a network with slow-fading channels, it is not meaningful to discuss about average performance since each observation period only goes through one fading state. In particular, a meaningful average measure stands for the expected (over random realizations) performance for a given observation period. However, since here we assume that each observation period only goes through one random state, we are not able to experience the “expected” performance regarding this particular event. On the contrary, if during one observation period, the random states change across all possible values, the system output will experience an “expected” version of the performance. In other words, in our setup, each observation period is either doomed by a bad random state or blessed by a good random state; in the former case, an outage occurs. Therefore, it is more meaningful to evaluate the outage performance across time. As such, similar to the outage concept developed in the wireless communication community, we define the detection outage probability as the fraction of observation periods that experience bad random states, which lead to low  $J$  values. Specifically, we have

$$p_{J_0} = \text{Prob}\{J(\cdot) < J_0\}$$

where  $J_0$  is a predefined threshold. For the proposed distributed detection system, the following theorem holds.

**Theorem 2.2.2** *In a homogeneous distributed detection system, with  $E[\eta_k]$  being finite, for  $J_0 < J_\infty$  and a sufficiently-large  $\mathcal{K}$ , the outage probability is given as*

$$p_{J_0} \sim \exp(-\mathcal{K}I_{\eta_k}(c))$$

or

$$-\log p_{J_0} \sim \mathcal{K}I_{\eta_k}(c)$$

where  $c = \frac{J_0}{P_{tot}}$  and  $I_{\eta_k}(c)$  is the rate function (defined next in (2.9)).

*Proof:* The detailed proof is given in Appendix A.4.

For a special case of i.i.d. Rayleigh fading channels between the sensors and the fusion center such that  $f_{\eta_k}(x) = \frac{1}{\delta^2} \exp\{-\frac{x}{\delta^2}\}$ , it can be shown [41] that

$$\begin{aligned} I_{\eta_k}(c) &= \sup_{\theta \in \mathcal{R}} [\theta c - \log M_{\eta_k}(\theta)] \\ &= \sup_{\theta \in \mathcal{R}} [\theta c + \log(1 - b\theta)] \\ &= \left(\frac{c}{b} - 1\right) + \log \frac{b}{c} \\ &= \frac{J_0}{P_{tot}\delta^2} - \log \frac{J_0}{P_{tot}\delta^2} - 1, \end{aligned} \tag{2.9}$$

where  $M_{\eta_k}(\theta) = \frac{1}{1-b\theta}$  is the moment generating function for  $\eta_k$  with  $b = \delta^2$ . Therefore, at high power regime where  $P_{tot} \gg \frac{J_0}{\delta^2}$ , we have  $I_{\eta_k}(c) \approx \log P_{tot}$  such that

$$\begin{aligned} &-\log p_{J_0} \sim \mathcal{K}I_{\eta_k}(c) \\ \implies &-\log p_{J_0} \sim \mathcal{K} \log P_{tot}. \end{aligned} \tag{2.10}$$

The above result implies that if we draw  $-\log p_{J_0}$  over  $\log P_{tot}$ , the slope is  $\mathcal{K}$  at high power regime, where  $\mathcal{K}$  is the total number of sensors. In other words, the outage probability  $p_{J_0}$  decreases exponentially over  $\mathcal{K}$ . Similar to the concept of spatial diversity in multiple antenna systems, where the diversity order is defined as the slope of the packet error rate curve (at high signal to noise ratio regime), we define the detection diversity order as the slope of the detection outage probability curve (at the high power regime). From the result in Theorem 2.2.2, we see that the proposed distributed detection scheme achieves a full diversity order, which is equal to the total number of nodes in the system.

We now further illustrate the concept of detection diversity via a numerical example. We set  $s_k = \frac{G_0}{\sigma_k^2 d_k^2} |r_k|^2$  where  $d_k$  is the transmission distance from sensor  $k$  to the fusion center ( $d_k = 100$  m for all  $k$ 's),  $\sigma_k^2 = -70$  dBm for all  $k$ 's is the channel noise power,  $G_0 = -30$  dB is the nominal gain at the unit distance  $d_k = 1$  m, and  $|r_k|$  is an i.i.d. Rayleigh fading random variable with unit variance. We further set  $P_D(k) = 0.5$  and  $P_F(k) = 0.01$  for all  $k$ 's, and set  $J_0 = 1.2$ .

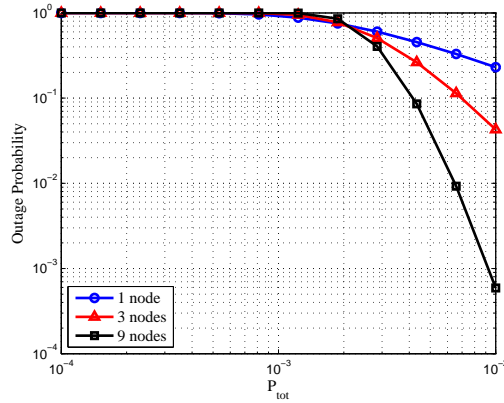


Figure 2.2: Outage probability versus  $P_{tot}$  with the uniform transmission strategy.

In Fig. 2.2, we draw the outage probability over  $P_{tot}$ , where we see that at reasonably high  $P_{tot}$  values, the slope of the three-node case is three times as large as that of the one-node case, while the slope of the nine-node case is nine times as large as that of the

one-node case. We conclude that even with the simple uniform transmission strategy with  $a_k^2 = \frac{P_{tot}}{\mathcal{K}}$ , the full detection diversity of order  $\mathcal{K}$  can be achieved.

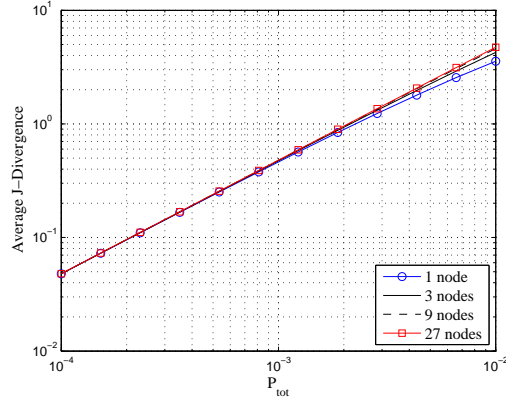


Figure 2.3: Average J-divergence versus  $P_{tot}$  with the uniform transmission strategy.

In Fig. 2.3, we draw the average J-divergence achieved with different numbers of nodes, where we see that the average performance shows no significant improvement when we increase the number of nodes for a fixed  $P_{tot}$ . While this verifies the results in Theorem 2.2.1, it cannot lead to the conclusion that a three-node system is as good as a nine-node system in a slow-fading environment. In such an environment, outage probability should be the correct criterion to use, and more nodes stand for a higher diversity gain that we could explore.

## 2.3 Optimal-power Transmission Strategy: Adaptive Power Gain and Full Diversity Gain

### 2.3.1 Optimal-power transmission with a total power constraint

If channel states are available to the fusion center and there exist downlink control channels from the fusion center to all the sensors, the fusion center could calculate and adjust the transmission power of each sensor node in an optimal way. Such a strategy can be

formulated as follows.

$$\begin{aligned}
 & \max && J(\cdot) \\
 & \text{s.t.} && \sum_{k=1}^{\mathcal{K}} a_k^2 \leq P_{tot}, \\
 & && a_k \geq 0, k = 1, \dots, \mathcal{K},
 \end{aligned} \tag{2.11}$$

where the design variables are  $a_k$ 's.

The convexity conditions and the optimal solution for (2.11) are given in [22], which is not the focus here. Instead, we are interested in the resulting outage performance. Using the same numerical setup as for calculating the data in Fig. 2.2, we draw the outage probability versus  $P_{tot}$  in Fig. 2.4, where we see that the resulting curve has the same slope at high power regime as the uniform transmission strategy. In other words, a full diversity order of  $\mathcal{K}$  is still achieved. Furthermore, the left shift of the outage probability curve stands for an adaptive power gain that is obtained by optimally controlling the transmission power at each node.

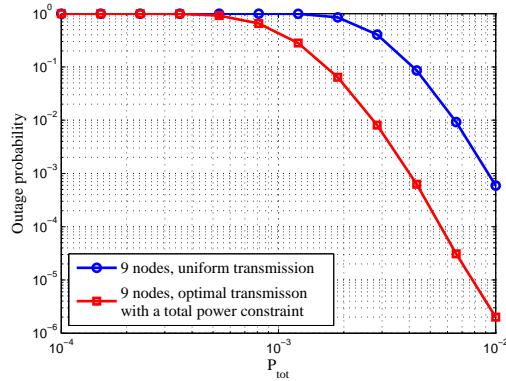


Figure 2.4: Outage probabilities for the power-optimal transmission strategy and the uniform transmission strategy.

### 2.3.2 Optimal-power transmission with both individual and total power constraints

In a practical system, each sensor may have an individual power constraint imposed by its battery or certain transmission regulations. As such, we can modify the optimization problem given in (2.11) as follows,

$$\begin{aligned}
 \max \quad & J(\cdot) \\
 \text{s.t.} \quad & \sum_{k=1}^{\mathcal{K}} a_k^2 \leq P_{tot}, \\
 & 0 \leq a_k \leq a_k^{max}, k = 1, \dots, \mathcal{K},
 \end{aligned} \tag{2.12}$$

where  $a_k^{max}$  is determined by the square root of the individual maximum allowed power. Since the added individual power constraints are linear constraints, there is no change on the convexity conditions given for (2.11) such that the optimal solution can be similarly solved. Sequential optimization techniques discussed in [41] can also be applied to solve (2.12), which is skipped in this chapter. Accordingly, we draw the detection outage probability in Fig. 2.5 for the case with both the individual and total constraints, against the case with only the total power constraint. We see that with extra individual power constraints, full diversity can still be achieved while the power gain is reduced compared with the case without individual power constraints.

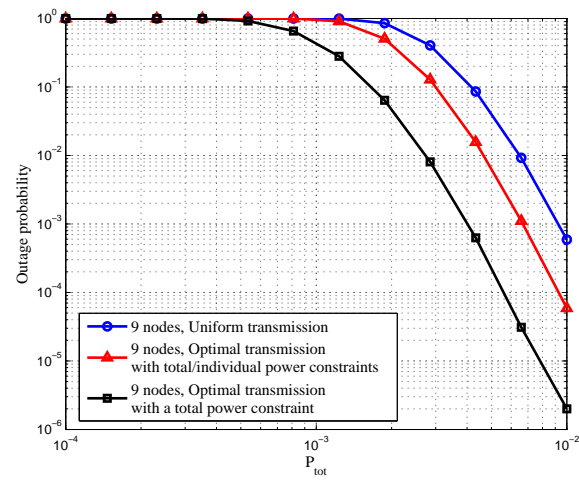


Figure 2.5: Outage probability versus  $P_{tot}$  for different scenarios.

## CHAPTER 3

POWER CONTROL STRATEGY FOR DISTRIBUTED MULTIPLE HYPOTHESIS  
DETECTION

In this chapter, we discuss power control strategy for multiple hypothesis system by using total J-divergence which is a pair-wise sum of individual J-divergences used in Chapter 2 and classification outage to evaluate the system performance of slow-fading WSNs.

## 3.1 System Model

We introduce a similar system setup as that used in Chapter 2, but the system model is extended for multiple hypothesis detection and includes a local modulation scheme  $f_s(\cdot)$  based on the local decision statistic for extra power gain. At some observation time, detection of the event source (for example, detection of multiple targets or target classification) can be abstracted as multiple hypotheses:  $H_1, H_2, \dots, H_{\mathcal{H}}$ . There are  $\mathcal{K}$  sensors that make local observations about the source and transmit local decisions to the fusion center. For simplicity, we assume that the local decisions are transmitted to the fusion center over orthogonal multiple access channels as shown in Figure 3.1. Specifically, sensor  $k$  collects a local observation  $x_k$  that has been corrupted by observation noise. Each sensor then makes its own local decision  $u_k$  according to a local decision rule denoted symbolically as

$$u_k = \gamma_k(x_k) \in \{1, 2, \dots, \mathcal{H}\}$$

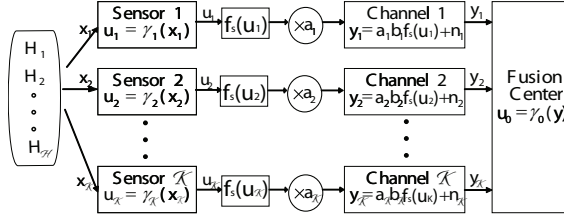


Figure 3.1: Distributed Multiple Hypothesis Detection with a Fusion Center.

Node  $k$  then maps its local decision  $u_k$  to a transmit symbol according to some modulation scheme,  $f_s(u_k)$ , amplifies the transmit symbol by  $a_k$ , and transmits the signal over the  $k$ th fading channel. In general, the modulation scheme  $f_s(u_k)$  can be scalar (e.g., amplitude modulation) or multi-dimensional (e.g., phase shift keying (PSK) or frequency shift keying (FSK)). The fading communication channel is modeled as a set of random amplitude gains  $b_k$ , and the signal received at the fusion center is corrupted by additive white Gaussian noise (AWGN). We assume that the  $b_k$ 's and  $n_k$ 's are both independent over  $k$ . Later in this chapter, we use outage probability which is introduced in Chapter 2 to quantify long-term system performance over multiple independent and identically distributed (i.i.d.) realizations of the channel gain coefficients.

From now on, we use a bold capital letter for a matrix and a bold lower-case letter for a vector.  $\mathcal{K}$  is the total number of channels.  $\mathcal{H}$  is the total number of hypotheses, and  $\mathcal{D}$  is the total number of dimensions used for the local modulation scheme.

We can characterize an individual sensor by its transition probabilities

$$P(u_k = l | H_i) \text{ for } l \text{ and } i = 1, 2, \dots, \mathcal{H}.$$

We further assume that the local observation  $x_k$ 's and the local decision  $u_k$ 's are independent over  $k$  when conditioned on a particular hypothesis [1] [42]. As such, the joint

conditional probability mass function of the local decisions is

$$P(\mathbf{u}|H_i) = \prod_{k=1}^{\mathcal{K}} P(u_k|H_i) \quad (3.1)$$

where the local decision vector is  $\mathbf{u} = [u_1, \dots, u_{\mathcal{K}}]^T$ . In Chapter 2, these transition probabilities can be represented only with respect to probability of detection,  $P_D(k)$  and probability of false alarm,  $P_F(k)$  due to two hypothesis problem. However, for multiple hypothesis problem, we use the above notation without using  $P_D(k)$  and  $P_F(k)$ .

We now allow each sensor to adaptively control its transmit symbol constellation according to its own unique decision statistics. Let  $\mathbf{f}_s(\cdot)$  be the optimization of the transmit symbol constellation. The optimized modulation symbols are transmitted to the fusion center over a fading channel. At the fusion center, the received signals are

$$\mathbf{y} = \mathbf{B}\mathbf{A}\mathbf{f}_s(\mathbf{u}) + \mathbf{n} \quad (3.2)$$

where  $\mathbf{B} = \text{diag}\{b_1, b_2, \dots, b_{\mathcal{K}}\}$ ,  $\mathbf{A} = \text{diag}\{a_1, a_2, \dots, a_{\mathcal{K}}\}$ ,  $\mathbf{f}_s(\mathbf{u}) = [f_s(u_1), f_s(u_2), \dots, f_s(u_{\mathcal{K}})]^T$ ,  $\mathbf{y} = [y_1, y_2, \dots, y_{\mathcal{K}}]^T$ , and where  $\mathbf{n} = \{n_1, n_2, \dots, n_{\mathcal{K}}\}^T$  is the additive noise vector. The conditional PDF of the received signal given hypothesis  $H_i$  is

$$p(\mathbf{y}|H_i) = \sum_{\mathbf{u}} p(\mathbf{y}|\mathbf{u})P(\mathbf{u}|H_i). \quad (3.3)$$

The final decision at the fusion center,  $u_0$ , is determined by a fusion rule denoted symbolically by  $u_0 = \gamma_0(\mathbf{y})$ .

In general,  $\{P(H_i)|i = 1, 2, \dots, \mathcal{H}\}$  may be unknown but can be estimated with a given local transition matrix and local decision probabilities. For example, a vector of a

priori source probabilities could be estimated according to  $\tilde{P}_k(\mathbf{H}) = \mathbf{L}_k^{-1} \cdot \tilde{P}(u_k)$  where

$$\mathbf{L}_k = \begin{pmatrix} P(u_k = 1|H_1) & \cdots & P(u_k = 1|H_{\mathcal{H}}) \\ P(u_k = 2|H_1) & \cdots & P(u_k = 2|H_{\mathcal{H}}) \\ \vdots & \ddots & \vdots \\ P(u_k = \mathcal{H}|H_1) & \cdots & P(u_k = \mathcal{H}|H_{\mathcal{H}}) \end{pmatrix}$$

is given.  $\tilde{P}(u_k)$  is a locally estimated vector of decision probabilities obtained by averaging in time over many independent decisions. For example, suppose the  $k$ th sensor operates for some period of time and makes 100 decisions. If the sensor chooses the second hypothesis ( $u_k = 2$ ) three times, then  $\tilde{P}(u_k = 2) = 3/100$ . However, this approach yields estimates of the decision probabilities, not of the source prior probabilities. The two are related by  $\tilde{P}_k(\mathbf{H}) = \mathbf{L}_k^{-1} \cdot \tilde{P}(u_k)$ , but unfortunately this can lead to negative prior probabilities because they are calculated from estimated  $\tilde{P}(u_k)$  rather than the true local decision probabilities. Therefore, we have instead performed a simple nonlinear optimization:

$$\begin{aligned} \min \quad & \|\tilde{P}_k(\mathbf{H}) - \mathbf{L}_k^{-1} \cdot \tilde{P}(u_k)\|^2 \\ \text{s.t.} \quad & \sum_{i=1}^{\mathcal{H}} \tilde{P}_k(H_i) = 1 \\ & \tilde{P}_k(H_i) \geq 0 \text{ for all } i. \end{aligned}$$

The final global estimate  $\tilde{P}(H_i)$  is calculated by averaging over the  $\mathcal{K}$  sensor estimates

$(\tilde{P}_k(H_i))$  according to

$$\tilde{P}(H_i) = \frac{1}{\mathcal{K}} \sum_{k=1}^{\mathcal{K}} \tilde{P}_k(H_i).$$

## 3.2 Power Control Strategy

### 3.2.1 Local Power Control

In this chapter, we adopt two power optimization strategies-local power control and channel power control-to take advantage of local sensor statistics and channel statistics. As for the local power control, we allow each sensor to adaptively control its transmit symbol constellation according to its own unique decision statistics. Let  $\{\delta_k(l) = f_s(u_k = l) | l = 1, 2, \dots, \mathcal{H}\}$  be the local modulation symbols of the  $k$ th sensor where the modulation function  $f_s(\cdot)$  adjusts the symbols based on local decision probabilities,  $\{P(u_k = h) | h = 1, 2, \dots, \mathcal{H}\}$ . The local power control minimizes the average power of the modulation constellation. Let this optimization for the  $k$ th sensor be stated as

$$\begin{aligned} \min_{\delta_k(1), \delta_k(2), \dots, \delta_k(\mathcal{H})} PW_k &= \sum_{l=1}^{\mathcal{H}} |\delta_k(l)|^2 \cdot P(u_k = l) \\ \text{s.t. } F(\cdot) &= \text{const} \end{aligned} \quad (3.4)$$

where the constraint  $F(\cdot)$  can be any communication performance measure between the  $k$ th sensor and the fusion center. The design variables are  $\delta_k(1), \delta_k(2), \dots, \delta_k(\mathcal{H})$ . The basic goal of the local power control is to find an optimal local constellation configuration that minimizes the average constellation power while keeping a given communication performance. We could let the constellation configuration be completely arbitrary, but this

approach would likely lead to impractical implementations and increases the optimization procedure considerably. Instead, we essentially implement  $F(\cdot)$  as a *structural constraint* by assuming a symbol constellation in the form of some traditional modulation scheme (QAM, FSK, etc) with spacing determined by desired communication performance. Once the constellation structure and size are set, we can then optimize the average power of the constellation through linear translations of the constellation and by judicious assignment of decisions to symbols in the constellation. Note that communication performance is determined by the constellation structure, not by its translational shifts. We later show an example using a translated QAM constellation including a fair comparison to the performance of a system with a non-translated QAM modulation transmission. This example is shown in 3.3.4 and 3.4. We also show an example using scalar modulation in 3.3.3. Optimization of the transmit symbol constellation as defined by (3.4) is referred to as *local power control* while optimization of the average power transmitted by each sensor based on channel states is referred to as *channel power control*.

### 3.2.2 System Performance Measure, Total J-divergence & Channel Power Control

Rather than considering optimal fusion rules, we instead argue that detection performance is generally improved if the total J-divergence  $J^{total}$  is maximized. In particular, the total J-divergence measure is defined by a weighted sum of pairwise J-divergences,  $J^{ij} = J(p(\mathbf{y}|H_i), p(\mathbf{y}|H_j))$ , according to [43][44]

$$\begin{aligned}
 J^{total} &= \frac{1}{2} \sum_{i=1}^{\mathcal{H}} \sum_{j=1}^{\mathcal{H}} J(p(\mathbf{y}|H_i), p(\mathbf{y}|H_j)) \cdot P(H_i) \cdot P(H_j) \\
 &= \sum_{i=1}^{\mathcal{H}} \sum_{j=i+1}^{\mathcal{H}} J(p(\mathbf{y}|H_i), p(\mathbf{y}|H_j)) \cdot P(H_i) \cdot P(H_j). \tag{3.5}
 \end{aligned}$$

The second line of (3.5) is possible because  $J^{ij}$  is symmetric and  $J^{ij} = 0$  when  $i = j$ . The pairwise J-divergences are  $J^{ij} = \frac{1}{2}[D(p_j||p_i)+D(p_i||p_j)]$  where  $D(p_i||p_j)$  is the Kullback-Leibler (KL) divergence measure between two probability density functions [36]. Thus, the J-divergence is a symmetric version of the more general KL distance measure.

The KL distance is the average of the difference between two log-likelihood functions. Let  $\log(p_1)$  and  $\log(p_2)$ , respectively, be the two conditional log-likelihood functions for hypotheses  $H_1$  and  $H_2$ . The KL distance is defined as  $E_1[\log(p_1) - \log(p_2)] = E_1[\log(\frac{p_1}{p_2})]$  where  $E_1[\cdot]$  is the expected value with respect to  $p_1$ . From this definition, the KL distance is interpreted as the average of the log likelihood ratio between two conditional PDFs  $p_1$  and  $p_2$ . Because the likelihood ratio is an optimal detection method that appears in both Neyman-Pearson and Bayesian detection, we can conclude that J-divergence is closely related with detection performance. In fact, for the binary Gaussian detection problem, J-divergence becomes the SNR at the receiver and the probability of error is  $Q\left(\sqrt{\frac{J(\cdot)}{2}}\right)$  where  $Q(\cdot)$  is the Gaussian Q-function and  $J(\cdot)$  is the J-divergence between the two hypotheses. Asymptotically, J-divergence determines the error exponent of the Chernoff bound from Stein's lemma [1], [36].

For distributed detection systems, the simulation results of [1] show that probability of detection can be enhanced by increasing J-divergence. Even though J-divergence is not a direct performance measure like probability of detection or probability of error, it usually leads to tractable analytical frameworks for distributed detection systems with Gaussian assumptions and has been adopted by many researchers, such as in [37], [39], [40], [38]. J-divergence also provides a lower bound to the probability of error by  $P_e > P(H_1)P(H_2)e^{-J/2}$  [37] in a binary detection system.

We apply J-divergence to the multiple-hypothesis detection system defined in (3.5) because of its relationship to detection performance and its ability to provide a tractable

analysis. For example,  $J^{total}$  provides a lower bound for the multiple-hypothesis detection problem by  $P_e^{bound} \geq C - J^{total}$  where  $P_e^{bound}$  is defined as pair-wise sum of the individual probabilities of error between two hypotheses ( $P_e^{bound} = \sum_{i=1}^{\mathcal{H}} \sum_{j=1}^{\mathcal{H}} P_e^{ij}$ ) and  $C$  is a constant ( $C = \sum_{i=1}^{\mathcal{H}} \sum_{j=1}^{\mathcal{H}} P(H_i)p(H_j)$ ). A detailed proof is given in Appendix B.1. We can reduce this bound by increasing the total J-divergence. Therefore, we can design a power control strategy by maximizing the total J-divergence according to

$$\begin{aligned} \max_{\mathbf{A}} \quad & J^{total}(\cdot) \\ \text{s.t.} \quad & Tr[\mathbf{A}\mathbf{A}^T] \leq P_{tot}, \\ & 0 \preceq \mathbf{A} \preceq \sqrt{\mathbf{P}_{max}}, \end{aligned}$$

where  $\mathbf{A}$  is the  $\mathcal{K} \times \mathcal{K}$  diagonal amplification matrix defined earlier,  $P_{tot}$  is a total power constraint, and  $\mathbf{P}_{max}$  is a  $\mathcal{K} \times \mathcal{K}$  diagonal matrix of individual power constraints.  $\sqrt{\mathbf{P}_{max}}$  denotes the component-wise square root of  $\mathbf{P}_{max}$ , and the inequality  $\preceq$  means that  $\sqrt{\mathbf{P}_{max}} - \mathbf{A}$  is positive semidefinite[1]. The individual diagonal terms of  $\sqrt{\mathbf{P}_{max}}$  need not be the same.

Note that the pairwise J-divergence  $J^{ij}$  directly depends only on the conditional probability densities,  $p(\mathbf{y}|H_i)$  and  $p(\mathbf{y}|H_j)$ , but apparently not at all on the source's prior probabilities. Hence, it is not immediately obvious how a local optimization procedure can have any effect on system performance. Closer inspection of (3.2) and (3.3), however, shows that  $J^{ij}$  does depend on the transmit symbol constellations used by the various sensors. Since the symbol constellations are the design variables in the local power optimization,  $J^{ij}$  does indeed depend on a priori probabilities of source. Essentially, the local optimization reduces the average power of the symbol constellation for each sensor,

which we will see later allows the amplification coefficients  $a_k$  to be increased while still meeting average power constraints on the total transmission power.

As discussed in Chapter 2, with AWGN channel noise, the system conditional probabilities can be approximated by Gaussian densities and the resulting individual J-divergence  $J^{ij}$  is given as

$$\begin{aligned} J(p(\mathbf{y}|H_i), p(\mathbf{y}|H_j)) = & \\ & \frac{1}{2} \text{Tr} \left[ [\mathbf{R} + \mathbf{B}\mathbf{A}(\mathbf{C}_i + \boldsymbol{\xi}\boldsymbol{\xi}^T)\mathbf{A}^T\mathbf{B}^T] \cdot [\mathbf{R} + \mathbf{B}\mathbf{A}\mathbf{C}_j\mathbf{A}^T\mathbf{B}^T]^{-1} \right] + \\ & \frac{1}{2} \text{Tr} \left[ [\mathbf{R} + \mathbf{B}\mathbf{A}(\mathbf{C}_j + \boldsymbol{\xi}\boldsymbol{\xi}^T)\mathbf{A}^T\mathbf{B}^T] \cdot [\mathbf{R} + \mathbf{B}\mathbf{A}\mathbf{C}_i\mathbf{A}^T\mathbf{B}^T]^{-1} \right] - \mathcal{K} \end{aligned} \quad (3.6)$$

where  $\mathbf{C}_i$  is a  $\mathcal{K} \times \mathcal{K}$  diagonal matrix with elements

$$C_i(k, k) = \sum_{l=1}^{\mathcal{H}} \sum_{m=l+1}^{\mathcal{H}} \{\delta_k(l) - \delta_k(m)\}^2 P(u_k = l|H_i) \cdot P(u_k = m|H_i)$$

for  $i \in \{1, 2, \dots, \mathcal{H}\}$ ,

$$\boldsymbol{\xi} = \boldsymbol{\xi}_0 - \boldsymbol{\xi}_1,$$

$$\boldsymbol{\xi}_i = \left[ \sum_{l=1}^{\mathcal{H}} \delta_1(l)P(u_1 = l|H_i), \sum_{l=1}^{\mathcal{H}} \delta_2(l)P(u_2 = l|H_i), \dots, \sum_{l=1}^{\mathcal{H}} \delta_{\mathcal{K}}(l)P(u_{\mathcal{K}} = l|H_i) \right]^T,$$

$\mathbf{R}$  is the covariance matrix of the channel noise vector  $\mathbf{n}$ , and  $\mathcal{K}$  is the dimension of the received signal vector  $\mathbf{y}$  at the fusion center. More details on the derivation of (3.6) are given in Appendix B.2. Finally, the total J-divergence of the system is obtained by applying (3.6) to (3.5). In the next section, we consider the performance benefit of local optimization when average transmit power is equal across sensors.

### 3.3 Classification Outage of a Uniform Transmit Strategy with Local Power Optimization

In this section, we discuss a scheme where local power optimization and uniform channel power control cooperate such that the average transmit power  $P_k$  of each sensor is equal, and analytically derive classification outage probability through asymptotic total J-divergence for distributed multiple hypothesis detection. We use the classification outage to evaluate the diversity gain of uniform transmit strategy in a multiple hypothesis system.

#### 3.3.1 Uniform Transmit Strategy

First, define the average power transmitted by the  $k$ th sensor as  $P_k = PW_k^* \cdot a_k^2$  where  $PW_k^* = \sum_{l=1}^{\mathcal{H}} |\delta_k^*(l)|^2 \cdot P(u_k = l)$  is the minimized constellation power obtained through local optimization. In other words, the asterisk denotes that the constellation has been selected to minimize the average power in the symbol constellation. The total transmit power constraint for all sensors is  $P_{tot} = \sum_{k=1}^{\mathcal{K}} P_k$ , which for uniform transmit power requires that  $P_k = \frac{P_{tot}}{\mathcal{K}}$ . Substituting for  $PW_k^*$ , the amplifying factor for the  $k$ th sensor is  $a_k = \sqrt{\frac{P_{tot}}{\mathcal{K} \cdot \sum_{l=1}^{\mathcal{H}} |\delta_k^*(l)|^2 \cdot P(u_k=l)}}$ . Note that the amplification factor is inversely related to the pre-amplification average power of the symbol constellation. Thus, the local optimization step has allowed larger  $a_k$  while still meeting the total power constraint. Even in Uniform Transmit Strategy, the local amplifying parameter  $a_k$  are not constant, while  $a_k = \sqrt{\frac{P_{tot}}{\mathcal{K}}}$  which is a constant in the power control strategy defined in Chapter 2 since there is no local optimization. The optimized constellation,  $\delta_k^*(\cdot)$  changes the system parameters in the total J-divergence measure as shown in (3.7). The dimension of the received signal vector  $\mathbf{y}$  at the fusion center is the same as the number of sensors  $\mathcal{K}$  since the system is modeled as having orthogonal communication channels. We set  $\mathbf{R} = \text{diag}\{\sigma_1^2, \sigma_2^2, \dots, \sigma_{\mathcal{K}}^2\}$  where

$\sigma_k^2$  is the noise power of the  $n$ th channel. From (3.5) and (3.6), the total J-divergence is

$$\begin{aligned}
J^{total} &= \frac{1}{2} \sum_{i=1}^{\mathcal{H}} \sum_{j=i+1}^{\mathcal{H}} \left[ \sum_{k=1}^{\mathcal{K}} \frac{a_k^2 b_k^2 \{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\}}{\sigma_k^2 + a_k^2 b_k^2 \psi_{ik}} + \sum_{k=1}^{\mathcal{K}} \frac{a_k^2 b_k^2 \{\psi_{ik} - \psi_{jk} + \phi_k^{ij}\}}{\sigma_k^2 + a_k^2 b_k^2 \psi_{jk}} \right] P(H_i) P(H_j) \\
\text{where } \psi_{ik} &= \sum_{l=1}^{\mathcal{H}} \sum_{m=l+1}^{\mathcal{H}} [(\delta_k^*(l) - \delta_k^*(m))^2 \cdot P_k(u_k = l|H_i) \cdot P_k(u_k = m|H_i)], \\
\phi_k^{ij} &= \sum_{l=1}^{\mathcal{H}} [\delta_k^*(l) \cdot (P_k(u_k = l|H_i) - P_k(u_k = l|H_j))]^2,
\end{aligned} \tag{3.7}$$

where  $i$  and  $j$  are hypothesis indices,  $k$  is the index of an individual sensor, and  $l, m$  are the indices of transmission symbols. Applying the uniform per-sensor power constraint and defining the  $k$ th channel's SNR as  $s_k = b_k^2/\sigma_k^2$ , the total J-divergence becomes

$$\begin{aligned}
J^{total} &= \frac{1}{2} \sum_{i=1}^{\mathcal{H}} \sum_{j=i+1}^{\mathcal{H}} \left[ \sum_{k=1}^{\mathcal{K}} \frac{P_{tot} s_k \{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\}}{\mathcal{K} \cdot \sum_{h=1}^{\mathcal{H}} |\delta_k^*(l)|^2 \cdot P(u_k = l) + P_{tot} s_k \{\psi_{ik}\}} \right. \\
&\quad \left. + \sum_{k=1}^{\mathcal{K}} \frac{P_{tot} s_k \{\psi_{ik} - \psi_{jk} + \phi_k^{ij}\}}{\mathcal{K} \cdot \sum_{h=1}^{\mathcal{H}} |\delta_k^*(l)|^2 \cdot P(u_k = l) + P_{tot} s_k \{\psi_{jk}\}} \right] \cdot P(H_i) \cdot P(H_j) \\
&= \frac{1}{2} \sum_{i=1}^{\mathcal{H}} \sum_{j=i+1}^{\mathcal{H}} \left[ \sum_{k=1}^{\mathcal{K}} \frac{P_{tot} s_k \{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\}}{\mathcal{K} \cdot PW_k^* + P_{tot} s_k \{\psi_{ik}\}} \right. \\
&\quad \left. + \sum_{k=1}^{\mathcal{K}} \frac{P_{tot} s_k \{\psi_{ik} - \psi_{jk} + \phi_k^{ij}\}}{\mathcal{K} \cdot PW_k^* + P_{tot} s_k \{\psi_{jk}\}} \right] \cdot P(H_i) \cdot P(H_j).
\end{aligned} \tag{3.8}$$

In the following, we will analytically derive the classification outage with (3.7), show a simulation result by the uniform transmit strategy, and then compare the analytical diversity gain of the classification outage probability and the gain obtained through the simulation result.

### 3.3.2 Asymptotic total J-divergence $J_\infty$ and Classification Outage

Intuitively, J-divergence can be increased by increasing the number of sensors. If a total power constraint is enforced, however, there is an asymptotic limit to the increase since a finite amount of power must be distributed among more and more sensors. In this section, we derive the asymptotic total J-divergence expression and show its relationship to classification outage probability in a homogeneous sensor environment. We begin by defining the two terms that constitute each individual J-divergence,  $J^{ij} = J(p(\mathbf{y}|H_i), p(\mathbf{y}|H_j)) = J_1^{ij} + J_2^{ij}$  in (3.5) and (3.7) as  $J_1^{ij} = \frac{1}{2} \sum_{k=1}^{\mathcal{K}} \frac{a_k^2 b_k^2 \{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\}}{\sigma_k^2 + a_k^2 b_k^2 \psi_{ik}}$  and  $J_2^{ij} = \frac{1}{2} \sum_{k=1}^{\mathcal{K}} \frac{a_k^2 b_k^2 \{\psi_{ik} - \psi_{jk} + \phi_k^{ij}\}}{\sigma_k^2 + a_k^2 b_k^2 \psi_{jk}}$ . We use these expressions to derive the upper and lower bounds for  $J^{ij}$ . We then show that  $J^{ij}$  converges to an asymptotic value by showing that the upper and lower bounds converge to the same value as  $\mathcal{K}$  goes to infinity. We first consider  $J_1^{ij}$ .

**Case 1:** For  $(\psi_{jk} - \psi_{ik} + \phi_k^{ij}) \geq 0$ , the lower and upper bounds for  $J_1^{ij}$  are

$$\begin{aligned} & \sum_{k=1}^{\mathcal{K}} \frac{P_{tot} s_k \{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\}}{\mathcal{K} \cdot PW_k^*} - \sum_{k=1}^{\mathcal{K}} \frac{P_{tot}^2 s_k^2 \psi_{ik} \{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\}}{\mathcal{K}^2 \cdot PW_k^{*2}} \\ & \leq 2J_1^{ij} \leq \sum_{k=1}^{\mathcal{K}} \frac{P_{tot} s_k \{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\}}{\mathcal{K} \cdot PW_k^*} \end{aligned} \quad (3.9)$$

where  $P_{tot}$ ,  $s_k$ , and  $\psi_{ik}$  are positive values. A detailed proof is given in Appendix B.3. Consider a homogeneous sensor network where  $\psi_{ik}$ ,  $\psi_{jk}$  and  $\phi_k^{ij}$  are the same for each sensor and the  $s_k$ 's are i.i.d., then by the law of large numbers (LLN), as  $\mathcal{K} \rightarrow \infty$  we obtain

$$\begin{aligned}
& \sum_{k=1}^{\mathcal{K}} \frac{P_{tot} s_k \{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\}}{\mathcal{K} \cdot PW_k^*} \rightarrow P_{tot} E[s_k \{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\} / PW_k^*], \\
& \sum_{k=1}^{\mathcal{K}} \frac{P_{tot}^2 s_k^2 \psi_{ik} \{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\}}{\mathcal{K} \cdot PW_k^*} \rightarrow P_{tot}^2 E[s_k^2 \psi_{ik} \{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\} / PW_k^*] \\
\text{such that } & \sum_{k=1}^{\mathcal{K}} \frac{P_{tot}^2 s_k^2 \psi_{ik} \{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\}}{\mathcal{K}^2 \cdot PW_k^{*2}} \rightarrow 0
\end{aligned} \tag{3.10}$$

because  $s_k$ ,  $\psi_{ik}$ ,  $\phi_k^{ij}$  are finite and  $PW_k^* \neq 0$ . Therefore, for  $(\psi_{jk} - \psi_{ik} + \phi_k^{ij}) \geq 0$ , we have

$$\lim_{\mathcal{K} \rightarrow \infty} J_1^{ij} = \frac{1}{2} P_{tot} E \left[ \frac{s_k \{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\}}{PW_k^*} \right].$$

**Case 2:** For  $(\psi_{jk} - \psi_{ik} + \phi_k^{ij}) \leq 0$ , the lower and upper bound for  $J_1^{ij}$  are

$$\begin{aligned}
& \sum_{k=1}^{\mathcal{K}} \frac{P_{tot} s_k \{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\}}{\mathcal{K} \cdot PW_k^*} \leq 2J_1^{ij} \leq \\
& \sum_{k=1}^{\mathcal{K}} \frac{P_{tot} s_k \{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\}}{\mathcal{K} \cdot PW_k^*} - \sum_{k=1}^{\mathcal{K}} \frac{P_{tot}^2 s_k^2 \psi_{ik} \{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\}}{\mathcal{K}^2 \cdot PW_k^{*2}}
\end{aligned} \tag{3.11}$$

where  $P_{tot}$ ,  $s_k$ ,  $\psi_{ik}$ , and  $PW_k^*$  are positive. A detailed proof is given in Appendix B.4.

Similar to Case 1, for  $(\psi_{jk} - \psi_{ik} + \phi_k^{ij}) \leq 0$ , we have

$$\lim_{\mathcal{K} \rightarrow \infty} J_1^{ij} = \frac{1}{2} P_{tot} E \left[ \frac{s_k \{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\}}{PW_k^*} \right].$$

Hence, the limit is valid for any  $(\psi_{jk} - \psi_{ik} + \phi_k^{ij})$ , and by similar procedures it can be shown that

$$\lim_{\mathcal{K} \rightarrow \infty} J_2^{ij} = \frac{1}{2} P_{tot} E \left[ \frac{s_k \{\psi_{ik} - \psi_{jk} + \phi_k^{ij}\}}{PW_k^*} \right].$$

Therefore, by combining the asymptotic expressions for  $J_1^{ij}$  and  $J_2^{ij}$ , we obtain the following theorems.

**Theorem 3.3.1** *In a homogeneous distributed detection system with finite  $E[\eta_k^{ij}]$  and nonzero  $PW_k^*$ , the large- $\mathcal{K}$  asymptotic pairwise J-divergence between  $p(\mathbf{y}|H_i)$  and  $p(\mathbf{y}|H_j)$  is*

$$J_\infty^{ij} := \lim_{\mathcal{K} \rightarrow \infty} J^{ij}(\cdot) = P_{tot} E[\eta_k^{ij}]$$

where  $\eta_k^{ij} = \frac{s_k \phi_k^{ij}}{PW_k^*}$ .

**Theorem 3.3.2** *In a homogeneous distributed detection system with finite  $E[\eta_k^{total}]$  and nonzero  $PW_k^*$ , the asymptotic total J-divergence is*

$$\begin{aligned} J_\infty^{total} &:= \sum_{i=1}^{\mathcal{H}} \sum_{j=i+1}^{\mathcal{H}} \left\{ \lim_{\mathcal{K} \rightarrow \infty} J^{ij}(\cdot) \right\} \cdot P(H_i) \cdot P(H_j) \\ &= \sum_{i=1}^{\mathcal{H}} \sum_{j=i+1}^{\mathcal{H}} J_\infty^{ij} \cdot P(H_i) \cdot P(H_j) \\ &= P_{tot} \sum_{i=1}^{\mathcal{H}} \sum_{j=i+1}^{\mathcal{H}} E[\eta_k^{ij}] \cdot P(H_i) \cdot P(H_j) \\ &= P_{tot} E \left[ \sum_{i=1}^{\mathcal{H}} \sum_{j=i+1}^{\mathcal{H}} \eta_k^{ij} \cdot P(H_i) \cdot P(H_j) \right] \\ &= P_{tot} E[\eta_k^{total}] \end{aligned} \tag{3.12}$$

where  $\eta_k^{total} = \sum_{i=1}^{\mathcal{H}} \sum_{j=i+1}^{\mathcal{H}} \eta_k^{ij} \cdot P(H_i) \cdot P(H_j)$ . That is to say,  $\eta_k^{total}$  is linear combination of random variables  $\eta_k^{ij}$  where  $P(H_i)$  are fixed values.

The asymptotic total J-divergence depends on the channel statistics, local statistics, and the local symbol constellation since  $\eta_k^{ij} = \frac{s_k \phi_k^{ij}}{PW_k^*}$  where  $s_k$  is the  $k$ th channel's SNR,

$\phi_k^{ij}$  is calculated by local statistics and symbol constellation, and  $PW_k^*$  is the minimized constellation power obtained through local optimization. In other words, The asymptotic total J-divergence increases with increased channel SNR and can be optimized by minimizing the power of the transmit constellation. From this result, we can say that the local-channel power control scheme can be used to increase instantaneous J-divergence.

### 3.3.3 Long-Term System Performance Measure

The total J-divergence measure in (3.7) is an instantaneous performance measure because it depends on an instantaneous realization of the channel gain coefficients. The measure can be used as an objective function for power control schemes. In this section, we now employ classification outage as a long-term system performance measure. Classification outage is derived by applying the total J-divergence measure to the definition of the outage probability,  $p_{J_0} = \text{Prob}\{J^{total} < J_0\}$  defined in Chapter 2. This enables us to assess the average performance of our power control strategies over many realizations of the fading communication channel. For the proposed distributed multiple hypothesis detection system, the following theorem holds.

**Theorem 3.3.3** *In a homogeneous distributed multiple hypothesis detection system with a finite  $E[\eta_k^{total}]$ , for  $J_0 < J_\infty^{total}$  and a sufficiently-large  $\mathcal{K}$ , the outage probability is given as [42][41][45][46]*

$$p_{J_0} \sim \exp(-\mathcal{K}I_\eta(c))$$

or

$$-\log p_{J_0} \sim \mathcal{K}I_\eta(c)$$

where  $c = \frac{J_0}{P_{tot}}$ ,  $I_\eta(c) = \sup_{t \in R}(tc - \log M_\eta(t))$  with  $M_\eta(t)$ , the moment generating function of  $\eta$  which is an i.i.d random variable  $\eta_k^{total}$  over  $k$ , and  $\mathcal{K}$  determines the classi-

fication diversity order of the system.

The rate function  $I_\eta(c)$  is related to the following expression. For  $\{\eta_k^{total} : k \in \{1, \dots, \mathcal{K}\}\}$  that are i.i.d random variables,  $c < E\{\eta_k^{total}\}$  when  $J_0 < J_\infty^{total}$ . We then have

$$\lim_{\mathcal{K} \rightarrow \infty} -\frac{1}{\mathcal{K}} \log \left[ \text{Prob} \left\{ \frac{1}{\mathcal{K}} \sum_{k=1}^{\mathcal{K}} \eta_k^{total} < c \right\} \right] = I_\eta(c).$$

A detailed proof is shown in [45], [46], [41]. From an example of a Rayleigh fading channel system, the diversity gain can be analyzed as follows. The rate function,  $I_\eta(c)$ , is defined to be

$$I_\eta(c) = \sup_{t \in R} (tc - \log M_\eta(t))$$

where

$$\begin{aligned} \eta_k^{total} &= \sum_{i=1}^{\mathcal{H}} \sum_{j=i+1}^{\mathcal{H}} \eta_k^{ij} \cdot P(H_i) \cdot P(H_j) \\ &= \sum_{i=1}^{\mathcal{H}} \sum_{j=i+1}^{\mathcal{H}} \frac{s_k \phi_k^{ij}}{PW_k^*} \cdot P(H_i) \cdot P(H_j) \\ &= \frac{s_k}{2} \cdot \sum_{i=1}^{\mathcal{H}} \sum_{j=i+1}^{\mathcal{H}} \frac{2\phi_k^{ij}}{PW_k^*} \cdot P(H_i) \cdot P(H_j) \\ &= \frac{s_k}{2} \cdot \theta_k. \end{aligned}$$

The variable  $\eta_k^{total}$  is a weighted sum of random variables  $\eta_k^{ij}$ . However, for given sensor statistics and a given modulation scheme, it becomes an exponential random variable due to the fixed value  $\theta_k$ . Therefore,  $\eta_k^{total}$  is exponentially distributed with the mean  $\sigma^2 \theta_k$  where  $\sqrt{s_k}$  is i.i.d. Rayleigh-distributed with  $p_{\sqrt{s_k}}(t) = \frac{t}{\sigma^2} \exp\{-\frac{t^2}{2\sigma^2}\}$ . Since  $I_\eta(c)$  is

nonnegative and convex over  $c$ , we obtain

$$\begin{aligned} I_\eta(c) &= \sup_{t \in R} [tc + \log(1 - \sigma^2\theta \cdot t)] \\ &= \left(\frac{1}{\sigma^2\theta} - \frac{1}{c}\right)c + \log\left(\frac{\sigma^2\theta}{c}\right) \\ &= \frac{J_0}{P_{tot}\sigma^2\theta} - \log\left(\frac{J_0}{P_{tot}\sigma^2\theta}\right) - 1. \end{aligned}$$

As  $P_{tot} \rightarrow \infty$ , then  $\frac{J_0}{P_{tot}\sigma^2\theta} \rightarrow 0$ . Therefore,

$$\log p_{J_0} \sim -\mathcal{K} \cdot \log(P_{tot}). \quad (3.13)$$

This result is verified in Figures 3.2, 3.3, 3.4, and 3.5. The proportional gain,  $\mathcal{K}$ , is the classification diversity order, and if we increase the total power constraint  $P_{tot}$ , the reduction in outage probability in log scale is proportional to  $\mathcal{K}$ .

Figure 3.2 shows classification outage probability versus the total power constraint for varying number of sensors in a four-hypothesis distributed detection system. From the figure, we can see the benefit of local power control since channel power control has not been optimized. Local power gain is shown as a left shift of the curve. The figure was generated by simulating  $10^5$  independent realizations of the discrete source and the fading channel coefficients (or equivalently, the channel SNR values). For each realization, total J-divergence was calculated using (3.8). Finally, we counted the number of times the total J-divergence fell below a specified threshold. The following parameters were used to generate the results shown in Figure 3.2. The channel SNR was set according to  $s_k = \frac{G_0}{\sigma_k^2 \epsilon_k^2} |r_k|^2$  where  $\epsilon_k$  is the transmission distance from user  $k$  to the fusion center ( $\epsilon_k = 100$  m for all  $k$ 's),  $\sigma_k^2 = -70$  dBm is the channel noise power for all  $k$ 's,  $G_0 = -30$  dB is the nominal gain at the unit distance  $\epsilon_k = 1$  m, and the  $|r_k|$ 's are i.i.d. Rayleigh

fading random variables with unit variance. We further set ten arbitrary, but different, local transition matrices  $\{\mathbf{L}_k | k = 1, 2, \dots, 9\}$ , which implies a heterogeneous sensor environment. The outage threshold was set to  $J_0 = 0.05$ . In this example, we considered amplitude modulation and four hypotheses. After estimating each of the local decision probabilities  $\tilde{P}(u_k)$  with 100 local decision samples, local power control is performed by minimizing the average constellation power according to

$$\min_{\delta_k(1), \delta_k(2), \dots, \delta_k(\mathcal{H})} PW_k = \sum_{l=1}^{\mathcal{H}} |\delta_k(l)|^2 \cdot \tilde{P}(u_k = l). \quad (3.14)$$

The symbol constellation was constrained such that the minimum distance among the four symbols was unity. Therefore, we can set  $[\delta_k(1), \delta_k(2), \delta_k(3), \delta_k(4)] = [\tau, \tau + 1, \tau + 2, \tau + 3]$  where the design parameter is  $\tau$ . We also ordered the symbols in descending order of estimated prior probability. This is a logical step since minimum power should be allocated to the most frequently occurring symbols and vice versa. After substituting the optimal symbol constellation into (3.8), results for five and ten nodes using only local power control were generated in Figure 3.2.

It is seen that, at reasonably high  $P_{tot}$ , the slope of the ten-node case is two times larger than the slope of the five-node case. In the next subsection, we specifically address local power control for a system with multi-dimensional local modulation scheme.

### 3.3.4 Total J-divergence with $\mathcal{D}$ -dimensional Local Modulation

For a more general case, we can use a multi-dimensional ( $\mathcal{D}$ -dimensional) local modulation scheme such as QAM, FSK, or any arbitrary modulation. To explicitly handle multi-dimensional modulation, we modify the observation vector  $\mathbf{y}$  to become  $\mathcal{K} \times \mathcal{D}$  matrix  $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{\mathcal{D}}]$  where  $\mathcal{K}$  is still the number of sensors and  $\mathcal{D}$  is the maximum dimension of transmitted symbols. Thus, a matrix-variate normal distribu-

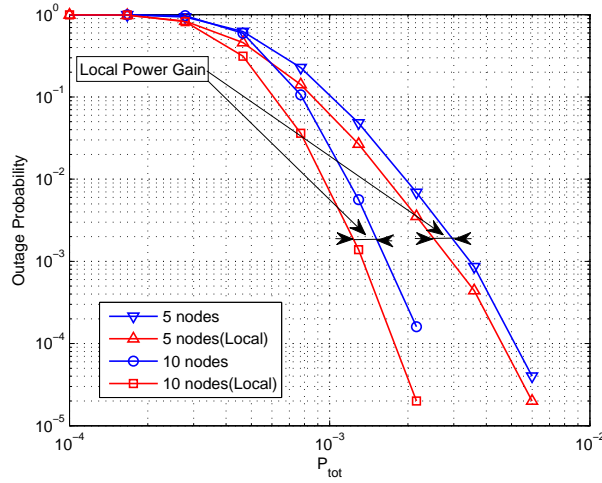


Figure 3.2: Outage probability versus  $P_{tot}$  with the uniform channel power transmission strategy in heterogeneous sensor configuration. "Local" in parenthesis means that Local power control is performed

tion for  $P(\mathbf{Y}|H_i)$  should be considered, but each column of the observation matrix  $\mathbf{Y}$  is independent of every other column under the Gaussian assumption used in (3.6). This implies  $J(p(\mathbf{Y}|H_i), p(\mathbf{Y}|H_j)) = \sum_{d=1}^{\mathcal{D}} J(p(\mathbf{y}_d|H_i), p(\mathbf{y}_d|H_j))$  since  $P(\mathbf{Y}|H_i) = P(\mathbf{y}_1|H_i)P(\mathbf{y}_2|H_i) \dots P(\mathbf{y}_{\mathcal{D}}|H_i)$  where the vector  $\mathbf{y}_d$  is the  $\mathcal{K}$ -dimensional observation.

Each optimal symbol component  $\delta_{(k,d)}^*(l)$  is transmitted with uniform power  $\frac{P_{tot}}{\mathcal{K} \cdot \mathcal{D}}$  since the dimensional components  $d$  are also independent  $\{P_{(k,d)} = \frac{P_{tot}}{\mathcal{K} \cdot \mathcal{D}} | k = 1, 2, \dots, \mathcal{K}, d = 1, 2, \dots, \mathcal{D}\}$ .  $l$  is the local decision index at each sensor. The system works like a system with  $\mathcal{K} \cdot \mathcal{D}$  independent components due to independent channels and independent symbol dimensional components and still satisfies the classification diversity  $\mathcal{K}$  for the given local modulation scheme ( $\mathcal{D}$  is a fixed value). For a given local power constraint, we can get better local power gain than that of the scalar modulation because the multi-dimensional modulation provides more geometrical distances between symbols. Recently, software defined radio furnishes flexible modulation schemes that give additional source power gain. Finally, from  $J(p(\mathbf{Y}|H_i), p(\mathbf{Y}|H_j)) = \sum_{d=1}^{\mathcal{D}} J(p(\mathbf{y}_d|H_i), p(\mathbf{y}_d|H_j))$

and (3.16), the total J-divergence with multi-dimensional local modulation is

$$\begin{aligned}
J^{total} &= \frac{1}{2} \sum_{d=1}^{\mathcal{D}} \sum_{i=1}^{\mathcal{H}} \sum_{j=i+1}^{\mathcal{H}} \left[ \sum_{k=1}^{\mathcal{K}} \frac{P_{totS_k} \{ \psi_{j,(k,d)} - \psi_{i,(k,d)} + \phi_{(k,d)}^{ij} \}}{\mathcal{K} \cdot \mathcal{D} \cdot \sum_{l=1}^{\mathcal{H}} |\delta_{(k,d)}^*(l)|^2 \cdot P(u_k = l) + P_{totS_k} \{ \psi_{i,(k,d)} \}} \right. \\
&\quad \left. + \sum_{k=1}^{\mathcal{K}} \frac{P_{totS_k} \{ \psi_{i,(k,d)} - \psi_{j,(k,d)} + \phi_{(k,d)}^{ij} \}}{\mathcal{K} \cdot \mathcal{D} \cdot \sum_{l=1}^{\mathcal{H}} |\delta_{(k,d)}^*(l)|^2 \cdot P(u_k = l) + P_{totS_k} \{ \psi_{j,(k,d)} \}} \right] \cdot P(H_i) \cdot P(H_j) \\
&= \frac{1}{2} \sum_{i=1}^{\mathcal{H}} \sum_{j=i+1}^{\mathcal{H}} \sum_{d=1}^{\mathcal{D}} \sum_{k=1}^{\mathcal{K}} \left[ \frac{P_{totS_k} \{ \psi_{j,(k,d)} - \psi_{i,(k,d)} + \phi_{(k,d)}^{ij} \}}{\mathcal{K} \cdot \mathcal{D} \cdot PW_{(k,d)}^* + P_{totS_k} \{ \psi_{i,(k,d)} \}} \right. \\
&\quad \left. + \frac{P_{totS_k} \{ \psi_{i,(k,d)} - \psi_{j,(k,d)} + \phi_{(k,d)}^{ij} \}}{\mathcal{K} \cdot \mathcal{D} \cdot PW_{(k,d)}^* + P_{totS_k} \{ \psi_{j,(k,d)} \}} \right] \cdot P(H_i) \cdot P(H_j) \\
\text{where } \psi_{i,(k,d)} &= \sum_{l=1}^{\mathcal{H}} \sum_{m=l+1}^{\mathcal{H}} [(\delta_{(k,d)}^*(l) - \delta_{(k,d)}^*(m))^2 \cdot P(u_k = l|H_i) \cdot P(u_k = m|H_i)], \\
\phi_{(k,d)}^{ij} &= \sum_{l=1}^{\mathcal{H}} [\delta_{(k,d)}^*(l) \cdot (P(u_k = l|H_i) - P(u_k = l|H_j))]^2,
\end{aligned} \tag{3.15}$$

$PW_{(k,d)}^* = \sum_{l=1}^{\mathcal{H}} |\delta_{(k,d)}^*(l)|^2 P(u_k = l)$ ,  $d$  is the dimensional index of the transmission symbol,  $i, j$  are the indices of multiple hypotheses,  $k$  is the sensor or channel node index, and  $l, m$  are the hypothesis indices of transmission symbols.

Similar system parameters from Figure 3.2 were used to generate Figure 3.3 for a multi-dimensional system. Differences include the threshold  $J_0 = 0.1$  and the 2D symbol constellation (QAM), which has less average symbol power than the 1D-symbol transmission system. To make a fair and simple comparison, we apply a simple structural constraint to the local power control by setting  $\delta_k(l = 1) = (\zeta - \tau, \zeta - \tau)$ ,  $\delta_k(l = 2) = (-\zeta - \tau, \zeta - \tau)$ ,  $\delta_k(l = 3) = (-\zeta - \tau, -\zeta - \tau)$ , and  $\delta_k(l = 4) = (\zeta - \tau, -\zeta - \tau)$  where  $\zeta$  is a given constant and the design parameter is  $\tau$ . The square QAM structure can be translated based on the local decision probabilities.  $\tau$  can be obtained based on local

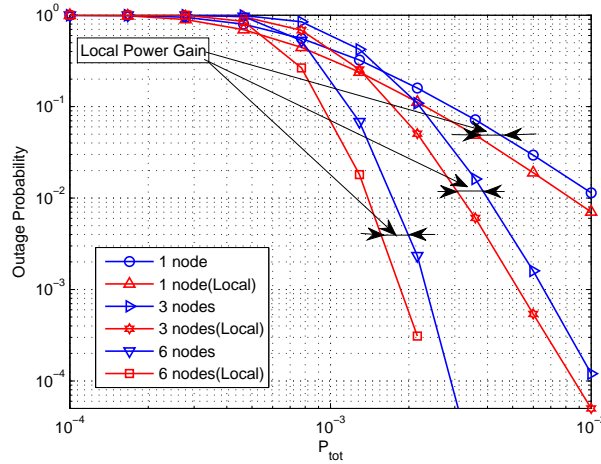


Figure 3.3: Outage probability versus  $P_{tot}$  with the uniform channel power transmission, 2D-local modulation, and heterogeneous sensor configuration

decision statistics  $P(u_k = l)$ , which then defines the optimal symbol constellation.

In Figure 3.3, it is seen that, at reasonably high  $P_{tot}$ , the slope of the outage probability is proportional to the number of nodes in the system. The power gain achieved by local optimization manifests itself as a shift of the curve. An interesting point is that this local power gain is achievable even for a single-node system, and the shifts for one, three, and six nodes are approximately equal. That is to say, the gain obtained through local power control is not related to the number of nodes, but instead is strongly related to the a priori source statistics. However, full classification diversity order is observed even under uniform average power allocation. Although classification diversity order is related to large- $\mathcal{K}$  asymptotic divergence, it is seen in Figure 3.3 that even systems with small  $\mathcal{K}$  (such as one or three) achieve full diversity order. Although we have only derived diversity order using asymptotic total J-divergence for a homogeneous sensor environment, it can be seen that classification diversity also applies to a heterogeneous sensor environment. In the next section, we simultaneously apply both power control strategies - local power control and channel power control - to a distributed multiple hypothesis detection system.

### 3.4 Local-Channel Power Transmission Strategy

#### 3.4.1 Optimization with a Total Power Constraint

If local decision probabilities are available at each local sensor, the sensor can adjust its local transmitting symbol in an optimal way. On the other hand, if the propagation channel states are also known, then the optimized symbol constellation of each sensor can also be optimally amplified within given total and individual power constraints. Such a strategy is formulated as follows. Recall that the average transmit power transmitted by the  $k$ th sensor is  $\{P_{(k,d)}|d = 1, 2, \dots, \mathcal{D}\} = \{PW_{(k,d)}^* \cdot a_{(k,d)}^2|d = 1, 2, \dots, \mathcal{D}\} = \{\sum_{l=1}^{\mathcal{H}} |\delta_{(k,d)}^*(l)|^2 \cdot P(u_k = l) a_{(k,d)}^2|d = 1, 2, \dots, \mathcal{D}\}$  where  $\mathcal{D}$  is the dimensionality of the symbol constellation. In the previous section, the average power  $P_k$  of each sensor was constrained to be equal, meaning the amplifying factor was decided based on the optimized symbol constellation. In this section, we use multi-dimensional average powers  $P_{(k,d)}$  of each sensor as additional design variables. This additional freedom leads to an optimization problem defined as

$$\begin{aligned}
\max \quad J^{total}(\cdot) &= \frac{1}{2} \sum_{d=1}^{\mathcal{D}} \sum_{i=1}^{\mathcal{H}} \sum_{j=i+1}^{\mathcal{H}} \left[ \sum_{k=1}^{\mathcal{K}} \frac{P_{(k,d)} s_k \{\psi_{j,(k,d)} - \psi_{i,(k,d)} + \phi_{(k,d)}^{ij}\}}{\sum_{l=1}^{\mathcal{H}} |\delta_{(k,d)}^*(l)|^2 \cdot P(u_k = l) + P_{(k,d)} s_k \{\psi_{i,(k,d)}\}} \right. \\
&\quad \left. + \sum_{k=1}^{\mathcal{K}} \frac{P_{tot} s_k \{\psi_{i,(k,d)} - \psi_{j,(k,d)} + \phi_{(k,d)}^{ij}\}}{\sum_{l=1}^{\mathcal{H}} |\delta_{(k,d)}^*(l)|^2 \cdot P(u_k = l) + P_{(k,d)} s_k \{\psi_{j,(k,d)}\}} \right] \cdot P(H_i) \cdot P(H_j) \\
&= \frac{1}{2} \sum_{i=1}^{\mathcal{H}} \sum_{j=i+1}^{\mathcal{H}} \sum_{d=1}^{\mathcal{D}} \sum_{k=1}^{\mathcal{K}} \left[ \frac{P_{(k,d)} s_k \{\psi_{j,(k,d)} - \psi_{i,(k,d)} + \phi_{(k,d)}^{ij}\}}{PW_{(k,d)}^* + P_{(k,d)} s_k \{\psi_{i,(k,d)}\}} \right. \\
&\quad \left. + \frac{P_{(k,d)} s_k \{\psi_{i,(k,d)} - \psi_{j,(k,d)} + \phi_{(k,d)}^{ij}\}}{PW_{(k,d)}^* + P_{(k,d)} s_k \{\psi_{j,(k,d)}\}} \right] \cdot P(H_i) \cdot P(H_j) \quad (3.16)
\end{aligned}$$

$$\text{s.t.} \quad \sum_{d=1}^{\mathcal{D}} \sum_{k=1}^{\mathcal{K}} P_{(k,d)} \leq P_{tot},$$

$$P_{(k,d)} \geq 0, k = 1, \dots, \mathcal{K} \quad \text{and} \quad d = 1, \dots, \mathcal{D}$$

$$\begin{aligned}
\text{where} \quad \psi_{i,(k,d)} &= \sum_{l=1}^{\mathcal{H}} \sum_{m=l+1}^{\mathcal{H}} [(\delta_{(k,d)}^*(l) - \delta_{(k,d)}^*(m))^2 \cdot P(u_k = l | H_i) \cdot P(u_k = m | H_i)], \\
\phi_{(k,d)}^{ij} &= \sum_{l=1}^{\mathcal{H}} [\delta_{(k,d)}^*(l) \cdot (P(u_k = l | H_i) - P(u_k = l | H_j))]^2.
\end{aligned}$$

The optimized constellation power of the  $k$ th sensor is  $\{PW_{(k,d)}^* | d = 1, 2, \dots, \mathcal{D}\}$ . The average transmitted power in the  $d$ th dimension is  $P_{(k,d)}^* = a_{(k,d)}^2 \cdot PW_{(k,d)}^*$  where the asterisk means optimized value. Therefore, the amplifying factor  $a_{(k,d)}$  is controlled by  $\delta_{(k,d)}(l)^*$  and  $P_{(k,d)}^*$ , which are respectively the local power control and channel power control. Therefore, we can optimize system performance by performing both local power control and channel power control.

For a scalar modulation system, once the  $P_k^*$ 's are found, the amplification factors  $a_k$ 's can be found. The objective function is non-decreasing with increasing  $P_{tot}$  because

$$\left\{ \frac{\partial}{\partial P_k} J^{total}(P_1, P_2, \dots, P_{\mathcal{K}}) \geq 0 | k = 0, 1, \dots, \mathcal{K} \right\} \quad (3.17)$$

where the  $P_k$ 's are non-negative. A detailed proof is given in Appendix B.5. In the multi-dimensional modulation case, the total J-divergence  $J^{total}(P_{(1,d)}, P_{(2,d)}, \dots, P_{(\mathcal{K},d)})$ , is still non-decreasing since it is a linear sum of non-decreasing scalar modulation systems. Therefore, we can get the optimal power allocation at the boundary of the power constraint,  $\sum_{k=1}^{\mathcal{K}} \sum_{d=1}^{\mathcal{D}} P_{(k,d)} = P_{tot}$  by a convex optimization algorithm for the concave region of the object function[1] [22]. Though the objective function is not necessarily concave in general, for the parameters of interest in this chapter all of the objective functions are concave and we apply the following convex optimization algorithm to find the optimal power allocation.

To get the optimal solution, we apply Lagrange multipliers to (3.16) to obtain

$$L(\{P_{(k,d)}|k = 1, 2, \dots, \mathcal{L}, d = 1, 2, \dots, \mathcal{D}\}) = J^{total}(\cdot) - \lambda \cdot \left( \sum_{d=1}^{\mathcal{D}} \sum_{k=1}^{\mathcal{K}} P_{(k,d)} - P_{tot} \right) + \sum_{d=1}^{\mathcal{D}} \sum_{k=1}^{\mathcal{K}} \nu_{d,k} P_{(k,d)} \quad (3.18)$$

where  $\lambda$  and  $\{\nu_{d,k}|k = 1, 2, \dots, \mathcal{L}, d = 1, 2, \dots, \mathcal{D}\}$  are Lagrange multipliers. After applying the derivative to (3.18), we can get the Karush-Kuhn-Tucker conditions and calculate the optimal power in the same manner with [1]. In cases where the objective function is not concave, another technique such as the interior point method [1][47][48] must be used.

For a simulation result, we use the same setup as that of Figure 3.3, except we assume

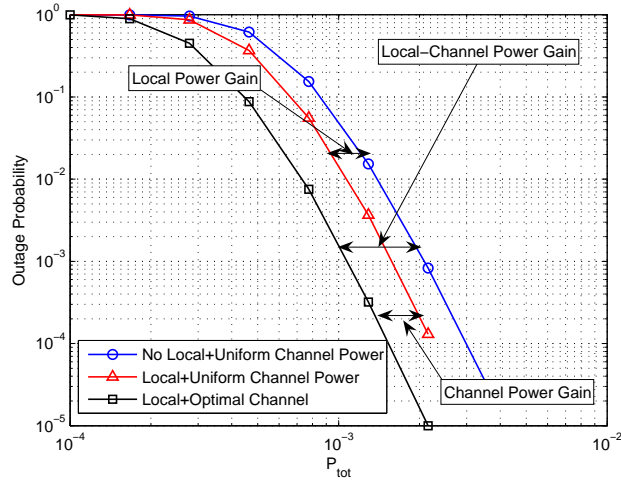


Figure 3.4: Outage probability versus  $P_{tot}$ . local and channel power control strategies are applied in a six-node system.

homogeneous local sensor statistics according to the transition matrices

$$\mathbf{L}_k = \begin{bmatrix} 0.7 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.7 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.7 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.7 \end{bmatrix} \text{ for all } k,$$

and we optimize the local+channel power control according to (3.16).

Figure 3.4 shows numerically generated classification outage probability versus the total power constraint for the three different optimization strategies in a six-node system. The three strategies are: 1) no power control, 2) local power control (symbol constellation optimization) with equal transmit power (from the previous section), and 3) optimization of both the symbol constellations and the amplification factors (local and channel power optimization). Note that all three curves have the same asymptotic slope but have different translations due to different power optimization schemes. Therefore, each system has the same diversity order, but for a given value of the power constraint, the curve

corresponding to local-channel power control has the lowest probability of outage.

### 3.4.2 Optimization with Individual and Total Power Constraints

In a practical system, each sensor may have an individual power constraint imposed by its battery or certain transmission regulations. The total power constraint above does not model this scenario; therefore, we now modify the optimization problem to include individual power constraints as well. Define a new optimization problem as

$$\begin{aligned}
 & \max && J^{total}(\cdot) \\
 & \text{s.t.} && \sum_{k=1}^{\mathcal{D}} \sum_{d=1}^{\mathcal{K}} P_{(k,d)} \leq P_{tot}, \\
 & && 0 \leq P_{(k,d)} \leq P_k^{max}(d), \quad k = 1, \dots, \mathcal{K} \text{ and } d = 1, \dots, \mathcal{D} \quad (3.19)
 \end{aligned}$$

where the optimization is again over the  $P_{(k,d)}$ 's and  $P_k^{max}(d)$  is the individual maximum allowed power. Since the additional power constraints on the individual sensors are linear constraints, there is no change on the conditions given for (3.16), such that the optimal solution can be similarly solved.

To find the solution to (3.19), the  $P_{(k,d)}$ 's are first optimized with the total power constraint, but without the individual power constraints. If some of the  $P_{(k,d)}$ 's are more than their upper limits,  $P_k^{max}(d)$ , then the optimal power for that sensor lies on the boundary of the individual power constraint and they are forced to equal  $P_k^{max}(d)$ . These sensors are then removed from  $P_{tot}$  and the optimization procedure continues with the remaining sensors. This iteration continues until all power constraints are satisfied and a global optimum is obtained [41]. Figure 3.5 compares classification outage probability for the case with both total and individual power constraints to the case with only the total power constraints in a nine-node system. We see that with the additional individual power con-

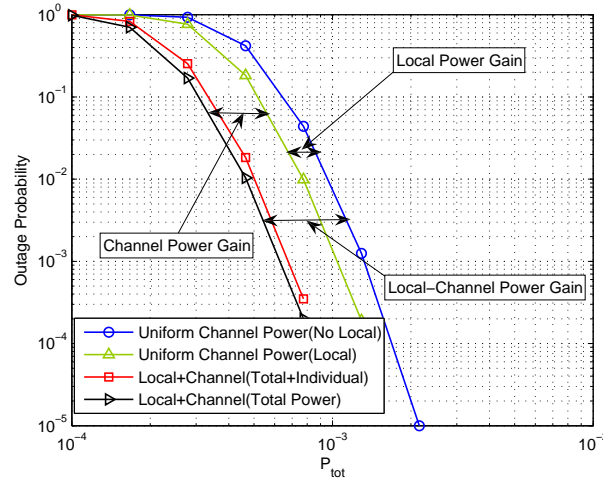


Figure 3.5: Outage probability of an optimal system with individual power constraints and a total power constraint in a nine-node system.

straints, full diversity is still achieved but the power gain is reduced compared with the case without individual power constraints. We consider the power allocation to each sensor in the next subsection.

### 3.4.3 Power distribution across sensors

In Table 3.1, homogeneous sensor statistics are adopted in order to clearly show the effect of the local power control. We define five different total power constraints in increasing order according to  $P_1 < P_2 < P_3 < P_4 < P_5$ . We also order 10 sensors in decreasing order of channel strength. In other words, the channel gain coefficient is strongest for sensor 1 and weakest for sensor 10. When we apply the lowest power constraint, all of the transmitted power is focused on sensor 1 because power is extremely limited and the first sensor's channel has the highest SNR. As more power is allowed, the power begins to be distributed to the next highest SNR channel, then the next highest, and so on. Finally, for the highest total power constraint, sensor 10 is utilized, but only for the system with local power optimization (LP). In fact, at every power level except the weakest, the system

Table 3.1: Power distribution by local power control in percentage of total power allocated to each sensor

$P_{tot}$	Scheme	Sensor										
		1	2	3	4	5	6	7	8	9	10	
P1	NLP	100	0	0	0	0	0	0	0	0	0	0
	LP	100	0	0	0	0	0	0	0	0	0	0
P2	NLP	43.17	34.41	22.42	0	0	0	0	0	0	0	0
	LP	37.47	32.01	23.50	7.02	0	0	0	0	0	0	0
P3	NLP	28.04	25.76	20.98	11.21	6.79	3.82	3.39	0	0	0	0
	LP	23.64	22.50	19.17	12.04	8.76	6.55	6.23	1.11	0	0	0
P4	NLP	21.32	20.71	18.07	12.22	9.50	7.66	7.40	3.12	0	0	0
	LP	18.15	18.22	16.50	12.36	10.39	9.05	8.85	5.71	0.78	0	0
P5	NLP	17.71	17.86	16.25	12.33	10.45	9.17	8.98	5.98	1.26	0	0
	LP	15.04	15.60	14.62	11.90	10.56	9.63	9.50	7.30	3.83	2.03	0

with local power control is able to exploit one additional sensor compared to the system with no local power control (NLP). Power distribution in the absence of local power control is explained in detail in [1]. The main goal of Table 3.1 is to show the effect of the local power control. By comparing LP with NLP for the same total power constraint, we can see that the system with local power control can often exploit more sensors than the system without local power control, thereby achieving higher power gain and improved outage probability. The reason is that the optimized local constellation consumes lower power for the same detection performance.

### 3.5 Conclusion of Power Optimization for Sensor Network

Chapters 2 and 3 consider sensor power optimization problem and outage probability as a long-term performance measure. In Chapter 2, we investigated a distributed binary detection system in a slow-fading environment, where each sensor amplifies and transmits the local decision to a fusion center. We introduced two new concepts termed as outage

probability and its diversity to quantify the long-term system performance. The power control strategy was analyzed with asymptotic J-divergence and the relationship between asymptotic J-divergence and detection outage probability was showed analytically. This relationship leads to the classification diversity gain under the homogenous sensor environment. We proved that for a homogeneous system with a uniform transmission strategy, a full diversity gain of order  $\mathcal{K}$  is achievable where  $\mathcal{K}$  is the total number of nodes in the network. We further showed that by optimally adjusting the transmission power for each node, an adaptive power gain is obtained in addition to the full diversity gain. We also showed that with extra individual power constraints, the power gain will be reduced while the full diversity gain is still achievable.

In Chapter 3, a local-channel power control scheme applicable to distributed multiple hypothesis detection systems in slow-fading environments is introduced. Instead of using J-divergence, we defined and used total J-divergence measure for distributed multiple hypothesis system optimization. We generalized the outage probability of Chapter 2 for the multiple-hypothesis problem and showed that the classification diversity still holds. We showed via simulation that classification diversity is observed even in heterogeneous sensor environment. Then, using the classification outage probability as a long-term performance measure, we showed that the distributed multiple hypothesis system is efficiently improved through both local power optimization and channel power optimization. The local power optimization is based on local decision statistics and results in optimized transmit symbol constellations while the channel power optimization is based on channel fading states and results in optimized amplifying factors. Individual power constraints were also considered.

## CHAPTER 4

## WAVEFORM DESIGN WITH TASK-SPECIFIC INFORMATION

In this chapter, we discuss Waveform Design by using Task-Specific Information which is an efficient and direct classification measure for radar system.

## 4.1 System Model

Consider a discrete, frequency-domain measurement model where the  $l^{th}$  measurement is

$$y(l) = w(l)g(l) + n(l), \quad (4.1)$$

where  $w(l)$  is the  $l^{th}$  frequency-domain waveform coefficient,  $g(l)$  is the  $l^{th}$  target transfer function coefficient, and  $n(l)$  is additive Gaussian noise. For measurements taken at  $\mathcal{L}$  frequencies, the  $\mathcal{L} \times 1$  measurement vector is then  $\mathbf{y} = \mathbf{W}\mathbf{g} + \mathbf{n}$  where  $\mathbf{y} = [y(1), y(2), \dots, y(\mathcal{L})]^T$ ,  $\mathbf{g} = [g(1), g(2), \dots, g(\mathcal{L})]^T$ , the  $\mathcal{L} \times \mathcal{L}$  waveform matrix  $\mathbf{W}$  is diagonal with entries  $w(1), w(2), \dots, w(\mathcal{L})$ , and  $\mathbf{n}$  is a vector of independent Gaussian noise samples. For the case where we are trying to identify one of  $\mathcal{H}$  known hypotheses, the target transfer function can take one of  $\mathcal{H}$  possible realizations. Let these known realizations be  $\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_{\mathcal{H}}$ , and define the  $\mathcal{L} \times \mathcal{H}$  target transfer matrix as  $\mathbf{G} = [\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_{\mathcal{H}}]$ . The measurements can now be described by  $\mathbf{y} = \mathbf{W}\mathbf{G}\mathbf{x} + \mathbf{n}$ . The vector  $\mathbf{x}$  is an indicator vector  $\{\mathbf{x} = \mathbf{e}_h | h = 1, 2, \dots, \mathcal{H}\}$  where  $\mathbf{e}_h$  is a vector with a one in the  $h^{th}$  position and zeros elsewhere. For example, if the first hypothesis is the true hypothesis, then  $\mathbf{x} = \mathbf{e}_1$ , which selects the first column of  $\mathbf{G}$ . The pdf of the measurement

vector conditioned on  $\mathbf{x}$  is

$$\begin{aligned} p(\mathbf{y}|\mathbf{x}) &= \frac{1}{(2\pi\sigma^2)^{\frac{L}{2}}} \exp \left[ -\frac{(\mathbf{y} - \mathbf{W}\mathbf{G}\mathbf{x})^T (\mathbf{y} - \mathbf{W}\mathbf{G}\mathbf{x})}{2\sigma^2} \right] \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{L}{2}}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^{\mathcal{L}} (y(i) - w(i)\{\mathbf{G}\mathbf{x}\}(i))^2 \right]. \end{aligned}$$

There are  $\mathcal{H}$  possible realizations of  $\mathbf{x}$  - one realization for each target hypothesis. Thus, the virtual source variable  $\mathbf{x}$  encodes the classification task into the signal model and correctly detecting  $\mathbf{x}$  is equivalent to correctly detecting which target is present. If the hypotheses are equally likely, then the task-specific entropy prior to taking any measurements is  $H(\mathbf{x}) = \log_2(\mathcal{H})$  bits, which is the maximum information that can be produced by any observation. Earlier information-based waveform design methods based for estimation use the mutual information between  $\mathbf{y}$  and the target transfer function ensemble  $\mathbf{G}$ . While it is also true, for our signal model, that correctly estimating the transfer function is equivalent to correctly detecting which target is present, the MI between measurements and transfer function ensemble assumes that all frequency components carry independent information. For stationary Gaussian random processes, this is true, but there are an infinite number of realizations of a Gaussian random process and the entropy is infinite. In contrast, a scenario where the decision to be made is from a finite set of hypotheses implies that some measurements carry redundant information. For example, suppose a particular element of the measurement vector can be measured with such high fidelity that the correct target hypothesis can be detected with negligible error. The remaining elements of the measurement vector add no additional *task-specific* information because they can be inferred from the decision and known structure of the finite number of signals. By introducing a source variable, we enable the correlation between measurements that

is present in a finite-hypothesis scenario.

Using the above measurement model, TSI is defined as the mutual information between measurements  $\mathbf{y}$  and source variable  $\mathbf{x}$ ,  $TSI = I(\mathbf{y}; \mathbf{x})$ , and the waveform optimization problem is

$$\begin{aligned} \max \quad & TSI = I(\mathbf{y}; \mathbf{x}) \\ \text{s.t.} \quad & Tr[\mathbf{W}\mathbf{W}^T] \leq E_w, \end{aligned} \quad (4.2)$$

where  $Tr[\cdot]$  is the matrix trace and  $E_w$  is a constraint on the allowable waveform energy. There is no closed-form solution to this optimization problem because  $\mathbf{y}$  is a Gaussian mixture. To find the waveform  $\mathbf{W}$  that maximizes TSI, we adopt the gradient search method using the gradient of TSI with respect to the waveform as described in the next section.

#### 4.2 Waveform design with Task-Specific Information for Monostatic Radar

In this section, we focus on waveform design with TSI for monostatic radar. In the optimization problem of (4.2), TSI is used as the objective function. We use a gradient search method to find the waveform energy spectrum that maximizes TSI. The derivative of  $TSI$  with respect to the  $i^{th}$  frequency component,  $w(i)$  is

$$\begin{aligned} \frac{\partial TSI}{\partial w(i)} &= \frac{\partial [h(\mathbf{y}) - h(\mathbf{y}|\mathbf{x})]}{\partial w(i)} \\ &= \frac{\partial [-E[\log_2 p(\mathbf{y})] + E[\log_2 p(\mathbf{y}|\mathbf{x})]]}{\partial w(i)} \\ &= -\frac{\partial E[\log_2 p(\mathbf{y})]}{\partial w(i)}. \end{aligned} \quad (4.3)$$

Continuing with (4.3) for the given monostatic system model, we obtain

$$\frac{\partial TSI}{\partial \widetilde{\mathbf{W}}} = \frac{\mathbf{W}}{\sigma^2} \mathbb{E}_{\mathbf{x}, \mathbf{y}} [\mathit{diag}\{\mathbf{G}\mathbf{x}\} \cdot \mathbf{G}\mathbf{x} - \mathbb{E}_{\mathbf{x}|\mathbf{y}} [\mathit{diag}\{\mathbf{G}\mathbf{x}\}] \cdot \mathbb{E}_{\mathbf{x}|\mathbf{y}} [\mathbf{G}\mathbf{x}]] \quad (4.4)$$

where  $\frac{\partial TSI}{\partial \widetilde{\mathbf{W}}} = \left[ \frac{\partial TSI}{\partial w(1)}, \frac{\partial TSI}{\partial w(2)}, \dots, \frac{\partial TSI}{\partial w(L)} \right]^T$ ,  $\mathit{diag}\{\mathbf{G}\mathbf{x}\}$  is a diagonal matrix including the vector  $\mathbf{G}\mathbf{x}$  on the main diagonal,  $\mathbb{E}_{\mathbf{x}|\mathbf{y}} [\mathbf{G}\mathbf{x}]$  is calculated by

$$\mathbb{E}_{\mathbf{x}|\mathbf{y}} [\mathbf{G}\mathbf{x}] = \frac{\sum_{h=1}^{\mathcal{H}} \{\mathbf{G}\mathbf{x}_h\} p(\mathbf{y}|\mathbf{x}_h) P(\mathbf{x}_h)}{\sum_{h=1}^{\mathcal{H}} p(\mathbf{y}|\mathbf{x}_h) P(\mathbf{x}_h)},$$

and  $\mathbb{E}_{\mathbf{x}|\mathbf{y}} [\mathit{diag}\{\mathbf{G}\mathbf{x}\}]$  is similarly calculated. The detailed proof is given in Appendix C.1. As we see in (4.4), waveform design with TSI takes advantage of the correlated information between different frequency components by considering the overall shape of a target transfer vector  $\mathbf{G}\mathbf{x}$  triggered by the virtual source  $\mathbf{x}$ .

The primary difficulty with (4.4) is that there is no closed-form solution for the required expected values since the distribution of  $\mathbf{y}$  is a Gaussian Mixture. However, we can numerically calculate the expected value through Monte Carlo methods because we know the conditional probability distribution,  $p(\mathbf{y}|\mathbf{x})$  and a priori probability,  $P(\mathbf{x})$ . Thus, we can evaluate

$$\begin{aligned} & \frac{\partial TSI}{\partial \widetilde{\mathbf{W}}} \\ &= \frac{\mathbf{W}}{N_{total}\sigma^2} \sum_{\mathbf{y}_0, \mathbf{x}_0} \{ \mathit{diag}\{\mathbf{G}\mathbf{x}_0\} \cdot \{\mathbf{G}\mathbf{x}_0\} - \mathbb{E}_{\mathbf{x}|\mathbf{y}_0} [\mathit{diag}\{\mathbf{G}\mathbf{x}\}] \cdot \mathbb{E}_{\mathbf{x}|\mathbf{y}_0} [\mathbf{G}\mathbf{x}] \} \end{aligned} \quad (4.5)$$

where  $N_{total}$  is the number of Monte Carlo realizations necessary for estimating the required expected value,  $\mathbf{x}_0$  is one of the  $N_{total}$  samples drawn from  $P(\mathbf{x})$ ,  $\mathbf{y}_0$  is one of the

$N_{total}$  samples drawn from  $p(\mathbf{y})$ , and

$$\mathbb{E}_{\mathbf{x}|\mathbf{y}_0} [\mathbf{G}\mathbf{x}] = \frac{\sum_{h=1}^{\mathcal{H}} \{G\mathbf{x}_h\} p(\mathbf{y}_0|\mathbf{x}_h) P(\mathbf{x}_h)}{\sum_{h=1}^{\mathcal{H}} p(\mathbf{y}_0|\mathbf{x}_h) P(\mathbf{x}_h)}. \quad (4.6)$$

Equation (4.4) does not include the waveform energy constraint  $E_w$ . We next consider an  $\mathcal{L}$ -dimensional spherical coordinate system for defining an unconstrained optimization problem.

#### 4.2.1 $\mathcal{L}$ -dimensional spherical coordinate

The optimum waveform under a fixed energy constraint will lie on an  $\mathcal{L}$ -dimensional spheroid. We can exploit this fact by converting the waveform to an  $\mathcal{L}$ -dimensional spherical coordinate system according to

$$\begin{aligned} w(1) &= \sqrt{E_w} \cos \theta_1 \\ w(2) &= \sqrt{E_w} \sin \theta_1 \cos \theta_2 \\ &\vdots \\ w_{\mathcal{L}-1} &= \sqrt{E_w} \prod_{i=1}^{\mathcal{L}-2} \sin \theta_i \cos \theta_{\mathcal{L}-1} \\ w(\mathcal{L}) &= \sqrt{E_w} \prod_{i=1}^{\mathcal{L}-1} \sin \theta_i \end{aligned} \quad (4.7)$$

where  $E_w = \text{Tr}[\mathbf{W}\mathbf{W}^T] = \sum_{i=1}^{\mathcal{L}} (w(i))^2$ . After applying the coordinate transform of (4.7), we can parameterize the waveform coefficients with angles and the radius  $\{(\theta_i, E_w) | i = 1, 2, \dots, \mathcal{L} - 1\}$ . If we fix the energy constraint  $E_w$ , a spherical search space is parameterized only by  $(\mathcal{L} - 1)$  angles  $\{\theta_i | i = 1, 2, \dots, \mathcal{L} - 1\}$ . For example, a two-tap waveform is defined by one angle to a point on a circle with fixed radius by  $E_w$ , and a three-tap waveform is defined by two angles describing a point on the sphere of

radius  $E_w$ . Alternatively, we can view the spherical coordinate strategy as a projection of the original design parameters  $[w(1), w(2), \dots, w(\mathcal{L})]^T$  onto the fixed spherical surface. Therefore, we can establish an unconstrained optimization problem by applying (4.7) to the objective function of (4.2). However, we need to calculate the derivative of TSI with respect to the waveform angles, so we define a Jacobian transform from (4.7) and apply it to  $\frac{\partial TSI}{\partial \mathbf{W}}$  to get  $\frac{\partial TSI}{\partial \Theta}$  where  $\frac{\partial TSI}{\partial \theta} = [\frac{\partial TSI}{\partial \theta_1}, \frac{\partial TSI}{\partial \theta_2}, \dots, \frac{\partial TSI}{\partial \theta_{\mathcal{L}-1}}]^T$  and  $\Theta = [\theta_1, \theta_2, \dots, \theta_{\mathcal{L}-1}]^T$ . The unconstrained optimization problem can be solved by a gradient search algorithm. After obtaining the optimal angles  $\Theta$ , we convert them back to the  $\mathcal{L}$ -dimensional waveform  $\mathbf{W}$  by (4.7). The Jacobian matrix is

$$\mathbf{J}(\theta_1, \theta_2, \dots, \theta_{\mathcal{L}-1}, E_w) = \begin{pmatrix} \frac{\partial w(1)}{\partial \theta_1} & \frac{\partial w(1)}{\partial \theta_2} & \cdots & \frac{\partial w(1)}{\partial \theta_{\mathcal{L}-1}} & \frac{\partial w(1)}{\partial E_w} \\ \frac{\partial w(2)}{\partial \theta_1} & \frac{\partial w(2)}{\partial \theta_2} & \cdots & \frac{\partial w(2)}{\partial \theta_{\mathcal{L}-1}} & \frac{\partial w(2)}{\partial E_w} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{\partial w(\mathcal{L})}{\partial \theta_1} & \frac{\partial w(\mathcal{L})}{\partial \theta_2} & \cdots & \frac{\partial w(\mathcal{L})}{\partial \theta_{\mathcal{L}-1}} & \frac{\partial w(\mathcal{L})}{\partial E_w} \end{pmatrix}, \quad (4.8)$$

which when substituting (4.7) yields the lower-diagonal Jacobian matrix

$$\mathbf{J}(\theta_1, \theta_2, \dots, \theta_{\mathcal{L}-1}, E_w) = \begin{pmatrix} \frac{\partial w(1)}{\partial \theta_1} & 0 & \cdots & 0 \\ \frac{\partial w(2)}{\partial \theta_1} & \frac{\partial w(2)}{\partial \theta_2} & 0 & 0 \\ \vdots & \vdots & \ddots & 0 \\ \frac{\partial w(\mathcal{L})}{\partial \theta_1} & \frac{\partial w(\mathcal{L})}{\partial \theta_2} & \cdots & \frac{\partial w(\mathcal{L})}{\partial \theta_{\mathcal{L}-1}} \end{pmatrix}. \quad (4.9)$$

We then have  $\frac{\partial TSI}{\partial \Theta} = \mathbf{J}(\theta_1, \theta_2, \dots, \theta_{\mathcal{L}-1}, E_w)^T \cdot \frac{\partial TSI}{\partial \mathbf{W}}$ . The optimum waveform can now be found via gradient update programming according to

$$\Theta_{j+1} = \Theta_j + \lambda \left[ \frac{\partial TSI}{\partial \Theta} \right]_j \quad (4.10)$$

where  $\Theta_j = [\theta_j(1), \theta_j(2), \dots, \theta_j(\mathcal{L})]^T$ ,  $0 \leq \theta_j(l) \leq \pi/2$ , and  $j$  is an update-index of the gradient search method. This gradient update method is a unconstrained optimization since the energy constraint is already applied in the Jacobian matrix and  $\frac{\partial TSI}{\partial \Theta}$  is defined only on the energy-constrained sphere.  $\theta_j(l)$  is a non-negative angle less than  $\pi/2$  due to the optimality characteristics discussed in the next section.

We should note that we could perform the gradient search in the waveform coefficient domain according to

$$\begin{aligned} \widetilde{\mathbf{W}}_{j+1} &= \widetilde{\mathbf{W}}_j + \lambda \left[ \frac{\partial TSI}{\partial \widetilde{\mathbf{W}}} \right]_j \\ \text{s.t.} \quad &Tr[\mathbf{W}\mathbf{W}^T] \leq E_w. \end{aligned} \quad (4.11)$$

This approach, however, does not implement a waveform energy constraint and the waveform must be normalized back to  $E_w$  after each gradient update. The angle-based update method yields faster convergence, but we must be careful of numerical errors that appear during the Jacobian calculation and the transformation between the waveform coefficients and angles.

#### 4.2.2 Global Optimality of TSI Waveform Design

An important feature of the TSI-based waveform design metric is the ability to find a global maximum. In order to prove global optimality, we need to show that TSI is concave or has one continuous and smooth peak within the constrained feasible region. TSI can

be rewritten as follows.

$$\begin{aligned}
TSI &= I(\mathbf{y}; \mathbf{x}) = h(\mathbf{y}) - h(\mathbf{y}|\mathbf{x}) \\
&= -E[\log_2 p(\mathbf{y})] + E[\log_2 p(\mathbf{y}|\mathbf{x})] \\
&= -E[\log_2 p(\mathbf{y})] - \frac{1}{2} \log_2 [(2\pi e)^n |\mathbf{K}|]
\end{aligned} \tag{4.12}$$

where  $p(\mathbf{y}|\mathbf{x})$  is a multivariate normal distribution with covariance matrix  $\mathbf{K}$ . The second term  $h(\mathbf{y}|\mathbf{x})$  is a constant for the given covariance matrix. Thus, the shape of TSI depends only on  $h(\mathbf{y})$ , which depends on  $p(\mathbf{y})$ . Therefore, we need to consider the entropy of  $\mathbf{y}$  to prove global optimality, but  $h(\mathbf{y})$  is the entropy of a Gaussian Mixture Model (GMM), which is intractable. In this case, one can calculate the upper and lower bounds of the entropy, obtain approximated values by Taylor-series expansion [49], or simply employ a single Gaussian to approximate the GMM for a closed-form solution[1][42][50]. These methods provide approximate solutions based on techniques not used in the gradient search procedure. Thus, here we show empirical evidence of the global optimality of our TSI calculations. At first, we will demonstrate global optimality characteristics by showing TSI surfaces versus angles for two- and three-tap waveforms, then we provide some intuition as to why higher-dimensional TSI surfaces follow the same characteristics.

Figures 4.1 and 4.2 show examples of TSI numerically evaluated over the entire design space. The waveform of Figure 4.1 is defined by only two coefficients ( $\mathcal{L} = 2$ ), and with the spherical coordinate system, the two-dimensional waveform is parameterized by one angle. The waveform of Figure 4.2 is defined by only three coefficients ( $\mathcal{L} = 3$ ), which are parameterized by two angles in spherical coordinates. In Figure 4.2 we plot TSI versus  $(\theta_1, \theta_2)$  with ranges of  $([0^\circ, 360^\circ], [0^\circ, 180^\circ])$ , respectively. We only have to check  $[0^\circ, 180^\circ]$  for one of the angles to check the given entire energy sphere region. There are

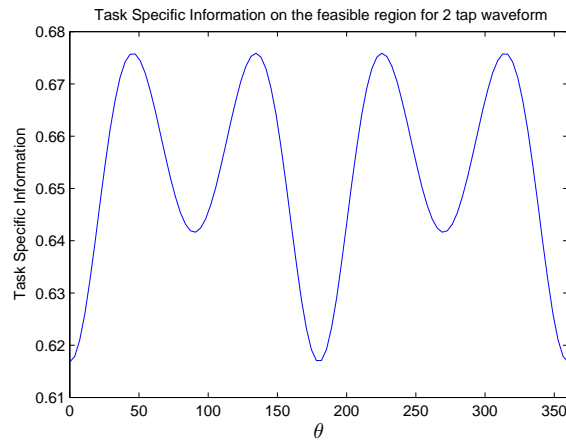


Figure 4.1: Plot of TSI on the feasible energy region for two tap waveform

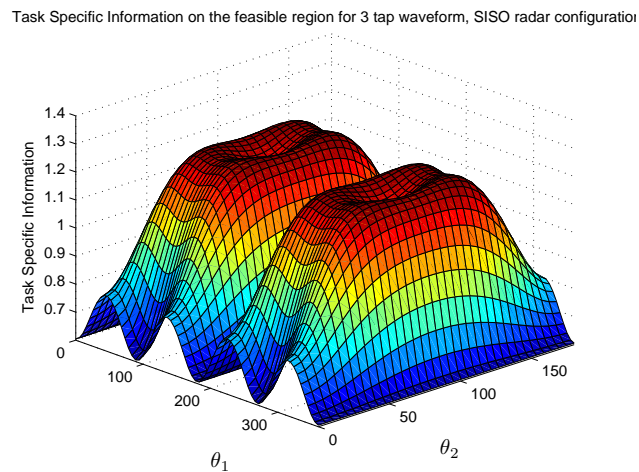


Figure 4.2: Plot of TSI on the feasible energy region for three tap waveform

four equally likely target hypotheses in both cases. The maximum TSI possible is two bits, and the waveform design spaces for the two figures are a circle and a sphere, respectively, both with radius  $\sqrt{E_w}$ . The maximum entropies of Figures 4.1 and 4.2 are located at approximately  $45^\circ$  and  $(63^\circ, 58^\circ)$ , respectively. The TSI that is achieved depends on the waveform energy constraint, target characteristics, and noise power.

The circular design space has four unique quadrants, and the spherical design space has eight unique quadrants. The waveform designs in the different quadrants differ only in

the signs of the waveform coefficients, and we see from the surfaces that TSI is symmetric in the sign of the waveform coefficients. We also see in Figures 4.1 and 4.2 that each quadrant has a single well-behaved optimum point. Based on the full two- and three-dimensional TSI surfaces, the TSI objective function is not concave, but the single peaks in each quadrant are symmetric. Empirically, it seems that we can restrict our search to the first quadrant of an appropriately dimensioned search space and expect to find a globally optimum solution.

### 4.3 Waveform design with Task-Specific Information for MIMO Radar

In this section, we describe waveform design with TSI for MIMO radar. Henceforth, we use the hat symbol  $\hat{\cdot}$  for a vector or matrix when denoting a single-input, multiple-output (SIMO) system and the tilde symbol  $\tilde{\cdot}$  for a MIMO system. We also use the symbol  $\otimes$  to denote the Kronecker product and  $\oplus$  to denote a direct sum operation.

We first develop TSI-based waveform design for a SIMO configuration and then extend the result to MIMO configurations. The SIMO system model is

$$\hat{\mathbf{y}} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_{N_{\text{Rx}}} \end{bmatrix} = \widehat{\mathbf{W}}\widehat{\mathbf{G}}\mathbf{x} + \widehat{\mathbf{n}} = \begin{bmatrix} \mathbf{W}\mathbf{G}_1\mathbf{x} + \mathbf{n}_1 \\ \mathbf{W}\mathbf{G}_2\mathbf{x} + \mathbf{n}_2 \\ \vdots \\ \mathbf{W}\mathbf{G}_{N_{\text{Rx}}}\mathbf{x} + \mathbf{n}_{N_{\text{Rx}}} \end{bmatrix} \quad (4.13)$$

where the  $\mathcal{L}N_{\text{Rx}} \times \mathcal{L}N_{\text{Rx}}$  matrix  $\widehat{\mathbf{W}} = (\mathbf{I}_{N_{\text{Rx}}} \otimes \mathbf{W})$ , the  $\mathcal{L}N_{\text{Rx}} \times \mathcal{H}$  matrix  $\widehat{\mathbf{G}} = [\mathbf{G}_1^T, \mathbf{G}_2^T, \dots, \mathbf{G}_{N_{\text{Rx}}}^T]^T$ ,  $\mathbf{G}_m$  is an  $\mathcal{L} \times \mathcal{H}$  matrix of transfer functions for each hypothesis and receiver index  $m = 1, 2, \dots, N_{\text{Rx}}$ , the  $\mathcal{H} \times 1$  indicator vector  $\mathbf{x}$  is the same as for the monostatic configuration, and the noise vector  $\widehat{\mathbf{n}}$  is a  $\mathcal{L}N_{\text{Rx}} \times 1$ .  $\mathbf{I}_{N_{\text{Rx}}}$  is an  $N_{\text{Rx}} \times N_{\text{Rx}}$  identity matrix and  $\mathbf{1}_{N_{\text{Rx}}}$  is an  $N_{\text{Rx}} \times 1$  vector of ones. The diagonal entries of the wave-

form matrix  $\mathbf{W}$  are the design parameters of the single transmitter.  $\mathbf{G}_m$  is the target transfer matrix that defines the path between the transmitter and the  $m^{\text{th}}$  radar receiver. The resulting SIMO TSI optimization problem is

$$\begin{aligned} \max \quad & TSI = I(\hat{\mathbf{y}}; \mathbf{x}) \\ \text{s.t.} \quad & Tr[WW^T] \leq E_w, \end{aligned} \quad (4.14)$$

where  $I(\hat{\mathbf{y}}; \mathbf{x}) = h(\hat{\mathbf{y}}) + h(\hat{\mathbf{y}}|\mathbf{x})$ . The procedure is similar to the monostatic case except that the waveform is optimized based on multiple target transfer functions arising from multiple bistatic paths, which yields the different conditional pdf:

$$\begin{aligned} p(\hat{\mathbf{y}}|\mathbf{x}) &= \left( \frac{1}{2\pi\sigma^2} \right)^{\frac{\mathcal{L}N_{\text{Rx}}}{2}} \exp \left[ -\frac{(\hat{\mathbf{y}} - \widehat{\mathbf{W}}\widehat{\mathbf{G}}\mathbf{x})^T (\hat{\mathbf{y}} - \widehat{\mathbf{W}}\widehat{\mathbf{G}}\mathbf{x})}{2\sigma^2} \right] \\ &= \left( \frac{1}{2\pi\sigma^2} \right)^{\frac{\mathcal{L}N_{\text{Rx}}}{2}} \exp \left[ -\frac{1}{2\sigma^2} \left\{ \sum_{m=1}^{N_{\text{Rx}}} \sum_{i=1}^{\mathcal{L}} \left( \hat{y}(i + \mathcal{L}(m-1)) - w(i) \cdot \{\widehat{\mathbf{G}}\mathbf{x}\}(i + \mathcal{L}(m-1)) \right)^2 \right\} \right] \end{aligned} \quad (4.15)$$

where  $\{\widehat{\mathbf{G}}\mathbf{x}\}(i)$  means the  $i^{\text{th}}$  component of the  $\mathcal{L}N_{\text{Rx}} \times 1$  vector  $\widehat{\mathbf{G}}\mathbf{x}$ . For example, a SIMO configuration with two receiving radars yields

$$\hat{\mathbf{y}} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} W & \mathbf{0} \\ \mathbf{0} & W \end{bmatrix} \begin{bmatrix} \mathbf{G}_1\mathbf{x} \\ \mathbf{G}_2\mathbf{x} \end{bmatrix} + \begin{bmatrix} \mathbf{n}_1 \\ \mathbf{n}_2 \end{bmatrix} \quad (4.16)$$

where the subscript denotes the receiver index,  $W$  is the  $\mathcal{L} \times \mathcal{L}$  waveform matrix, and  $\{\mathbf{y}_m, \mathbf{G}_m\mathbf{x}, \mathbf{n}_m | m = 1, 2\}$  are  $\mathcal{L} \times 1$  vectors. As seen in (4.16), the single transmit waveform affects both bistatic paths.

After differentiating the TSI function  $I(\widehat{\mathbf{y}}; \mathbf{x})$  with respect to  $w(i)$  and doing some algebra, we obtain the TSI gradient for the SIMO configuration as

$$\begin{aligned} & \frac{\partial TSI}{\partial \widetilde{\mathbf{W}}} \\ &= \frac{\mathbf{W}}{\sigma^2} \mathbb{E}_{\mathbf{x}, \mathbf{y}_m} \left[ \sum_{m=1}^{N_{\text{Rx}}} \left\{ \text{diag}\{\mathbf{G}_m \mathbf{x}\} \cdot \{\mathbf{G}_m \mathbf{x}\} - \mathbb{E}_{\mathbf{x}|\mathbf{y}_m} [\text{diag}\{\mathbf{G}_m \mathbf{x}\}] \cdot \mathbb{E}_{\mathbf{x}|\mathbf{y}_m} [\mathbf{G}_m \mathbf{x}] \right\} \right] \end{aligned} \quad (4.17)$$

where  $\frac{\partial TSI}{\partial \widetilde{\mathbf{W}}} = \left[ \frac{\partial TSI}{\partial w(1)}, \frac{\partial TSI}{\partial w(2)}, \dots, \frac{\partial TSI}{\partial w(L)} \right]^T$ ,  $\mathbb{E}_{\mathbf{x}|\mathbf{y}_m} [\mathbf{G}_m \mathbf{x}]$  is calculated by

$$\mathbb{E}_{\mathbf{x}|\mathbf{y}_m} [\mathbf{G}_m \mathbf{x}] = \frac{\sum_{h=1}^{\mathcal{H}} \{\mathbf{G}_m \mathbf{x}_h\} p(\mathbf{y}_m | \mathbf{x}_h) P(\mathbf{x}_h)}{\sum_{h=1}^{\mathcal{H}} p(\mathbf{y}_m | \mathbf{x}_h) P(\mathbf{x}_h)},$$

and  $\mathbb{E}_{\mathbf{x}|\mathbf{y}_m} [\text{diag}\{\mathbf{G}_m \mathbf{x}\}]$  is similarly calculated. In (4.17), all of the  $N_{\text{Rx}}$  bistatic paths factor into the gradient calculation for each waveform coefficient. The detailed proof is given in Appendix C.2.

Once again, the expected values must be calculated via Monte Carlo integration, which can be expressed as

$$\begin{aligned} & \frac{\partial TSI}{\partial \widetilde{\mathbf{W}}} \\ &= \frac{\mathbf{W}}{N_{\text{total}} \sigma^2} \sum_{\mathbf{x}_0, \mathbf{y}_{m_0}} \left\{ \sum_{m=1}^{N_{\text{Rx}}} \left\{ \text{diag}\{\mathbf{G}_m \mathbf{x}_0\} \cdot \{\mathbf{G}_m \mathbf{x}_0\} - \mathbb{E}_{\mathbf{x}|\mathbf{y}_{m_0}} [\text{diag}\{\mathbf{G}_m \mathbf{x}\}] \cdot \mathbb{E}_{\mathbf{x}|\mathbf{y}_{m_0}} [\mathbf{G}_m \mathbf{x}] \right\} \right\} \end{aligned} \quad (4.18)$$

where  $\widehat{\mathbf{y}}_0 = [\mathbf{y}_{1_0}, \mathbf{y}_{2_0}, \dots, \mathbf{y}_{0N_{\text{Rx}}}]^T$  and

$$\mathbb{E}_{\mathbf{x}|\mathbf{y}_{m_0}} [\mathbf{G}_m \mathbf{x}] = \frac{\sum_{h=1}^{\mathcal{H}} \{\mathbf{G}_m \mathbf{x}_h\} p(\mathbf{y}_{m_0} | \mathbf{x}_h) P(\mathbf{x}_h)}{\sum_{h=1}^{\mathcal{H}} p(\mathbf{y}_{m_0} | \mathbf{x}_h) P(\mathbf{x}_h)}.$$

Next, we consider MIMO radar waveform design. One important consideration in MIMO radar waveform design is how to make the waveforms orthogonal such that reflected signals due to different transmitted waveforms can be separated at the receivers. For example, one approach is to transmit waveforms on distinct frequency bands such that the receivers can separate the waveforms using bandpass filters. Other techniques are also possible, and MIMO radar waveform design is currently an active research area [7][51][52]. At this point, however, we wish to demonstrate the highest potential benefit of MIMO radar waveform design via the TSI metric. We do not explicitly consider or enforce waveform orthogonality, though in the future this is certainly an issue that must be incorporated into MIMO waveform design by TSI or any other design metric. We assume here that the individual waveforms can be separated at the receivers and that the target transfer functions are equal for both directions of a particular propagation path. This reciprocity would not hold true for transmitters that operate on different frequency bands.

We extend the SIMO system model by adding the transmitter indices  $n = 1, 2, \dots, N_{\text{Tx}}$  according to

$$\tilde{\mathbf{y}} = \begin{bmatrix} \dot{y}_1 \\ \dot{y}_2 \\ \vdots \\ \dot{y}_{N_{\text{Rx}}} \end{bmatrix} = \tilde{\mathbf{W}}\tilde{\mathbf{G}}\mathbf{x} + \tilde{\mathbf{n}} = \begin{bmatrix} \dot{\mathbf{W}}\dot{\mathbf{G}}_1\mathbf{x} + \dot{\mathbf{n}}_1 \\ \dot{\mathbf{W}}\dot{\mathbf{G}}_2\mathbf{x} + \dot{\mathbf{n}}_2 \\ \vdots \\ \dot{\mathbf{W}}\dot{\mathbf{G}}_{N_{\text{Rx}}}\mathbf{x} + \dot{\mathbf{n}}_{N_{\text{Rx}}} \end{bmatrix} \quad (4.19)$$

where the  $\mathcal{L}N_{\text{Rx}}N_{\text{Tx}} \times \mathcal{L}N_{\text{Rx}}N_{\text{Tx}}$  diagonal matrix  $\tilde{\mathbf{W}} = (\mathbf{I}_{N_{\text{Rx}}} \otimes \dot{\mathbf{W}})$ , the  $\mathcal{L}N_{\text{Rx}}N_{\text{Tx}} \times \mathcal{H}$  matrix  $\tilde{\mathbf{G}} = [\dot{\mathbf{G}}_1^T, \dot{\mathbf{G}}_2^T, \dots, \dot{\mathbf{G}}_{N_{\text{Rx}}}^T]^T$ ,  $\mathbf{x}$  is an  $(\mathcal{H} \times 1)$  indicator vector, and the noise vector  $\tilde{\mathbf{n}}$  is  $\mathcal{L}N_{\text{Rx}}N_{\text{Tx}} \times 1$ . The multi-transmitter waveform matrix is formed via a matrix direct

sum according to  $\hat{\mathbf{W}} = (\mathbf{W}_1 \oplus \mathbf{W}_2 \oplus \dots \oplus \mathbf{W}_{N_{\text{Tx}}})$ . The multi-input, single-output transfer function matrix for the  $m$ th receiver is  $\hat{\mathbf{G}}_m = [\mathbf{G}_{m1}^T, \mathbf{G}_{m2}^T, \dots, \mathbf{G}_{mN_{\text{Tx}}}^T]^T$  with each block matrix entry  $\mathbf{G}_{mn}$  of size  $\mathcal{L} \times \mathcal{H}$  for  $n = 1, 2, \dots, N_{\text{Tx}}$ . The noise vector is  $\hat{\mathbf{n}} = [\mathbf{n}_1^T, \mathbf{n}_2^T, \dots, \mathbf{n}_{N_{\text{Tx}}}^T]^T$ . For a simple  $2 \times 2$  case, the MIMO model is

$$\begin{bmatrix} \mathbf{y}_{11} \\ \mathbf{y}_{12} \\ \mathbf{y}_{21} \\ \mathbf{y}_{22} \end{bmatrix} = \begin{bmatrix} W_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & W_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & W_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & W_2 \end{bmatrix} \begin{bmatrix} \mathbf{G}_{11}\mathbf{x} \\ \mathbf{G}_{12}\mathbf{x} \\ \mathbf{G}_{21}\mathbf{x} \\ \mathbf{G}_{22}\mathbf{x} \end{bmatrix} + \begin{bmatrix} \mathbf{n}_{11} \\ \mathbf{n}_{12} \\ \mathbf{n}_{21} \\ \mathbf{n}_{22} \end{bmatrix}. \quad (4.20)$$

On inspecting (4.13) and (4.19), we see that the MIMO system model is similar to the monostatic system model with the exception of replacing  $\mathbf{W}$ ,  $\mathbf{G}$ , and  $\mathbf{n}$  with the multi-transmitter versions  $\hat{\mathbf{W}}$ ,  $\hat{\mathbf{G}}$ , and  $\hat{\mathbf{n}}$ . The difference will be in how the waveform energy constraints are applied.

The optimization statement is now

$$\begin{aligned} \max \quad & TSI = I(\tilde{\mathbf{y}}; \mathbf{x}) \\ \text{s.t.} \quad & \{Tr[W_n W_n^T] \leq E_{w,n} | n = 1, 2, \dots, N_{\text{Tx}}\}. \end{aligned} \quad (4.21)$$

In the MIMO configuration, we must design waveforms for multiple transmitters. In theory, we can either set individual energy constraints for each individual radar waveform as shown in (4.21) or set a total energy constraint  $E_{w,total}$  for the  $N_{\text{Tx}}$  waveforms combined according to  $\sum_{n=1}^{N_{\text{Tx}}} Tr[W_n W_n^T] \leq E_{w,total}$ . However, the former scheme is more practical for distributed MIMO radar systems, which are the focus here.

In the next section, we search the first quadrant for the optimal waveform and compare the performance with other waveform designs.

#### 4.4 Simulation Result

In this section we present simulation results that indicate the potential benefit of using TSI as a waveform design metric. We show the results of monostatic and MIMO radar configurations with TSI-based waveforms optimized using the gradient search method. Figures 4.3 to 4.6 demonstrate performance in the monostatic configuration while Figures 4.7 and 4.9 demonstrate performance for MIMO configurations.

We evaluate the probability of error in determining the true target transfer function with three different waveforms to compare their performances. The first is a wideband waveform which a flat energy distribution across the transmission band, and the second is a waterfilling waveform that is optimized based on the mutual information metric between the received measurements and the variance of the target transfer functions [8]. The last is the TSI waveform presented in this chapter. The number of target hypotheses is  $\mathcal{H} = 4$ . The waveform dimension is  $\mathcal{L} = 10$  for all cases except the one shown in Figure 4.3, which was generated with a 5 tap waveform. The measurement noise power is normalized to  $\sigma^2 = 1$ , the waveform energy allocation varies from  $10^{-4}$  to  $10^0$  energy units, and probability of error is calculated over 20000 trials for each of 100 different target sets. The deterministic target transfer functions were randomly generated as realizations of Gaussian random vectors, but once generated were treated as known for the duration of the classification experiment. We used both non-colored and colored distributions for generating the target transfer functions. Figures 4.3 to 4.8 show the benefits of TSI-based waveform design. The most improvement is seen in the target sets generated with colored Gaussian characteristics because these cases provide more instances where a few specific frequencies provide information that distinguish the hypotheses. Thus, waveform spectral shape is more important.

Both the TSI-based waveform and the waterfilling waveform outperform the wide-

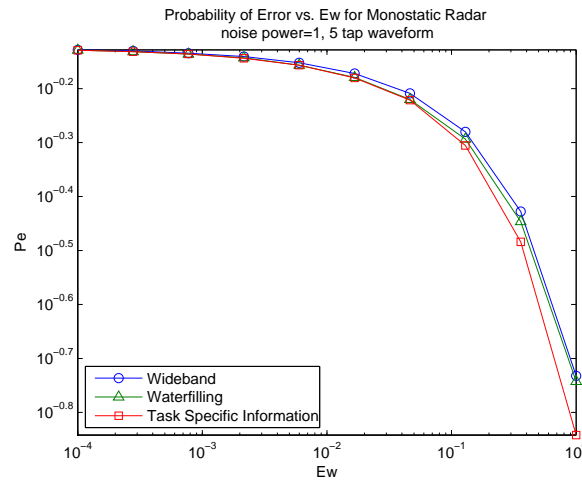


Figure 4.3: Performance comparison of various waveform design methods in maximum likelihood detection in case of 5 tap waveform and non-colored target sets

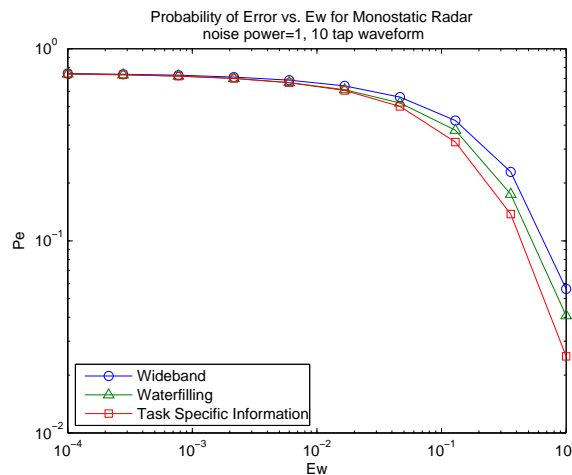


Figure 4.4: Performance comparison of various waveform design methods in maximum likelihood detection in case of 10 tap waveform and non-colored target sets

band waveform. This fact is easy to understand since these two waveforms are the only ones that specifically take target characteristics into account. Figures 4.3 and 4.4 show the results for five-tap and 10-tap waveforms, respectively, with non-colored target sets. Figures 4.5 and 4.6 show results for 10-tap waveforms in the monostatic radar configuration for target sets colored with two different sets of weighting functions. Figures 4.7 and 4.8

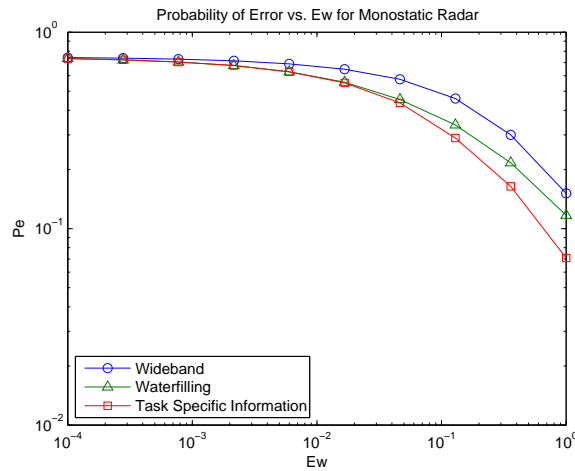


Figure 4.5: Performance comparison of various waveform design methods in maximum likelihood detection in case of 10 tap waveform with colored target set I

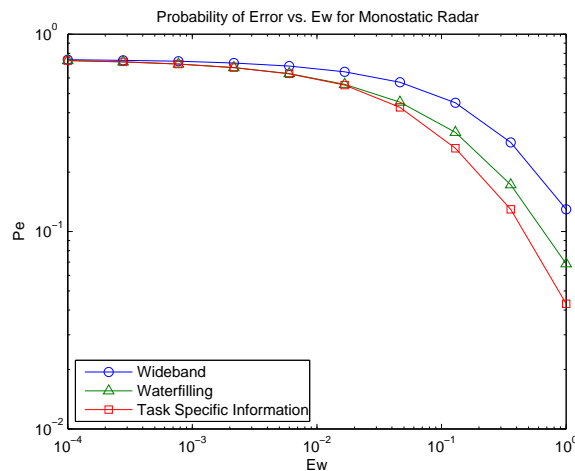


Figure 4.6: Performance comparison of various waveform design methods in maximum likelihood detection in case of 10 tap waveform with colored target set II

show results for MIMO radar with two different sets of coloring functions. The results of colored sets show the wider performance gap compared to the non-colored result of Figures 4.3 and 4.4. They show that TSI-based design takes efficient advantage of target spectral characteristics. All decisions are based on maximum likelihood detection using pdf's derived earlier.

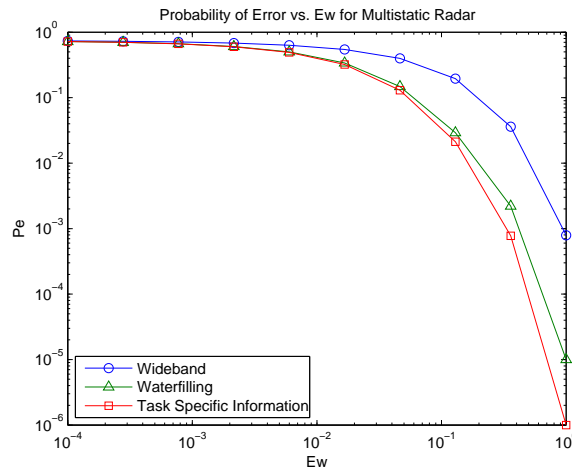


Figure 4.7: Performance comparison of various waveform design methods in maximum likelihood detection in case of 10 tap waveform with colored target set I for MIMO radar system of three radars

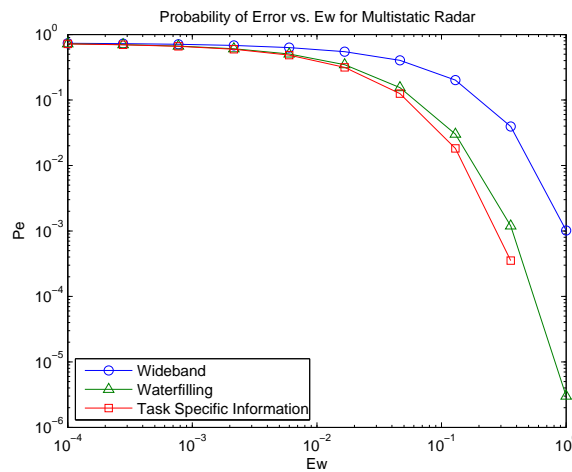


Figure 4.8: Performance comparison of various waveform design methods in maximum likelihood detection in case of 10 tap waveform with colored target set II for MIMO radar system of three radars

Note that the asymptotic slopes seen in Figures 4.7 and 4.8 are steeper than we observe for the monostatic cases due to the diversity gain provided by multiple platforms. Figure 4.9 shows various diversity gains which are proportional to the total number of radars. The slope of the MIMO configuration with three radars is three times steeper than that of

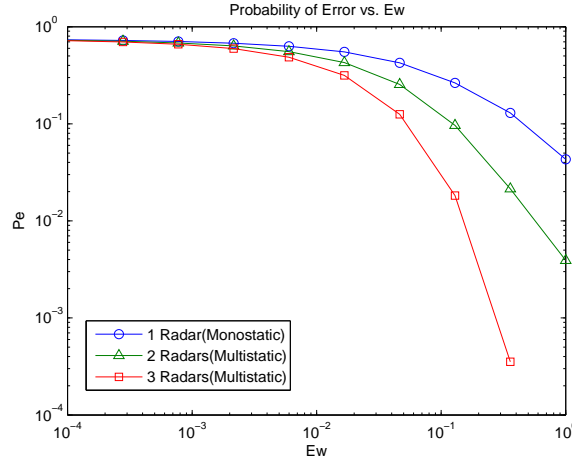


Figure 4.9: Diversity gain by MIMO radar configuration using TSI waveform design in case of 10 tap waveform with colored target set II

the monostatic result.

#### 4.4.1 Energy Distribution in Waveform

It is interesting to evaluate how TSI-based waveform design distributes waveform energy across the allowable frequency band in comparison to other matched waveform techniques. We observe that for low energy constraints, the TSI waveform design distributes energy similarly to the MI-based waterfilling approach. These methods tend to distribute energy into just a few frequency components. However, as the given waveform energy increases, the TSI-based approach begins to distribute the energy across more frequency components than the the waterfilling approach. At extremely high energy allocations, both methods will produce flat energy distributions. Table 4.1 shows the energy distributions of the waterfilling approach versus TSI-based design as the energy constraint increases for one set of target transfer functions taken from the target spectra used in Figure 4.10. We define five different energy constraints in increasing order according to  $E1 < E2 < E3 < E4 < E5$ . Table 4.1 shows the percentage of energy distributed

to each of the 10 frequency components for all five energy constraints for both waveform design techniques. At the lowest energy level, most of the energy is allocated to the 9<sup>th</sup> frequency component by both design methods. At low energy, this is the best frequency component for having any chance of a correct decision. At energy constraint  $E_2$ , the TSI waveform allocates 12.63 percent of its energy to the 4<sup>th</sup> frequency band while the waterfilling approach still allocates all energy to the 9<sup>th</sup> component. It turns out in this case that the 4<sup>th</sup> frequency component was important for distinguishing between the second and fourth targets. The waterfilling waveform does not allocate energy to the 4<sup>th</sup> frequency component until the highest energy constraint  $E_5$  while by that point the TSI-based design allocates energy to several frequency components. The differing energy allocations are a direct result of accounting for the correlation between frequency components under a finite-hypothesis decision-making task. The MI-based design approach assumes independence between Gaussian-distributed frequency components; thus, it continues to allocate energy to the 9<sup>th</sup> frequency component in order to reduce the variance across the transfer functions at that frequency. The TSI-based approach, however, recognizes the diminishing returns provided by that frequency component with respect to the classification task and begins exploiting additional frequencies.

#### 4.5 Conclusion of Radar Waveform Design

A new waveform design method based on the concept of task-specific information is introduced. To evaluate TSI, it is necessary to introduce source variables that directly encode the task at hand. For example, in the  $\mathcal{H}$ -class identification problem, the source variable is an indicator vector that takes one of  $\mathcal{H}$  possible realizations. By optimizing over the mutual information between the measurements and this source variable rather than the mutual information between the measurements and target transfer functions, it is possible

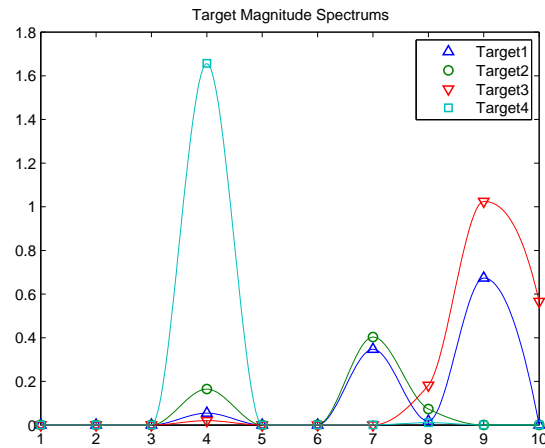


Figure 4.10: Target Magnitude Spectrums used for the Comparison of Waveform Energy Distribution

to further improve performance beyond what has been demonstrated in earlier work. The TSI-based objective function has a single maxima in each quadrant of a spheroidal search space under a fixed energy constraint, and we showed that we need only search over a single quadrant of the design space; thus, gradient-based search techniques lead to a globally optimum solution. The performance improvement of the TSI-based approach is evident in our simulation results. We also verified that TSI shows the best performance both in monostatic and MIMO radar configurations through the simulation results. However, the computational load of TSI-based design is significant due to the need to evaluate expected values for the gradient search technique via Monte Carlo methods.

Table 4.1: Comparison of Waveform Energy Distribution

$E_w$	Scheme	Waveform Energy Distribution, Unit: %									
		1	2	3	4	5	6	7	8	9	10
$E1$	Waterfilling	0	0	0	0	0	0	0	0	100	0
	TSI	0	0	0	0	0	0	0	0	100	0
$E2$	Waterfilling	0	0	0	0	0	0	0	0	100	0
	TSI	0	0	0	12.6272	0	0	0	0	87.3728	0
$E3$	Waterfilling	0	0	0	0	0	0	0	0	100	0
	TSI	0	0	0	9.9246	0	0	4.5054	0	85.5700	0
$E4$	Waterfilling	0	0	0	0	0	0	0	0	100	0
	TSI	0.1001	0.1004	0.1000	28.1495	0.1002	0.1035	17.8631	2.3393	47.5294	3.6145
$E5$	Waterfilling	0	0	0	8.1607	0	0	0	0	91.8393	0
	TSI	0.0683	0.0683	0.0681	23.6220	0.0683	0.0714	22.6484	2.2160	47.9172	3.2518

## CHAPTER 5

## FINAL CONCLUSIONS AND FUTURE WORKS

We discussed sensor power optimization for wireless sensor networks (WSNs) using J-divergence. Detection/classification outage was proposed as a long-term performance measure for WSNs communicating over slow-fading channels. The long-term performance measure was used to compare the performances of different power control schemes. In addition, we discussed a radar waveform optimization with a new information measure: Task-Specific Information (TSI). Probability of error was numerically calculated to compare the performances of various waveforms. In these optimizations, we took advantage of the statistics of the given WSN or radar system model to get a better power transmission scheme for WSNs or a better radar waveform.

The design parameters of WSNs are the amplifying parameters and symbol constellations of the multiple sensors. The parameters impact the battery powers of the individual sensors since the value of a parameter means how much a sensor amplifies and forwards its sensor decision in the optimized constellation through each channel. The transmit constellation for a sensor can be a  $\mathcal{D}$ -dimensional vector depending on the given system modulation method. Each sensor can have a different sensor measurement noise. Some of the communication paths from the sensors to the fusion center can have high instantaneous channel SNR, while other paths can have low instantaneous channel SNR. Overall, the system statistics are composed of a priori probabilities of binary/multiple-hypotheses, sensor measurement noises, additive channel noises, and fading channel realizations. The power optimization with J-divergence or total J-divergence provides us with an optimal power control strategy based on the system statistics.

The design parameters of the radar system are the spectral coefficients of the waveform. The target signatures and the channel noises make a heterogeneous SNR variation across the frequency spectrum. The waveform optimization scheme with TSI provides us with an optimal waveform based on the heterogeneous SNR variation across the frequency components. Here, the TSI optimization algorithm is performed by a gradient search algorithm and the TSI gradient has a nice and compact form of MMSE expectation [11].

From the computational point of view, incrementing the total number of hypotheses  $\mathcal{H}$  results in an increment of the dimension of the source variable  $\mathbf{x}$ . In this case, linear sums in the MMSE expectation of the TSI gradient increases proportionally. However, if the dimension of design parameter increases, the computational load increases exponentially since we should dramatically increase the total size of samples for more precise Monte Carlo expectation. In summary, the computational load and complexity of TSI optimization are less sensitive to the total number of hypotheses while they are more sensitive to the dimension of design parameter. On the other hand, the computational load and complexity of total J-divergence optimization are more sensitive to the total number of hypotheses while they are less sensitive to the dimension of design parameters.

When J-divergence was used for an objective function, observations at the fusion center were assumed to be Gaussian distributed since it is not possible to derive the analytical J-divergence under a Gaussian mixture model. In addition, for the multiple-hypothesis problem with total J-divergence measure, we need to increase the total number of individual J-divergences with the number of total hypotheses. Therefore, the amplifying parameters are controlled by more individual J-divergence equations. The number of convex conditions shown in Appendix B.6 increases and the objective function becomes more complex due to non-linearity with respect to the design parameters. In this case, we

need to consider numerical errors for a higher number of hypotheses. In the long run, it becomes very hard to find an optimal solution especially with the interior point method for non-convex case.

Therefore, for the next topic, the TSI measure for the sensor power optimization problem is promising. In a practical wireless sensor network problem, we might use a few number of sensors. Even though the computational load of TSI optimization can be more intensive than that of J-divergence, the numerical error of TSI optimization is less sensitive to the total number of target classes(hypotheses) and TSI model does not require the Gaussian assumption. In short, in a slow-fading WSNs with few sensors, the computational load is not a big issue and we expect that TSI optimization gives a better result since TSI is mathematically well-defined in the form of information measure without any mathematical approximation.

In Chapter 4, a radar waveform optimization based on deterministic target signatures was discussed. In the target signature model, the orientation of the target with respect to the radar was not considered. This assumption means that each target class was perfectly characterized by a single deterministic target signature. For the future research topic, radar waveform design with simple random target models can be investigated, and then a more realistic and sophisticated random target model covering the 360-degree angle of targets can be developed. The 360-degree angle model can be treated by a limited number of perspectives along with approximations for the angles between these perspectives. In other words, it is an interpolated model with a limited number of perspectives for the full 360-degree angle. In this realistic model, we need to also make a target model library including the model parameters of the 360-degree perspective models through case-dependent studies.

## APPENDIX A

## DERIVATIONS AND PROOFS FOR CHAPTER 2

## A.1 Derivation of (2.3)

Let us consider  $J_1(\cdot)$  for a large  $K$  and  $0 < \alpha_F(k) - \beta_F(k) < 1$ . We have

$$\begin{aligned} & \frac{P_{tot}s_k[\alpha_F(k) - \beta_F(k)]}{K} - \frac{P_{tot}s_k[\alpha_F(k) - \beta_F(k)]}{K + P_{tot}s_k\beta_F(k)} \\ &= \frac{P_{tot}^2s_k^2\beta_F(k)[\alpha_F(k) - \beta_F(k)]}{K^2 + K \cdot P_{tot}s_k\beta_F(k)} \\ &\leq \frac{P_{tot}^2s_k^2\beta_F(k)[\alpha_F(k) - \beta_F(k)]}{K^2} \end{aligned}$$

since  $K, P_{tot}, s_k$ , and  $\beta_F(k) = P_D(k) - P_D(k)^2$  are positive.

The inequality can be arranged as follows

$$\begin{aligned} & \frac{P_{tot}s_k[\alpha_F(k) - \beta_F(k)]}{K} - \frac{P_{tot}^2s_k^2\beta_F(k)[\alpha_F(k) - \beta_F(k)]}{K^2} \\ &\leq \frac{P_{tot}s_k[\alpha_F(k) - \beta_F(k)]}{K + P_{tot}s_k\beta_F(k)} \leq \frac{P_{tot}s_k[\alpha_F(k) - \beta_F(k)]}{K}. \\ \Rightarrow & \sum_{k=1}^K \frac{P_{tot}s_k[\alpha_F(k) - \beta_F(k)]}{K} \\ & - \sum_{k=1}^K \frac{P_{tot}^2s_k^2\beta_F(k)[\alpha_F(k) - \beta_F(k)]}{K^2} \\ &\leq J_1(\cdot) \leq \sum_{k=1}^K \frac{P_{tot}s_k[\alpha_F(k) - \beta_F(k)]}{K} \end{aligned}$$

since  $\frac{P_{tot}s_k[\alpha_F(k) - \beta_F(k)]}{K + P_{tot}s_k\beta_F(k)} \leq \frac{P_{tot}s_k[\alpha_F(k) - \beta_F(k)]}{K}$ .

## A.2 Derivation of (2.6)

Let us consider  $J_1(\cdot)$  for a large  $K$  and  $-1 < \alpha_F(k) - \beta_F(k) < 0$ . We have

$$\begin{aligned} & \frac{P_{tot}s_k[\alpha_F(k) - \beta_F(k)]}{K} - \frac{P_{tot}s_k[\alpha_F(k) - \beta_F(k)]}{K + P_{tot}s_k\beta_F(k)} \\ &= \frac{P_{tot}^2s_k^2\beta_F(k)[\alpha_F(k) - \beta_F(k)]}{K^2 + K \cdot P_{tot}s_k\beta_F(k)} \\ &\geq \frac{P_{tot}^2s_k^2\beta_F(k)[\alpha_F(k) - \beta_F(k)]}{K^2} \end{aligned}$$

since  $K, P_{tot}, s_k$ , and  $\beta_F(k) = P_D(k) - P_D(k)^2$  are positive.

The inequality can be arranged as follows

$$\begin{aligned} & \frac{P_{tot}s_k[\alpha_F(k) - \beta_F(k)]}{K} - \frac{P_{tot}^2s_k^2\beta_F(k)[\alpha_F(k) - \beta_F(k)]}{K^2} \\ &\geq \frac{P_{tot}s_k[\alpha_F(k) - \beta_F(k)]}{K + P_{tot}s_k\beta_F(k)} \geq \frac{P_{tot}s_k[\alpha_F(k) - \beta_F(k)]}{K}. \\ \implies & \sum_{k=1}^K \frac{P_{tot}s_k[\alpha_F(k) - \beta_F(k)]}{K} \leq J_1(\cdot) \leq \\ & \sum_{k=1}^K \frac{P_{tot}s_k[\alpha_F(k) - \beta_F(k)]}{K} - \sum_{k=1}^K \frac{P_{tot}^2s_k^2\beta_F(k)[\alpha_F(k) - \beta_F(k)]}{K^2} \end{aligned}$$

since  $\frac{P_{tot}s_k[\alpha_F(k) - \beta_F(k)]}{K} \leq \frac{P_{tot}s_k[\alpha_F(k) - \beta_F(k)]}{K + P_{tot}s_k\beta_F(k)}$ .

## A.3 Derivation of the two conditions in Theorem 3.1

The two conditions,  $|\alpha_F(k) - \beta_F(k)| < 1$  and  $|\alpha_D(k) - \beta_D(k)| < 1$  are easily verified to hold in our system as follows:

1.  $\alpha_F(k) - \beta_F(k) = P_F(k) + P_D(k)^2 - 2P_D(k)P_F(k) - P_D(k) + P_D(k)^2 = (2P_D(k) - 1)(P_D(k) - P_F(k))$

$$\implies |\alpha_F(k) - \beta_F(k)| < 1, \text{ since } |2P_D(k) - 1| < 1 \text{ and } |P_D(k) - P_F(k)| < 1.$$

$$2. \alpha_D(k) - \beta_D(k) = P_D(k) + P_F(k)^2 - 2P_D(k)P_F(k) - P_F(k) + P_F(k)^2 = (2P_F(k) - 1)(P_F(k) - P_D(k))$$

$$\implies |\alpha_F(k) - \beta_F(k)| < 1, \text{ since } |2P_F(k) - 1| < 1 \text{ and } |P_F(k) - P_D(k)| < 1.$$

#### A.4 Proof of Theorem 3.2

We start from the definition of outage probability as follows,

$$\begin{aligned} p_{J_0} &= \text{Prob}\{J(\cdot) < J_0\} \\ &= \text{Prob}\left\{\frac{1}{2}[J_1(\cdot) + J_2(\cdot)] < J_0\right\}. \end{aligned}$$

Based on the inequalities given in (2.3) and (2.6), and the counterpart inequalities for  $J_2(\cdot)$ , for large  $\mathcal{K}$ , we asymptotically have

$$\begin{aligned} J_1(\cdot) &\sim \sum_{k=1}^{\mathcal{K}} \frac{P_{tot} s_k [\alpha_F(k) - \beta_F(k)]}{\mathcal{K}} \\ \text{and } J_2(\cdot) &\sim \sum_{k=1}^{\mathcal{K}} \frac{P_{tot} s_k [\alpha_D(k) - \beta_D(k)]}{\mathcal{K}}. \end{aligned}$$

As such, we obtain

$$\begin{aligned}
& \lim_{\mathcal{K} \rightarrow \infty} -\frac{1}{\mathcal{K}} \log p_{J_0} \\
&= \lim_{\mathcal{K} \rightarrow \infty} -\frac{1}{\mathcal{K}} \log \left[ \text{Prob} \left\{ \frac{1}{2} \left[ \sum_{k=1}^{\mathcal{K}} \frac{s_k [\alpha_F(k) - \beta_F(k)]}{\mathcal{K}} \right] + \right. \right. \\
&\quad \left. \left. \frac{1}{2} \left[ \sum_{k=1}^{\mathcal{K}} \frac{s_k [\alpha_D(k) - \beta_D(k)]}{\mathcal{K}} \right] < \frac{J_0}{P_{tot}} \right\} \right] \\
&= \lim_{\mathcal{K} \rightarrow \infty} -\frac{1}{\mathcal{K}} \log \left[ \text{Prob} \left\{ \frac{1}{\mathcal{K}} \sum_{k=1}^{\mathcal{K}} \eta_k < c \right\} \right],
\end{aligned}$$

where  $c = \frac{J_0}{P_{tot}}$  and  $\eta_k = \frac{s_k(\alpha_F(k) + \alpha_D(k) - \beta_F(k) - \beta_D(k))}{2}$ .

By Lemma 6.1 in [41], for  $\{\eta_k : k \in \{1, \dots, \mathcal{K}\}\}$  that are i.i.d random variables, and  $c < E\{\eta_k\}$  that is satisfied when  $J_0 < J_\infty$ , we have

$$\lim_{\mathcal{K} \rightarrow \infty} -\frac{1}{\mathcal{K}} \log \left[ \text{Prob} \left\{ \frac{1}{\mathcal{K}} \sum_{k=1}^{\mathcal{K}} \eta_k < c \right\} \right] = I_\eta(c),$$

where the rate function  $I_\eta(c) = \sup_{\theta \in R} (\theta c - \log M_\eta(\theta))$  with  $M_\eta(\theta)$  the moment generating function of  $\eta_k$ .

Therefore, for large  $\mathcal{K}$ , we asymptotically have

$$\begin{aligned}
& \log p_{J_0} \sim -\mathcal{K} I_\eta(c) \\
& \implies p_{J_0} \sim \exp(-\mathcal{K} I_\eta(c)).
\end{aligned} \tag{A.1}$$

where  $\eta$  is the generalized representation of  $\eta_k$ .

## APPENDIX B

## DERIVATIONS AND PROOFS FOR CHAPTER 3

## B.1 The Relationship between Bound Probability of Error and Total J-divergence

The inequality  $P_e > P(H_1)P(H_2)e^{-J/2}$  [37] is valid between two hypotheses with a priori probabilities whose sum are less than one. This can be proved by starting with

$$P_e^{12} = P(u|H_1)\dot{P}(H_1) + P(u|H_2)\dot{P}(H_2) > \dot{P}(H_1)\dot{P}(H_2)e^{-J_{12}/2}.$$

Let  $\dot{P}(H_1) = \frac{P(H_1)}{a}$  and  $\dot{P}(H_2) = \frac{P(H_2)}{b}$  be new probabilities where  $a$  and  $b \geq 1$ . Then,

$$\begin{aligned} P(u|H_1)\frac{P(H_1)}{a} + P(u|H_2)\frac{P(H_2)}{b} &> \frac{P(H_1)}{a}\frac{P(H_2)}{b}e^{-J_{12}/2} \\ \iff P(u|H_1)P(H_1)b + P(u|H_2)P(H_2)a &> P(H_1)P(H_2)e^{-J_{12}/2}. \end{aligned}$$

Now, we extend the inequality for the multiple-hypotheses problem by defining an upper bound to the bound probability of error,  $P_e^{bound} = \sum_{i=1}^{\mathcal{H}} \sum_{j=1}^{\mathcal{H}} P_e^{ij}$ , and a probability-weighted total J-divergence,  $J^{total} = \sum_{i=1}^{\mathcal{H}} \sum_{j=i+1}^{\mathcal{H}} J^{ij} \cdot P(H_i) \cdot P(H_j)$ . At first, by summing the inequalities, we get

$$\sum_{i=1}^{\mathcal{H}} \sum_{j=1}^{\mathcal{H}} P_e^{ij} > \sum_{i=1}^{\mathcal{H}} \sum_{j=1}^{\mathcal{H}} P(H_i)P(H_j)e^{-J^{ij}/2}.$$

By an inequality,  $e^x > 1 + x$ , which is derived via the Taylor series of the exponential function, we get

$$\begin{aligned}
\sum_{i=1}^{\mathcal{H}} \sum_{j=1}^{\mathcal{H}} P_e^{ij} &> \sum_{i=1}^{\mathcal{H}} \sum_{j=1}^{\mathcal{H}} P(H_i)P(H_j)e^{-\frac{J^{ij}}{2}} \\
&> \sum_{i=1}^{\mathcal{H}} \sum_{j=1}^{\mathcal{H}} P(H_i)P(H_j)\left(1 - \frac{J^{ij}}{2}\right). \\
&= \sum_{i=1}^{\mathcal{H}} \sum_{j=1}^{\mathcal{H}} P(H_i)P(H_j) - J^{total}
\end{aligned}$$

since  $J^{total} = \frac{1}{2} \sum_{i=1}^{\mathcal{H}} \sum_{j=1}^{\mathcal{H}} J^{ij} \cdot P(H_i) \cdot P(H_j)$ . Finally, we get

$$P_e^{bound} > C - J^{total}$$

where  $C = \sum_{i=1}^{\mathcal{H}} \sum_{j=1}^{\mathcal{H}} P(H_i)P(H_j)$ .

## B.2 Derivation of (3.6)

We now derive an approximate total J-divergence measure that explicitly includes the optimized transmit symbol constellation. The measure is approximate because the signals received at the fusion center are distributed according to a Gaussian mixture, which doesn't lead to closed-form expressions. Therefore, we adopt the strategy used in [22] [42], which is to approximate the received conditional probabilities in (2.1) by Gaussian densities with the same mean and covariance as the mixture.

As a first step, we must first derive the mean and covariance of  $\mathbf{y}$  under  $H_i$  to find the

multivariate conditional PDF  $p(\mathbf{y}|H_i)$ . The mean vector is given as

$$\begin{aligned}
\boldsymbol{\mu}_i &= \int_{\mathbf{y}} \mathbf{y} p(\mathbf{y}|H_i) d\mathbf{y} \\
&= \int_{\mathbf{y}} \mathbf{y} \sum_{\mathbf{u}} p(\mathbf{y}|\mathbf{u}) P(\mathbf{u}|H_i) d\mathbf{y} \\
&= \sum_{\mathbf{u}} P(\mathbf{u}|H_i) \int_{\mathbf{y}} \mathbf{y} p(\mathbf{y}|\mathbf{u}) d\mathbf{y} \\
&= \sum_{\mathbf{u}} P(\mathbf{u}|H_i) \mathbf{B} \mathbf{A} \mathbf{f}_s(\mathbf{u}) \\
&= \mathbf{B} \mathbf{A} \begin{bmatrix} \sum_{l=1}^{\mathcal{H}} \delta_1(l) P(u_1 = l|H_i) \\ \sum_{l=1}^{\mathcal{H}} \delta_2(l) P(u_2 = l|H_i) \\ \vdots \\ \sum_{l=1}^{\mathcal{H}} \delta_{\mathcal{K}}(l) P(u_{\mathcal{K}} = l|H_i) \end{bmatrix} \\
&= \mathbf{B} \mathbf{A} \boldsymbol{\xi}_i
\end{aligned} \tag{B.1}$$

where  $i \in \{1, 2, \dots, \mathcal{H}\}$ ,  $\delta_k(l) = f_s(u_k = l)$  and

$$\boldsymbol{\xi}_i = \begin{bmatrix} \sum_{l=1}^{\mathcal{H}} \delta_1(l) P(u_1 = l|H_i) \\ \sum_{l=1}^{\mathcal{H}} \delta_2(l) P(u_2 = l|H_i) \\ \vdots \\ \sum_{l=1}^{\mathcal{H}} \delta_{\mathcal{K}}(l) P(u_{\mathcal{K}} = l|H_i) \end{bmatrix}.$$

The covariance matrix is given as

$$\begin{aligned}
\boldsymbol{\Sigma}_i &= \int_{\mathbf{y}} [\mathbf{y} - \boldsymbol{\mu}_i][\mathbf{y} - \boldsymbol{\mu}_i]^T p(\mathbf{y}|H_i) d\mathbf{y} \\
&= \int_{\mathbf{y}} \mathbf{y} \mathbf{y}^T p(\mathbf{y}|H_i) d\mathbf{y} - \boldsymbol{\mu}_i \boldsymbol{\mu}_i^T.
\end{aligned}$$

From (B.1),  $\boldsymbol{\mu}_i \boldsymbol{\mu}_i^T = \mathbf{B} \mathbf{A} \boldsymbol{\xi}_i \boldsymbol{\xi}_i^T \mathbf{B} \mathbf{A}$ . Hence, after some algebra with  $\sum_{u_k=1}^{\mathcal{H}} P(u_k|H_i) = 1$ , we obtain

$$\begin{aligned} \boldsymbol{\Sigma}_i &= \mathbf{B} \mathbf{A} \sum_{\mathbf{u}} \mathbf{f}_s(\mathbf{u}) \mathbf{f}_s(\mathbf{u})^T P(\mathbf{u}|H_i) \mathbf{A}^T \mathbf{B}^T - \mathbf{B} \mathbf{A} \boldsymbol{\xi}_i \boldsymbol{\xi}_i^T \mathbf{A}^T \mathbf{B}^T + \mathbf{R} \\ &= \mathbf{B} \mathbf{A} \left[ \sum_{\mathbf{u}} \mathbf{f}_s(\mathbf{u}) \mathbf{f}_s(\mathbf{u})^T P(\mathbf{u}|H_i) - \boldsymbol{\xi}_i \boldsymbol{\xi}_i^T \right] \mathbf{A}^T \mathbf{B}^T + \mathbf{R} \\ &= \mathbf{B} \mathbf{A} \mathbf{C}_i \mathbf{A}^T \mathbf{B}^T + \mathbf{R} \end{aligned} \quad (\text{B.2})$$

where  $\mathbf{R}$  is the covariance matrix of the channel noise vector  $\mathbf{n}$ . In (B.2),  $\mathbf{C}_i$  is a  $\mathcal{K} \times \mathcal{K}$  diagonal matrix with elements

$$C_i(k, k) = \sum_{l=1}^{\mathcal{H}} \sum_{m=l+1}^{\mathcal{H}} \{\delta_k(l) - \delta_k(m)\}^2 P(u_k = l|H_i) \cdot P(u_k = m|H_i)$$

where  $i \in \{1, 2, \dots, \mathcal{H}\}$ . Since the matrix  $\mathbf{C}_i$  is diagonalized,  $J^{total}$  becomes a simple form even in multiple-hypotheses problem. We define multivariate normal distributions with the mean vectors,  $\boldsymbol{\mu}_i$ ,  $i \in \{1, 2, \dots, \mathcal{H}\}$ , and covariance matrices,  $\boldsymbol{\Sigma}_i$ ,  $i \in \{1, 2, \dots, \mathcal{H}\}$ . Then, by substituting the distributions into the definition of J-divergence, we get

$$\begin{aligned} &J(p(\mathbf{y}|H_i), p(\mathbf{y}|H_j)) \\ &= \frac{1}{2} \text{Tr} [\boldsymbol{\Sigma}_i^{-1} \boldsymbol{\Sigma}_j + \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\Sigma}_i] + \frac{1}{2} \text{Tr} [(\boldsymbol{\Sigma}_i^{-1} + \boldsymbol{\Sigma}_j^{-1})(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^T] - \mathcal{K} \end{aligned} \quad (\text{B.3})$$

where  $\mathcal{K}$  is the dimension of the covariance matrix [53].

Now using the mean matrices and covariance matrices from (B.1) and (B.2), we can

define the following terms in (B.3),

$$\begin{aligned}\Sigma_i^{-1}\Sigma_j &= (\mathbf{R} + \mathbf{BAC}_i\mathbf{A}^T\mathbf{B}^T)^{-1}(\mathbf{R} + \mathbf{BAC}_j\mathbf{A}^T\mathbf{B}^T), \\ \Sigma_j^{-1}\Sigma_i &= (\mathbf{R} + \mathbf{BAC}_j\mathbf{A}^T\mathbf{B}^T)^{-1}(\mathbf{R} + \mathbf{BAC}_i\mathbf{A}^T\mathbf{B}^T), \text{ and} \\ (\Sigma_i^{-1} + \Sigma_j^{-1})(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^T \\ &= \{(\mathbf{R} + \mathbf{BAC}_i\mathbf{A}^T\mathbf{B}^T) + (\mathbf{R} + \mathbf{BAC}_j\mathbf{A}^T\mathbf{B}^T)^{-1}\}\mathbf{BA}\boldsymbol{\xi}\boldsymbol{\xi}^T\mathbf{A}^T\mathbf{B}^T\end{aligned}$$

where  $\boldsymbol{\xi} = \boldsymbol{\xi}_i - \boldsymbol{\xi}_j$ .

Therefore, we have

$$\begin{aligned}J(p(\mathbf{y}|H_i), p(\mathbf{y}|H_j)) &= J^{ij} = \\ &= \frac{1}{2}\text{Tr} [(\mathbf{R} + \mathbf{BA}(\mathbf{C}_i + \boldsymbol{\xi}\boldsymbol{\xi}^T)\mathbf{A}^T\mathbf{B}^T) \cdot (\mathbf{R} + \mathbf{BAC}_j\mathbf{A}^T\mathbf{B}^T)^{-1}] + \\ &= \frac{1}{2}\text{Tr} [(\mathbf{R} + \mathbf{BA}(\mathbf{C}_j + \boldsymbol{\xi}\boldsymbol{\xi}^T)\mathbf{A}^T\mathbf{B}^T) \cdot (\mathbf{R} + \mathbf{BAC}_i\mathbf{A}^T\mathbf{B}^T)^{-1}] - \mathcal{K}\end{aligned}$$

where  $\text{Tr}[\cdot]$  denotes the matrix trace operation.

### B.3 Derivation of Case 1, (3.9)

In (3.8), let us consider  $J_1^{ij}$  for a large  $\mathcal{K}$ ,  $(\psi_{jk} - \psi_{ik} + \phi_k^{ij}) \geq 0$ , and  $PW_k^* > 0$ . We have

$$\begin{aligned}&= \frac{P_{tot} s_k \{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\}}{\mathcal{K} \cdot PW_k^*} - \frac{P_{tot} s_k \{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\}}{\mathcal{K} \cdot PW_k^* + P_{tot} s_k \psi_{ik}} \\ &= \frac{P_{tot}^2 s_k^2 \psi_{ik} \{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\}}{\mathcal{K}^2 \cdot PW_k^{*2} + \mathcal{K} \cdot PW_k^* P_{tot} s_k \psi_{ik}} \leq \frac{P_{tot}^2 s_k^2 \psi_{ik} \{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\}}{\mathcal{K}^2 \cdot PW_k^{*2}}\end{aligned}$$

where  $PW_k^* = \sum_{h=1}^{\mathcal{H}} \|\boldsymbol{\delta}_h\|^2 \cdot P(u_k = h)$ , since  $P_{tot}$ ,  $s_k$ ,  $\psi_{ik}$ , and  $PW_k^*$  are positive.

The inequality is arranged as follows

$$\begin{aligned}
& \frac{P_{tot} s_k \{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\}}{\mathcal{K} \cdot PW_k^*} - \frac{P_{tot}^2 s_k^2 \psi_{ik} \{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\}}{\mathcal{K}^2 \cdot PW_k^{*2}} \\
& \leq \frac{P_{tot} s_k \{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\}}{\mathcal{K} \cdot PW_k^* + P_{tot} s_k \psi_{ik}} \leq \frac{P_{tot} s_k \{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\}}{\mathcal{K} \cdot PW_k^*} \\
\Rightarrow & \sum_{k=1}^{\mathcal{K}} \frac{P_{tot} s_k \{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\}}{\mathcal{K} \cdot PW_k^*} - \sum_{k=1}^{\mathcal{K}} \frac{P_{tot}^2 s_k^2 \psi_{ik} \{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\}}{\mathcal{K}^2 \cdot PW_k^{*2}} \\
& \leq 2J_1^{ij} \leq \sum_{k=1}^{\mathcal{K}} \frac{P_{tot} s_k \{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\}}{\mathcal{K} \cdot PW_k^*}
\end{aligned}$$

since  $\sum_{k=1}^{\mathcal{K}} \frac{P_{tot} s_k \{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\}}{\mathcal{K} \cdot PW_k^* + P_{tot} s_k \psi_{ik}} \leq \sum_{k=1}^{\mathcal{K}} \frac{P_{tot} s_k \{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\}}{\mathcal{K} \cdot PW_k^*}$ .

#### B.4 Derivation of Case 2, (3.11)

In (3.8), let us consider  $J_1(\cdot)$  for a large  $\mathcal{K}$ ,  $(\psi_{jk} - \psi_{ik} + \phi_k^{ij}) \leq 0$ , and  $PW_k^* > 0$ . We have

$$\begin{aligned}
& \frac{P_{tot} s_k \{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\}}{\mathcal{K} \cdot PW_k^*} - \frac{P_{tot} s_k \{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\}}{\mathcal{K} \cdot PW_k^* + P_{tot} s_k \psi_{ik}} \\
= & \frac{P_{tot}^2 s_k^2 \psi_{ik} \{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\}}{\mathcal{K}^2 \cdot PW_k^{*2} + \mathcal{K} \cdot PW_k^* P_{tot} s_k \psi_{ik}} \geq \frac{P_{tot}^2 s_k^2 \psi_{ik} \{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\}}{\mathcal{K}^2 \cdot PW_k^{*2}}
\end{aligned}$$

where  $PW_k^* = \sum_{h=1}^{\mathcal{H}} \|\delta_h\|^2 \cdot P(u_k = h)$ , since  $P_{tot}$ ,  $s_k$ ,  $\psi_{ik}$ , and  $PW_k^*$  are positive. The inequality is arranged as follows

$$\begin{aligned}
& \frac{P_{tot} s_k \{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\}}{\mathcal{K} \cdot PW_k^*} - \frac{P_{tot}^2 s_k^2 \psi_{ik} \{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\}}{\mathcal{K}^2 \cdot PW_k^{*2}} \\
& \geq \frac{P_{tot} s_k \{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\}}{\mathcal{K} \cdot PW_k^* + P_{tot} s_k \psi_{ik}} \geq \frac{P_{tot} s_k \{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\}}{\mathcal{K} \cdot PW_k^*} \\
\Rightarrow & \sum_{k=1}^{\mathcal{K}} \frac{P_{tot} s_k \{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\}}{\mathcal{K} \cdot PW_k^*} \leq 2J_1^{ij} \leq \\
& \sum_{k=1}^{\mathcal{K}} \frac{P_{tot} s_k \{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\}}{\mathcal{K} \cdot PW_k^*} - \sum_{k=1}^{\mathcal{K}} \frac{P_{tot}^2 s_k^2 \psi_{ik} \{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\}}{\mathcal{K}^2 \cdot PW_k^{*2}} \tag{B.4}
\end{aligned}$$

$$\text{since } \sum_{k=1}^{\mathcal{K}} \frac{P_{tot} s_k \{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\}}{\mathcal{K} \cdot PW_k^*} \leq \sum_{k=1}^{\mathcal{K}} \frac{P_{tot} s_k \{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\}}{\mathcal{K} \cdot PW_k^* + P_{tot} s_k \psi_{ik}}.$$

### B.5 Derivation of Nondecreasing Characteristic of Total J-Divergence

The total J-divergence,  $J^{total}$  is a linear combination of  $J^{ij}$ . So, if  $J^{ij}$  is nondecreasing, then  $J^{total}$  is nondecreasing. The first partial derivative of  $J^{ij}(\cdot)$  with respect to  $P_k$  is

$$\begin{aligned}
& \frac{\partial}{\partial P_k} \left[ \sum_{k=1}^{\mathcal{K}} \frac{P_k s_k \{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\}}{PW_k^* + P_k s_k \{\psi_{ik}\}} + \sum_{k=1}^{\mathcal{K}} \frac{P_k s_k \{\psi_{ik} - \psi_{jk} + \phi_k^{ij}\}}{PW_k^* + P_k s_k \{\psi_{jk}\}} \right] \\
& = \frac{PW_k^*}{(PW_k^* + P_k s_k \{\psi_{ik}\})^2 (PW_k^* + P_k s_k \{\psi_{jk}\})^2} \\
& \quad \times \left[ 2PW_k^{*2} (\phi_k^{ij}) + 2P_k PW_k^* (\psi_{ik} + \psi_{jk}) + P_k^2 (\psi_{ik}^2 + \psi_{jk}^2) \right].
\end{aligned}$$

Therefore, the first partial derivative is nonnegative since  $PW_k^*$ ,  $\phi_k^{ij}$  and  $\psi_{ik}$  are nonnegative from (3.7). The total J-divergence is nondecreasing.

## B.6 Concave Condition of Total J-Divergence

To investigate the concavity of total J-divergence, we need to check whether the following second order partial derivative of the total J-divergence is non-positive.

$$\frac{\partial^2 J^{total}(\cdot)}{\partial P_k^2} = \frac{1}{2} \sum_{i=1}^{\mathcal{H}} \sum_{j=1}^{\mathcal{H}} \frac{\partial^2 J^{ij}(\cdot)}{\partial P_k^2} \quad (\text{B.5})$$

where the second order partial derivative of  $J^{ij}(\cdot)$  with respect to  $P_k$  is

$$\frac{\partial^2 J^{ij}(\cdot)}{\partial P_k^2} = \frac{\partial^2}{\partial P_k^2} \left[ \sum_{k=1}^{\mathcal{K}} \frac{P_k s_k \{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\}}{PW_k^* + P_k s_k \{\psi_{ik}\}} + \sum_{k=1}^{\mathcal{K}} \frac{P_k s_k \{\psi_{ik} - \psi_{jk} + \phi_k^{ij}\}}{PW_k^* + P_k s_k \{\psi_{jk}\}} \right] \quad (\text{B.6})$$

$$= \frac{-2PW_k^*}{(PW_k^* + P_k s_k \{\psi_{ik}\})^3 (PW_k^* + P_k s_k \{\psi_{jk}\})^3} [G_0^{ij} + G_1^{ij} P_k + G_2^{ij} P_k^2 + G_3^{ij} P_k^3] \quad (\text{B.7})$$

where

$$G_0^{ij} = s_k [(\psi_{jk} + \psi_{ik}) \phi_k^{ij} - (\psi_{jk} - \psi_{ik})^2],$$

$$G_1^{ij} = 6PW_k^{*2} s_k \psi_{ik} \psi_{jk} \phi_k^{ij},$$

$$G_2^{ij} = 3PW_k^* s_k \psi_{ik} \psi_{jk} [(\psi_{jk} - \psi_{ik})^2 + (\psi_{jk} + \psi_{ik}) \phi_k^{ij}],$$

$$G_3^{ij} = s_k [(\psi_{jk} + \psi_{ik})(\psi_{jk} - \psi_{ik})^2 + (\psi_{jk}^2 + \psi_{ik}^2) \phi_k^{ij}].$$

The concave region of  $J^{ij}(\cdot)$  depends on  $G_0^{ij}$  since  $M_{ik}$ ,  $G_1^{ij}$ ,  $G_2^{ij}$  and  $G_3^{ij}$  are non-negative from the result of B.5 and (3.7). If  $G_0^{ij}$  for all  $i, j$  are positive or zero, the second order partial derivative of total J-divergence becomes non-positive and total J-divergence becomes concave. However, this is a sufficient condition for concavity.

## APPENDIX C

## DERIVATIONS FOR CHAPTER 4

## C.1 Waveform design for monostatic radar with TSI

The system model is  $\mathbf{y} = \mathbf{W}\mathbf{G}\mathbf{x} + \mathbf{n}$  in frequency domain and the waveform design scheme is described by

$$\begin{aligned} \max \quad & TSI = I(\mathbf{y}; \mathbf{x}) \\ \text{s.t.} \quad & Tr[\mathbf{W}\mathbf{W}^T] \leq E_w, \end{aligned} \quad (\text{C.1})$$

where  $I(\mathbf{y}; \mathbf{x}) = h(\mathbf{y}) - h(\mathbf{y}|\mathbf{x})$  and

$$\begin{aligned} p(\mathbf{y}|\mathbf{x}) &= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{\mathcal{L}}{2}} \exp\left[-\frac{(\mathbf{y} - \mathbf{W}\mathbf{G}\mathbf{x})^T(\mathbf{y} - \mathbf{W}\mathbf{G}\mathbf{x})}{2\sigma^2}\right] \\ &= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{\mathcal{L}}{2}} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{\mathcal{L}} (y(i) - w(i)\{\mathbf{G}\mathbf{x}\}_i)^2\right]. \end{aligned} \quad (\text{C.2})$$

$\mathbf{y}$  and  $\mathbf{n}$  are  $\mathcal{L} \times 1$  vectors.  $\mathbf{W}$  is a  $\mathcal{L} \times \mathcal{L}$  diagonal matrix,  $\mathbf{G}$  is a  $\mathcal{L} \times \mathcal{H}$  matrix and  $x$  is a  $\mathcal{H} \times 1$  vector for  $\mathcal{H}$ -hypothesis detection system.

The waveform matrix  $\mathbf{W}$  is respectively defined such as

$$\mathbf{W} = \begin{bmatrix} w(1) & 0 & \dots & 0 \\ 0 & w(2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & w(\mathcal{L}) \end{bmatrix}.$$

After applying derivative to  $TSI$  with respect to each individual  $w(i)$  where  $i$  is a frequency index, we get

$$\begin{aligned}
\frac{\partial TSI}{\partial w(i)} &= \frac{\partial [h(\mathbf{y}) - h(\mathbf{y}|\mathbf{x})]}{\partial w(i)} \\
&= \frac{-\partial [E[\log_2 p(\mathbf{y})] + E[\log_2 p(\mathbf{y}|\mathbf{x})]]}{\partial w(i)} \\
&= -\frac{\partial E[\log_2 p(\mathbf{y})]}{\partial w(i)} \tag{C.3}
\end{aligned}$$

since  $-E[\log_2 p(\mathbf{y}|\mathbf{x})]$  is a fixed differential entropy of Gaussian noise. Then, we develop the derivative of TSI as follows.

$$\begin{aligned}
\frac{\partial TSI}{\partial w(i)} &= -\frac{\partial E[\log_2 p(\mathbf{y})]}{\partial w(i)} \\
&= -\frac{\partial}{\partial w(i)} \left[ \int p(\mathbf{y}) \log_2 p(\mathbf{y}) d\mathbf{y} \right] \\
&= -\int (1 + \log_2 p(\mathbf{y})) \frac{\partial p(\mathbf{y})}{\partial w(i)} d\mathbf{y} \\
&= -\int (1 + \log_2 p(\mathbf{y})) \frac{\partial \sum_{h=1}^{\mathcal{H}} p(\mathbf{y}|\mathbf{x}_h) P(\mathbf{x}_h)}{\partial w(i)} d\mathbf{y} \\
&= -\int (1 + \log_2 p(\mathbf{y})) \sum_{h=1}^{\mathcal{H}} \frac{\partial p(\mathbf{y}|\mathbf{x}_h)}{\partial w(i)} P(\mathbf{x}_h) d\mathbf{y} \tag{C.4}
\end{aligned}$$

where  $\mathbf{x}_h$  means  $\mathbf{x} = \mathbf{e}_h$  and  $\mathbf{e}_h$  is the  $h^{th}$  indicator vector. (C.4) needs be integrated with respect to  $\mathbf{y}$ . Thus, let us derive  $\frac{\partial p(\mathbf{y}|\mathbf{x}_h)}{\partial \mathbf{w}}$  with respect to  $\frac{\partial p(\mathbf{y}|\mathbf{x}_h)}{\partial \mathbf{y}}$  by defining  $\frac{\partial p(\mathbf{y}|\mathbf{x}_h)}{\partial \mathbf{w}} = \left[ \frac{\partial p(\mathbf{y}|\mathbf{x}_h)}{\partial w(1)}, \frac{\partial p(\mathbf{y}|\mathbf{x}_h)}{\partial w(2)}, \dots, \frac{\partial p(\mathbf{y}|\mathbf{x}_h)}{\partial w(\mathcal{L})} \right]^T$  since we can not find the component-wise relationship between  $\frac{\partial p(\mathbf{y}|\mathbf{x}_h)}{\partial w(i)}$  and  $\frac{\partial p(\mathbf{y}|\mathbf{x}_h)}{y(i)}$ . This vector relationship is from the fact that the components of the vector  $\mathbf{y}$  are not independent since  $\mathbf{y}$  is Gaussian Mixture.

From (C.2), we obtain

$$\frac{\partial p(\mathbf{y}|\mathbf{x})}{\partial w(i)} = p(\mathbf{y}|\mathbf{x}) \left\{ \frac{1}{\sigma^2} (y(i) - w(i) \cdot \{\mathbf{G}\mathbf{x}\}(i)) \{\mathbf{G}\mathbf{x}\}(i) \right\}. \quad (\text{C.5})$$

For a given  $\mathbf{x} = \mathbf{e}_h$ , we get

$$\frac{\partial p(\mathbf{y}|\mathbf{x}_h)}{\partial w(i)} = p(\mathbf{y}|\mathbf{x}_h) \left\{ \frac{1}{\sigma^2} (y(i) - w(i) \cdot g_h(i)) g_h(i) \right\} \quad (\text{C.6})$$

where  $g_h(i) = \{\mathbf{G}\mathbf{x}_h\}(i)$  and  $\{\mathbf{G}\mathbf{x}_h\}$  is a  $\mathcal{L} \times 1$  vector. On the other hand, from (C.2) we get

$$\frac{\partial p(\mathbf{y}|\mathbf{x})}{\partial \mathbf{y}} = p(\mathbf{y}|\mathbf{x}) \left\{ \frac{1}{-\sigma^2} (\mathbf{y} - \mathbf{W}\mathbf{G}\mathbf{x}) \right\} \quad (\text{C.7})$$

where  $\frac{\partial p(\mathbf{y}|\mathbf{x})}{\partial \mathbf{y}} = \left[ \frac{\partial p(\mathbf{y}|\mathbf{x})}{\partial y(1)}, \frac{\partial p(\mathbf{y}|\mathbf{x})}{\partial y(2)}, \dots, \frac{\partial p(\mathbf{y}|\mathbf{x})}{\partial y(\mathcal{L})} \right]^T$  and

$$\frac{\partial p(\mathbf{y}|\mathbf{x})}{\partial y(i)} = p(\mathbf{y}|\mathbf{x}) \left\{ \frac{1}{-\sigma^2} (y(i) - w(i) \cdot \{\mathbf{G}\mathbf{x}\}(i)) \right\}.$$

For a given  $\mathbf{x} = \mathbf{e}_h$ , we get

$$\frac{\partial p(\mathbf{y}|\mathbf{x}_h)}{\partial y(i)} = p(\mathbf{y}|\mathbf{x}_h) \left\{ \frac{1}{-\sigma^2} (y(i) - w(i) \cdot g_h(i)) \right\}. \quad (\text{C.8})$$

Therefore, from (C.6), (C.7), and (C.8), we get the following relationship

$$\frac{\partial p(\mathbf{y}|\mathbf{x} = \mathbf{e}_h)}{\partial \widetilde{\mathbf{W}}} = - \left( \begin{array}{cccc} g_h(1) & 0 & \dots & 0 \\ 0 & g_h(2) & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \dots & 0 & g_h(\mathcal{L}) \end{array} \right) \frac{\partial p(\mathbf{y}|\mathbf{x}_h)}{\partial \mathbf{y}}. \quad (\text{C.9})$$

Now, we continue to develop (C.4) as follows.

$$\begin{aligned}
\frac{\partial TSI}{\partial w(i)} &= - \int (1 + \log_2 p(\mathbf{y})) \sum_{h=1}^{\mathcal{H}} \frac{\partial p(\mathbf{y}|\mathbf{x}_h)}{\partial w(i)} P(\mathbf{x}_h) d\mathbf{y} \\
&= \sum_{h=1}^{\mathcal{H}} \left[ \int (1 + \log_2 p(\mathbf{y})) - \frac{\partial p(\mathbf{y}|\mathbf{x}_h)}{\partial w(i)} P(\mathbf{x}_h) d\mathbf{y} \right] \\
&= \sum_{h=1}^{\mathcal{H}} \left[ \int (1 + \log_2 p(\mathbf{y})) \frac{\partial p(\mathbf{y}|\mathbf{x}_h)}{\partial y(i)} g_h(i) P(\mathbf{x}_h) d\mathbf{y} \right] \\
&= \sum_{h=1}^{\mathcal{H}} P(\mathbf{x}_h) \left[ g_h(i) \int (1 + \log_2 p(\mathbf{y})) \frac{\partial p(\mathbf{y}|\mathbf{x}_h)}{\partial y(i)} d\mathbf{y} \right]
\end{aligned}$$

and with (C.9) we can get

$$\begin{aligned}
\frac{\partial TSI}{\partial \widetilde{\mathbf{W}}} &= \left[ \frac{\partial TSI}{\partial w(1)}, \frac{\partial TSI}{\partial w(2)}, \dots, \frac{\partial TSI}{\partial w(\mathcal{L})} \right]^T \\
&= \sum_{h=1}^{\mathcal{H}} P(\mathbf{x}_h) \cdot \text{diag}\{\mathbf{g}_h\} \int (1 + \log_2 p(\mathbf{y})) \left[ \frac{\partial p(\mathbf{y}|\mathbf{x}_h)}{\partial y(1)}, \frac{\partial p(\mathbf{y}|\mathbf{x}_h)}{\partial y(2)}, \dots, \frac{\partial p(\mathbf{y}|\mathbf{x}_h)}{\partial y(\mathcal{L})} \right]^T d\mathbf{y} \\
&= \sum_{h=1}^{\mathcal{H}} P(\mathbf{x}_h) \cdot \text{diag}\{\mathbf{g}_h\} \int (1 + \log_2 p(\mathbf{y})) \frac{\partial p(\mathbf{y}|\mathbf{x}_h)}{\partial \mathbf{y}} d\mathbf{y} \tag{C.10}
\end{aligned}$$

where the  $\text{diag}\{\mathbf{g}_h\}$  is a diagonal matrix including the vector  $\mathbf{g}_h = [g_h(1), g_h(2), \dots, g_h(\mathcal{L})]^T$  on the main diagonal. Using integration by parts, the integration within (C.10) becomes

$$\begin{aligned}
&\int (1 + \log_2 p(\mathbf{y})) \frac{\partial p(\mathbf{y}|\mathbf{x}_h)}{\partial \mathbf{y}} d\mathbf{y} \\
&= (1 + \log_2 p(\mathbf{y})) p(\mathbf{y}|\mathbf{x}_h) - \int \frac{\partial(1 + \log_2 p(\mathbf{y}))}{\partial \mathbf{y}} p(\mathbf{y}|\mathbf{x}_h) d\mathbf{y}
\end{aligned}$$

$$\begin{aligned}
&= - \int \frac{1}{p(\mathbf{y})} \frac{\partial p(\mathbf{y})}{\partial \mathbf{y}} p(\mathbf{y}|\mathbf{x}_h) d\mathbf{y} \\
&= - \int \frac{p(\mathbf{y}|\mathbf{x}_h)}{p(\mathbf{y})} \frac{\partial p(\mathbf{y})}{\partial \mathbf{y}} d\mathbf{y}
\end{aligned} \tag{C.11}$$

since  $(1 + \log_2 p(\mathbf{y})) p(\mathbf{y}|\mathbf{x}_h)$  goes to 0 as  $\|\mathbf{y}\|$  goes to  $\infty$  [35][11].

By substituting (C.11) into (C.10), we get

$$\begin{aligned}
&\frac{\partial TSI}{\partial \widetilde{\mathbf{W}}} \\
&= - \sum_{h=1}^{\mathcal{H}} P(\mathbf{x}_h) \cdot \text{diag}\{\mathbf{g}_h\} \left[ \int \frac{p(\mathbf{y}|\mathbf{x}_h)}{p(\mathbf{y})} \frac{\partial p(\mathbf{y})}{\partial \mathbf{y}} d\mathbf{y} \right] \\
&= - \int \sum_{h=1}^{\mathcal{H}} \text{diag}\{\mathbf{g}_h\} \frac{\partial p(\mathbf{y})}{\partial \mathbf{y}} \left[ \frac{p(\mathbf{y}|\mathbf{x}_h)}{p(\mathbf{y})} \right] P(\mathbf{x}_h) d\mathbf{y} \\
&= - \int \mathbb{E}_{\mathbf{x}|\mathbf{y}} [\text{diag}\{\mathbf{G}\mathbf{x}\}] \frac{\partial p(\mathbf{y})}{\partial \mathbf{y}} d\mathbf{y} \\
&= - \int \mathbb{E}_{\mathbf{x}|\mathbf{y}} [\text{diag}\{\mathbf{G}\mathbf{x}\}] \sum_{h=1}^{\mathcal{H}} \frac{\partial p(\mathbf{y}|\mathbf{x}_h)}{\partial \mathbf{y}} P(\mathbf{x}_h) d\mathbf{y} \\
&= - \int \mathbb{E}_{\mathbf{x}|\mathbf{y}} [\text{diag}\{\mathbf{G}\mathbf{x}\}] \sum_{h=1}^{\mathcal{H}} p(\mathbf{y}|\mathbf{x}_h) \left\{ \frac{1}{-\sigma^2} (\mathbf{y} - \mathbf{W}\mathbf{G}\mathbf{x}_h) \right\} P(\mathbf{x}_h) d\mathbf{y} \\
&= - \int \mathbb{E}_{\mathbf{x}|\mathbf{y}} [\text{diag}\{\mathbf{G}\mathbf{x}\}] \sum_{h=1}^{\mathcal{H}} p(\mathbf{x}_h|\mathbf{y}) p(\mathbf{y}) \left\{ \frac{1}{-\sigma^2} (\mathbf{y} - \mathbf{W}\mathbf{G}\mathbf{x}_h) \right\} d\mathbf{y} \\
&= \frac{1}{\sigma^2} \int \mathbb{E}_{\mathbf{x}|\mathbf{y}} [\text{diag}\{\mathbf{G}\mathbf{x}\}] \mathbb{E}_{\mathbf{x}|\mathbf{y}} [p(\mathbf{y})\mathbf{y} - p(\mathbf{y})\mathbf{W}\mathbf{G}\mathbf{x}] d\mathbf{y} \\
&= \frac{1}{\sigma^2} \int \mathbb{E}_{\mathbf{x}|\mathbf{y}} [\text{diag}\{\mathbf{G}\mathbf{x}\}] [p(\mathbf{y})\mathbf{y} - p(\mathbf{y})\mathbf{W}\mathbb{E}_{\mathbf{x}|\mathbf{y}} [\mathbf{G}\mathbf{x}]] d\mathbf{y} \\
&= \frac{1}{\sigma^2} \int \mathbb{E}_{\mathbf{x}|\mathbf{y}} [\text{diag}\{\mathbf{G}\mathbf{x}\}] \cdot p(\mathbf{y})\mathbf{y} - \mathbb{E}_{\mathbf{x}|\mathbf{y}} [\text{diag}\{\mathbf{G}\mathbf{x}\}] \cdot p(\mathbf{y})\mathbf{W}\mathbb{E}_{\mathbf{x}|\mathbf{y}} [\mathbf{G}\mathbf{x}] d\mathbf{y} \tag{C.12}
\end{aligned}$$

The first part of (C.12) can be evaluated as follows.

$$\begin{aligned}
& \int \mathbb{E}_{\mathbf{x}|\mathbf{y}} [\text{diag}\{\mathbf{G}\mathbf{x}\}] \cdot p(\mathbf{y})\mathbf{y}d\mathbf{y} \\
&= \int \sum_{h=1}^{\mathcal{H}} p(\mathbf{x}_h|\mathbf{y}) \cdot \text{diag}\{\mathbf{G}\mathbf{x}_h\}p(\mathbf{y})\mathbf{y}d\mathbf{y} \\
&= \int \sum_{h=1}^{\mathcal{H}} p(\mathbf{x}_h, \mathbf{y}) \cdot \text{diag}\{\mathbf{G}\mathbf{x}_h\}\mathbf{y}d\mathbf{y} \\
&= \int \sum_{h=1}^{\mathcal{H}} p(\mathbf{y}|\mathbf{x}_h) \cdot \text{diag}\{\mathbf{G}\mathbf{x}_h\}P(\mathbf{x}_h)\mathbf{y}d\mathbf{y} \\
&= \sum_{h=1}^{\mathcal{H}} P(\mathbf{x}_h) \cdot \text{diag}\{\mathbf{G}\mathbf{x}_h\} \int p(\mathbf{y}|\mathbf{x}_h)\mathbf{y}d\mathbf{y} \\
&= \sum_{h=1}^{\mathcal{H}} P(\mathbf{x}_h) \cdot \text{diag}\{\mathbf{G}\mathbf{x}_h\} \mathbf{W}\{\mathbf{G}\mathbf{x}_h\} \\
&= \mathbf{W} \sum_{h=1}^{\mathcal{H}} P(\mathbf{x}_h) \cdot \text{diag}\{\mathbf{G}\mathbf{x}_h\} \cdot \{\mathbf{G}\mathbf{x}_h\} \\
&= \mathbf{W}\mathbb{E}_{\mathbf{x}} [\text{diag}\{\mathbf{G}\mathbf{x}_h\} \cdot \{\mathbf{G}\mathbf{x}_h\}]. \tag{C.13}
\end{aligned}$$

since  $\int p(\mathbf{y}|\mathbf{x}_h)\mathbf{y}d\mathbf{y} = \mathbf{W}\{\mathbf{G}\mathbf{x}_h\}$ .

The second part of (C.12) can be evaluated as follows.

$$\begin{aligned}
& \int \mathbb{E}_{\mathbf{x}|\mathbf{y}} [\text{diag}\{\mathbf{G}\mathbf{x}\}] \cdot p(\mathbf{y})\mathbf{W}\mathbb{E}_{\mathbf{x}|\mathbf{y}} [\mathbf{G}\mathbf{x}] d\mathbf{y} \\
&= \mathbf{W} \int \sum_{h=1}^{\mathcal{H}} p(\mathbf{y}, \mathbf{x}_h)\mathbb{E}_{\mathbf{x}|\mathbf{y}} [\text{diag}\{\mathbf{G}\mathbf{x}\}] \mathbb{E}_{\mathbf{x}|\mathbf{y}} [\mathbf{G}\mathbf{x}] d\mathbf{y} \\
&= \mathbf{W} \cdot \mathbb{E}_{\mathbf{x},\mathbf{y}} [\mathbb{E}_{\mathbf{x}|\mathbf{y}} [\text{diag}\{\mathbf{G}\mathbf{x}\}] \mathbb{E}_{\mathbf{x}|\mathbf{y}} [\mathbf{G}\mathbf{x}]] \tag{C.14}
\end{aligned}$$

since  $p(\mathbf{y}) = \sum_{h=1}^{\mathcal{H}} p(\mathbf{y}, \mathbf{x}_h)$  and  $\mathbf{W}$  is a diagonal matrix.

From (C.12), (C.13), and (C.14), we obtain

$$\begin{aligned}\frac{\partial TSI}{\partial \widetilde{\mathbf{W}}} &= \frac{\mathbf{W}}{\sigma^2} \left\{ \mathbb{E}_{\mathbf{x}} [\text{diag}\{\mathbf{G}\mathbf{x}\} \cdot \{\mathbf{G}\mathbf{x}\}] - \mathbb{E}_{\mathbf{x},\mathbf{y}} [\mathbb{E}_{\mathbf{x}|\mathbf{y}} [\text{diag}\{\mathbf{G}\mathbf{x}\}] \cdot \mathbb{E}_{\mathbf{x}|\mathbf{y}} [\mathbf{G}\mathbf{x}]] \right\} \\ &= \frac{\mathbf{W}}{\sigma^2} \mathbb{E}_{\mathbf{x},\mathbf{y}} [\text{diag}\{\mathbf{G}\mathbf{x}\} \cdot \{\mathbf{G}\mathbf{x}\} - \mathbb{E}_{\mathbf{x}|\mathbf{y}} [\text{diag}\{\mathbf{G}\mathbf{x}\}] \cdot \mathbb{E}_{\mathbf{x}|\mathbf{y}} [\mathbf{G}\mathbf{x}]]\end{aligned}\quad (\text{C.15})$$

where  $\frac{\partial TSI}{\partial \widetilde{\mathbf{W}}} = \left[ \frac{\partial TSI}{\partial w(1)}, \frac{\partial TSI}{\partial w(2)}, \dots, \frac{\partial TSI}{\partial w(\mathcal{L})} \right]^T$ ,  $\mathbb{E}_{\mathbf{x}|\mathbf{y}} [\mathbf{G}\mathbf{x}]$  is calculated by

$$\mathbb{E}_{\mathbf{x}|\mathbf{y}} [\mathbf{G}\mathbf{x}] = \frac{\sum_{h=1}^{\mathcal{H}} \{\mathbf{G}\mathbf{x}_h\} p(\mathbf{y}|\mathbf{x}_h) P(\mathbf{x}_h)}{\sum_{h=1}^{\mathcal{H}} p(\mathbf{y}|\mathbf{x}_h) P(\mathbf{x}_h)},$$

and  $\mathbb{E}_{\mathbf{x}|\mathbf{y}} [\text{diag}\{\mathbf{G}\mathbf{x}\}]$  is similarly calculated.

For numerical programming, (C.15) can be written as follows.

$$\frac{\partial TSI}{\partial \widetilde{\mathbf{W}}} = \frac{\mathbf{W}}{N_{total}\sigma^2} \sum_{\mathbf{y}_0, \mathbf{x}_0} \left\{ \text{diag}\{\mathbf{G}\mathbf{x}_0\} \cdot \{\mathbf{G}\mathbf{x}_0\} - \mathbb{E}_{\mathbf{x}|\mathbf{y}_0} [\text{diag}\{\mathbf{G}\mathbf{x}\}] \cdot \mathbb{E}_{\mathbf{x}|\mathbf{y}_0} [\mathbf{G}\mathbf{x}] \right\}\quad (\text{C.16})$$

where  $N_{total}$  is total iteration for estimating the expectations and

$$\mathbb{E}_{\mathbf{x}|\mathbf{y}_0} [\mathbf{G}\mathbf{x}] = \frac{\sum_{h=1}^{\mathcal{H}} \{\mathbf{G}\mathbf{x}_h\} p(\mathbf{y}_0|\mathbf{x}_h) P(\mathbf{x}_h)}{\sum_{h=1}^{\mathcal{H}} p(\mathbf{y}_0|\mathbf{x}_h) P(\mathbf{x}_h)}.\quad (\text{C.17})$$

## C.2 Waveform design for SIMO radar with TSI

The system model shown in (4.13) for SIMO radar configuration is

$$\hat{\mathbf{y}} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_{N_{\text{Rx}}} \end{bmatrix} = \widehat{\mathbf{W}}\widehat{\mathbf{G}}\mathbf{x} + \widehat{\mathbf{n}} = \begin{bmatrix} \mathbf{W}\mathbf{G}_1\mathbf{x} + \mathbf{n}_1 \\ \mathbf{W}\mathbf{G}_2\mathbf{x} + \mathbf{n}_2 \\ \vdots \\ \mathbf{W}\mathbf{G}_{N_{\text{Rx}}}\mathbf{x} + \mathbf{n}_{N_{\text{Rx}}} \end{bmatrix} \quad (\text{C.18})$$

From (C.3), we get

$$\begin{aligned} \frac{\partial TSI}{\partial w(i)} &= -\frac{\partial E[\log_2 p(\widehat{\mathbf{y}})]}{\partial w(i)} \\ &= -\frac{\partial}{\partial w(i)} \left[ \int p(\widehat{\mathbf{y}}) \log_2 p(\widehat{\mathbf{y}}) d\widehat{\mathbf{y}} \right] \\ &= -\int (1 + \log_2 p(\widehat{\mathbf{y}})) \frac{\partial p(\widehat{\mathbf{y}})}{\partial w(i)} d\widehat{\mathbf{y}} \\ &= -\int (1 + \log_2 p(\widehat{\mathbf{y}})) \frac{\partial \sum_{h=1}^{\mathcal{H}} p(\widehat{\mathbf{y}}|\mathbf{x}_h) P(\mathbf{x}_h)}{\partial w(i)} d\widehat{\mathbf{y}} \\ &= -\int (1 + \log_2 p(\widehat{\mathbf{y}})) \sum_{h=1}^{\mathcal{H}} \frac{\partial p(\widehat{\mathbf{y}}|\mathbf{x}_h)}{\partial w(i)} P(\mathbf{x}_h) d\widehat{\mathbf{y}} \end{aligned} \quad (\text{C.19})$$

where  $\mathbf{x}_h$  means  $\mathbf{x} = \mathbf{e}_h$  and  $\mathbf{e}_h$  is the  $h^{\text{th}}$  indicator vector.

Let us derive the  $\frac{\partial p(\widehat{\mathbf{y}}|\mathbf{x}_h)}{\partial \widehat{\mathbf{W}}}$  with respect to  $\frac{\partial p(\widehat{\mathbf{y}}|\mathbf{x}_h)}{\partial \widehat{\mathbf{y}}}$  by defining

$$\frac{\partial p(\widehat{\mathbf{y}}|\mathbf{x}_h)}{\partial \widehat{\mathbf{W}}} = \left[ \frac{\partial p(\widehat{\mathbf{y}}|\mathbf{x}_h)}{\partial w(1)}, \frac{\partial p(\widehat{\mathbf{y}}|\mathbf{x}_h)}{\partial w(2)}, \dots, \frac{\partial p(\widehat{\mathbf{y}}|\mathbf{x}_h)}{\partial w(\mathcal{L})} \right]^T. \quad (\text{C.20})$$

From (4.15), we obtain

$$\begin{aligned} \frac{\partial p(\hat{\mathbf{y}}|\mathbf{x}_h)}{\partial w(i)} = & \\ p(\hat{\mathbf{y}}|\mathbf{x}_h) \left[ \frac{1}{\sigma^2} \sum_{m=1}^{N_{\text{Rx}}} \left\{ \hat{y}(i + \mathcal{L} \cdot (m-1)) - w(i) \hat{g}_j(i + \mathcal{L} \cdot (m-1)) \right\} \hat{g}_j(i + \mathcal{L} \cdot (m-1)) \right] & \text{for all } i \end{aligned} \quad (\text{C.21})$$

where  $\hat{g}_j(l) = \{\widehat{\mathbf{G}}_{\mathbf{x}_h}\}(l)$  for  $l = 1, 2, \dots, \mathcal{L}N_{\text{Rx}}$  and  $\{\widehat{\mathbf{G}}_{\mathbf{x}_h}\}$  is a  $\mathcal{L}N_{\text{Rx}} \times 1$  vector.

On the other hand, from (4.15) we get

$$\frac{\partial p(\hat{\mathbf{y}}|\mathbf{x})}{\partial \hat{\mathbf{y}}} = p(\hat{\mathbf{y}}|\mathbf{x}) \left\{ \frac{1}{-\sigma^2} (\hat{\mathbf{y}} - \widehat{\mathbf{W}}\widehat{\mathbf{G}}_{\mathbf{x}}) \right\} \quad (\text{C.22})$$

where  $\frac{\partial p(\hat{\mathbf{y}}|\mathbf{x})}{\partial \hat{\mathbf{y}}} = \left[ \frac{\partial p(\hat{\mathbf{y}}|\mathbf{x})}{\partial \hat{y}(1)}, \frac{\partial p(\hat{\mathbf{y}}|\mathbf{x})}{\partial \hat{y}(2)}, \dots, \frac{\partial p(\hat{\mathbf{y}}|\mathbf{x})}{\partial \hat{y}(\mathcal{L}N_{\text{Rx}})} \right]^T$  and

$$\frac{\partial p(\hat{\mathbf{y}}|\mathbf{x}_h)}{\partial \hat{y}(l)} = p(\hat{\mathbf{y}}|\mathbf{x}_h) \left\{ \frac{1}{-\sigma^2} (\hat{y}(l) - w(l') \cdot \hat{g}_j(l)) \right\}$$

where  $l'$  is  $l$  modulo  $\mathcal{L}$ , and  $l = 1, 2, \dots, \mathcal{L}N_{\text{Rx}}$ .

Therefore, from (C.21) and (C.22), we get

$$\frac{\partial p(\hat{\mathbf{y}}|\mathbf{x}_h)}{\partial \widehat{\mathbf{W}}} = -\widehat{\mathbf{B}}_{\mathbf{x}_h} \frac{\partial p(\hat{\mathbf{y}}|\mathbf{x}_h)}{\partial \hat{\mathbf{y}}} \text{ for } \mathbf{x}_h \quad (\text{C.23})$$

where the block matrix  $\widehat{\mathbf{B}}_{\mathbf{x}_h} = [\mathbf{B}_{1,h} \mathbf{B}_{2,h} \dots \mathbf{B}_{N_{\text{Rx}},h}]$  and the matrix  $\mathbf{B}_{m,h}$  is as follows.

$$\mathbf{B}_{m,h} = \text{diag}\{\mathbf{G}_{m,\mathbf{x}_h}\} = \text{diag}\{\mathbf{g}_{m,h}\} = \begin{bmatrix} g_{m,h}(1) & 0 & \dots & 0 \\ 0 & g_{m,h}(2) & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \dots & 0 & g_{m,h}(\mathcal{L}) \end{bmatrix}, \quad (\text{C.24})$$

for  $m = 1, 2, \dots, N_{\text{Rx}}$ . The vector  $\mathbf{g}_{m,h}$  is  $\{\mathbf{G}_{m,\mathbf{x}_h}\}$  and the  $h^{\text{th}}$  target transfer vector triggered by  $\mathbf{x} = \mathbf{e}_h$  from the path between the single transmitter to the  $m^{\text{th}}$  receiver.

With (C.23) and (C.20), we continue to develop (C.19) as follows.

$$\begin{aligned} \frac{\partial TSI}{\partial \widetilde{\mathbf{W}}} &= - \int (1 + \log_2 p(\widehat{\mathbf{y}})) \sum_{h=1}^{\mathcal{H}} \frac{\partial p(\widehat{\mathbf{y}}|\mathbf{x}_h)}{\partial \widetilde{\mathbf{W}}} P(\mathbf{x}_h) d\widehat{\mathbf{y}} \\ &= \sum_{h=1}^{\mathcal{H}} \left[ \int (1 + \log_2 p(\widehat{\mathbf{y}})) \left\{ -\frac{\partial p(\widehat{\mathbf{y}}|\mathbf{x}_h)}{\partial \widetilde{\mathbf{W}}} \right\} P(\mathbf{x}_h) d\widehat{\mathbf{y}} \right] \\ &= \sum_{h=1}^{\mathcal{H}} \left[ \int (1 + \log_2 p(\widehat{\mathbf{y}})) \left\{ \widehat{\mathbf{B}}_{\mathbf{x}_h} \frac{\partial p(\widehat{\mathbf{y}}|\mathbf{x}_h)}{\partial \widehat{\mathbf{y}}} \right\} P(\mathbf{x}_h) d\widehat{\mathbf{y}} \right] \\ &= \sum_{h=1}^{\mathcal{H}} P(\mathbf{x}_h) \widehat{\mathbf{B}}_{\mathbf{x}_h} \left[ \int (1 + \log_2 p(\widehat{\mathbf{y}})) \frac{\partial p(\widehat{\mathbf{y}}|\mathbf{x}_h)}{\partial \widehat{\mathbf{y}}} d\widehat{\mathbf{y}} \right]. \end{aligned} \quad (\text{C.25})$$

Using integration by parts, the integration part of (C.25) becomes

$$\begin{aligned} &\int (1 + \log_2 p(\widehat{\mathbf{y}})) \frac{\partial p(\widehat{\mathbf{y}}|\mathbf{x}_h)}{\partial \widehat{\mathbf{y}}} d\widehat{\mathbf{y}} \\ &= (1 + \log_2 p(\widehat{\mathbf{y}})) p(\widehat{\mathbf{y}}|\mathbf{x}_h) - \int \frac{\partial(1 + \log_2 p(\widehat{\mathbf{y}}))}{\partial \widehat{\mathbf{y}}} p(\widehat{\mathbf{y}}|\mathbf{x}_h) d\widehat{\mathbf{y}} \\ &= - \int \frac{1}{p(\widehat{\mathbf{y}})} \frac{\partial p(\widehat{\mathbf{y}})}{\partial \widehat{\mathbf{y}}} p(\widehat{\mathbf{y}}|\mathbf{x}_h) d\widehat{\mathbf{y}} \\ &= - \int \frac{p(\widehat{\mathbf{y}}|\mathbf{x}_h)}{p(\widehat{\mathbf{y}})} \frac{\partial p(\widehat{\mathbf{y}})}{\partial \widehat{\mathbf{y}}} d\widehat{\mathbf{y}} \end{aligned} \quad (\text{C.26})$$

since  $(1 + \log_2 p(\widehat{\mathbf{y}})) p(\widehat{\mathbf{y}}|\mathbf{x}_h)$  goes to 0 as  $\|\widehat{\mathbf{y}}\|$  goes to  $\infty$  [35][11].

By substituting (C.26) into (C.25), we get

$$\begin{aligned}
& \frac{\partial TSI}{\partial \widehat{\mathbf{W}}} \\
&= - \sum_{h=1}^{\mathcal{H}} P(\mathbf{x}_h) \cdot \widehat{\mathbf{B}}\mathbf{x}_h \left[ \int \frac{p(\widehat{\mathbf{y}}|\mathbf{x}_h)}{p(\widehat{\mathbf{y}})} \frac{\partial p(\widehat{\mathbf{y}})}{\partial \mathbf{y}} d\mathbf{y} \right] \\
&= - \int \sum_{h=1}^{\mathcal{H}} \widehat{\mathbf{B}}\mathbf{x}_h \frac{\partial p(\widehat{\mathbf{y}})}{\partial \widehat{\mathbf{y}}} \left[ \frac{p(\widehat{\mathbf{y}}|\mathbf{x}_h)}{p(\widehat{\mathbf{y}})} \right] P(\mathbf{x}_h) d\widehat{\mathbf{y}} \\
&= - \int \mathbb{E}_{\mathbf{x}|\widehat{\mathbf{y}}} [\widehat{\mathbf{B}}\mathbf{x}] \frac{\partial p(\widehat{\mathbf{y}})}{\partial \widehat{\mathbf{y}}} d\widehat{\mathbf{y}} \\
&= - \int \mathbb{E}_{\mathbf{x}|\widehat{\mathbf{y}}} [\widehat{\mathbf{B}}\mathbf{x}] \sum_{h=1}^{\mathcal{H}} \frac{\partial p(\widehat{\mathbf{y}}|\mathbf{x}_h)}{\partial \widehat{\mathbf{y}}} P(\mathbf{x}_h) d\widehat{\mathbf{y}} \\
&= - \int \mathbb{E}_{\mathbf{x}|\widehat{\mathbf{y}}} [\widehat{\mathbf{B}}\mathbf{x}] \sum_{h=1}^{\mathcal{H}} p(\widehat{\mathbf{y}}|\mathbf{x}_h) \left\{ \frac{1}{-\sigma^2} (\widehat{\mathbf{y}} - \widehat{\mathbf{W}}\widehat{\mathbf{G}}\mathbf{x}_h) \right\} P(\mathbf{x}_h) d\widehat{\mathbf{y}} \\
&= - \int \mathbb{E}_{\mathbf{x}|\widehat{\mathbf{y}}} [\widehat{\mathbf{B}}\mathbf{x}] \sum_{h=1}^{\mathcal{H}} p(\mathbf{x}_h|\widehat{\mathbf{y}}) p(\widehat{\mathbf{y}}) \left\{ \frac{1}{-\sigma^2} (\widehat{\mathbf{y}} - \widehat{\mathbf{W}}\widehat{\mathbf{G}}\mathbf{x}_h) \right\} d\widehat{\mathbf{y}} \\
&= \frac{1}{\sigma^2} \int \mathbb{E}_{\mathbf{x}|\widehat{\mathbf{y}}} [\widehat{\mathbf{B}}\mathbf{x}] \mathbb{E}_{\mathbf{x}|\widehat{\mathbf{y}}} [p(\widehat{\mathbf{y}})\widehat{\mathbf{y}} - p(\widehat{\mathbf{y}})\widehat{\mathbf{W}}\widehat{\mathbf{G}}\mathbf{x}] d\widehat{\mathbf{y}} \\
&= \frac{1}{\sigma^2} \int \mathbb{E}_{\mathbf{x}|\widehat{\mathbf{y}}} [\widehat{\mathbf{B}}\mathbf{x}] \left[ p(\widehat{\mathbf{y}})\widehat{\mathbf{y}} - p(\widehat{\mathbf{y}})\widehat{\mathbf{W}}\mathbb{E}_{\mathbf{x}|\widehat{\mathbf{y}}} [\widehat{\mathbf{G}}\mathbf{x}] \right] d\widehat{\mathbf{y}} \\
&= \frac{1}{\sigma^2} \int \mathbb{E}_{\mathbf{x}|\widehat{\mathbf{y}}} [\widehat{\mathbf{B}}\mathbf{x}] \cdot p(\widehat{\mathbf{y}})\widehat{\mathbf{y}} - \mathbb{E}_{\mathbf{x}|\widehat{\mathbf{y}}} [\widehat{\mathbf{B}}\mathbf{x}] \cdot p(\widehat{\mathbf{y}})\widehat{\mathbf{W}}\mathbb{E}_{\mathbf{x}|\widehat{\mathbf{y}}} [\widehat{\mathbf{G}}\mathbf{x}] d\widehat{\mathbf{y}} \quad (\text{C.27})
\end{aligned}$$

The first part of (C.27) can be evaluated as follows.

$$\int \mathbb{E}_{\mathbf{x}|\widehat{\mathbf{y}}} [\widehat{\mathbf{B}}\mathbf{x}] \cdot p(\widehat{\mathbf{y}})\widehat{\mathbf{y}} d\widehat{\mathbf{y}}$$

$$\begin{aligned}
&= \int \sum_{h=1}^{\mathcal{H}} p(\mathbf{x}_h|\hat{\mathbf{y}}) \widehat{\mathbf{B}}_{\mathbf{x}_h} p(\hat{\mathbf{y}}) \hat{\mathbf{y}} d\hat{\mathbf{y}} \\
&= \int \sum_{h=1}^{\mathcal{H}} p(\hat{\mathbf{y}}|\mathbf{x}_h) \widehat{\mathbf{B}}_{\mathbf{x}_h} P(\mathbf{x}_h) \hat{\mathbf{y}} d\hat{\mathbf{y}} \\
&= \sum_{h=1}^{\mathcal{H}} P(\mathbf{x}_h) \widehat{\mathbf{B}}_{\mathbf{x}_h} \int p(\hat{\mathbf{y}}|\mathbf{x}_h) \hat{\mathbf{y}} d\hat{\mathbf{y}} \\
&= \sum_{h=1}^{\mathcal{H}} P(\mathbf{x}_h) \widehat{\mathbf{B}}_{\mathbf{x}_h} \widehat{\mathbf{W}} \{\widehat{\mathbf{G}}_{\mathbf{x}_h}\} \\
&= \mathbb{E}_x \left[ \widehat{\mathbf{B}}_{\mathbf{x}} \cdot \widehat{\mathbf{W}} \cdot \{\widehat{\mathbf{G}}_{\mathbf{x}}\} \right] \\
&= \mathbf{W} \cdot \mathbb{E}_x \left[ \widehat{\mathbf{B}}_{\mathbf{x}} \{\widehat{\mathbf{G}}_{\mathbf{x}}\} \right] \tag{C.28}
\end{aligned}$$

since  $\int p(\hat{\mathbf{y}}|\mathbf{x}_h) \hat{\mathbf{y}} d\hat{\mathbf{y}} = \widehat{\mathbf{W}} \{\widehat{\mathbf{G}}_{\mathbf{x}_h}\}$ . We can pull out  $\mathbf{W}$  from the center between  $\mathcal{L} \times \mathcal{L} N_{\mathbf{R}\mathbf{x}}$  matrix  $\widehat{\mathbf{B}}_{\mathbf{x}_h}$  and  $\mathcal{L} N_{\mathbf{R}\mathbf{x}} \times 1$  vector  $\{\widehat{\mathbf{G}}_{\mathbf{x}_h}\}$  by changing  $\widehat{\mathbf{W}}$  to  $\mathbf{W}$  since  $\widehat{\mathbf{W}}$  is  $\mathbf{I}_{N_{\mathbf{R}\mathbf{x}}} \otimes \mathbf{W}$ , the block matrix  $\widehat{\mathbf{B}}_{\mathbf{x}_h} = [\mathbf{B}_{1 \cdot h} \mathbf{B}_{2 \cdot h} \dots \mathbf{B}_{N_{\mathbf{R}\mathbf{x}} \cdot h}]$ , and  $\mathbf{W}$  is  $\mathcal{L} \times \mathcal{L}$  diagonal matrix.

The second part of (C.27) can be evaluated as follows.

$$\begin{aligned}
&\int \mathbb{E}_{\mathbf{x}|\hat{\mathbf{y}}} \left[ \widehat{\mathbf{B}}_{\mathbf{x}} \right] \cdot p(\hat{\mathbf{y}}) \widehat{\mathbf{W}} \mathbb{E}_{\mathbf{x}|\hat{\mathbf{y}}} \left[ \widehat{\mathbf{G}}_{\mathbf{x}} \right] d\hat{\mathbf{y}} \\
&= \mathbf{W} \cdot \int \sum_{h=1}^{\mathcal{H}} p(\hat{\mathbf{y}}, \mathbf{x}_h) \mathbb{E}_{\mathbf{x}|\hat{\mathbf{y}}} \left[ \widehat{\mathbf{B}}_{\mathbf{x}_h} \right] \mathbb{E}_{\mathbf{x}|\hat{\mathbf{y}}} \left[ \widehat{\mathbf{G}}_{\mathbf{x}} \right] d\hat{\mathbf{y}} \\
&= \mathbf{W} \cdot \mathbb{E}_{\mathbf{x}, \hat{\mathbf{y}}} \left[ \mathbb{E}_{\mathbf{x}|\hat{\mathbf{y}}} \left[ \widehat{\mathbf{B}}_{\mathbf{x}} \right] \mathbb{E}_{\mathbf{x}|\hat{\mathbf{y}}} \left[ \widehat{\mathbf{G}}_{\mathbf{x}} \right] \right] \tag{C.29}
\end{aligned}$$

since  $p(\mathbf{y}) = \sum_{h=1}^{\mathcal{H}} p(\hat{\mathbf{y}}, \mathbf{x}_h)$  and  $\mathbf{W}$  is a diagonal matrix.

From (C.28), (C.29) and (C.27), we obtain

$$\begin{aligned}
\frac{\partial TSI}{\partial \widehat{\mathbf{W}}} &= \frac{\mathbf{W}}{\sigma^2} \left[ \mathbb{E}_{\mathbf{x}} \left[ \widehat{\mathbf{B}}_{\mathbf{x}} \cdot \{\widehat{\mathbf{G}}_{\mathbf{x}}\} \right] - \mathbb{E}_{\mathbf{x}, \hat{\mathbf{y}}} \left[ \mathbb{E}_{\mathbf{x}|\hat{\mathbf{y}}} \left[ \widehat{\mathbf{B}}_{\mathbf{x}} \right] \cdot \mathbb{E}_{\mathbf{x}|\hat{\mathbf{y}}} \left[ \widehat{\mathbf{G}}_{\mathbf{x}} \right] \right] \right] \\
&= \frac{\mathbf{W}}{\sigma^2} \mathbb{E}_{\mathbf{x}, \hat{\mathbf{y}}} \left[ \widehat{\mathbf{B}}_{\mathbf{x}} \cdot \{\widehat{\mathbf{G}}_{\mathbf{x}}\} - \mathbb{E}_{\mathbf{x}|\hat{\mathbf{y}}} \left[ \widehat{\mathbf{B}}_{\mathbf{x}} \right] \cdot \mathbb{E}_{\mathbf{x}|\hat{\mathbf{y}}} \left[ \widehat{\mathbf{G}}_{\mathbf{x}} \right] \right] \tag{C.30}
\end{aligned}$$

where

$$\mathbb{E}_{\mathbf{x}|\hat{\mathbf{y}}}\left[\widehat{\mathbf{G}}\mathbf{x}\right] = \frac{\sum_{h=1}^{\mathcal{H}}\{\widehat{\mathbf{G}}\mathbf{x}_h\}p(\hat{\mathbf{y}}|\mathbf{x}_h)P(\mathbf{x}_h)}{\sum_{h=1}^{\mathcal{H}}p(\hat{\mathbf{y}}|\mathbf{x}_h)P(\mathbf{x}_h)},$$

and  $\mathbb{E}_{\mathbf{x}|\hat{\mathbf{y}}}\left[\widehat{\mathbf{B}}\mathbf{x}\right]$  can be similarly calculated.

We can simplify (C.30) by inspecting the block matrix structures of  $\hat{\mathbf{y}} = [\mathbf{y}_1^T, \mathbf{y}_2^T, \dots, \mathbf{y}_{N_{\text{Rx}}}^T]^T$ ,  $\widehat{\mathbf{B}}\mathbf{x}$ , and  $\{\widehat{\mathbf{G}}\mathbf{x}\}$ . We then, obtain the following result similar with the result of monostatic radar.

$$\begin{aligned} & \frac{\partial TSI}{\partial \widetilde{\mathbf{W}}} \\ &= \frac{\mathbf{W}}{\sigma^2} \mathbb{E}_{\mathbf{x}, \mathbf{y}_m} \left[ \sum_{m=1}^{N_{\text{Rx}}} \left\{ \text{diag}\{\mathbf{G}_m \mathbf{x}\} \cdot \{\mathbf{G}_m \mathbf{x}\} - \mathbb{E}_{\mathbf{x}|\mathbf{y}_m} [\text{diag}\{\mathbf{G}_m \mathbf{x}\}] \cdot \mathbb{E}_{\mathbf{x}|\mathbf{y}_m} [\mathbf{G}_m \mathbf{x}] \right\} \right] \end{aligned} \quad (\text{C.31})$$

where  $\frac{\partial TSI}{\partial \widetilde{\mathbf{W}}} = \left[ \frac{\partial TSI}{\partial w(1)}, \frac{\partial TSI}{\partial w(2)}, \dots, \frac{\partial TSI}{\partial w(\mathcal{L})} \right]^T$ ,  $\mathbb{E}_{\mathbf{x}|\mathbf{y}_m} [\mathbf{G}_m \mathbf{x}]$  is calculated by

$$\mathbb{E}_{\mathbf{x}|\mathbf{y}_m} [\mathbf{G}_m \mathbf{x}] = \frac{\sum_{h=1}^{\mathcal{H}}\{\mathbf{G}_m \mathbf{x}_h\}p(\mathbf{y}_m|\mathbf{x}_h)P(\mathbf{x}_h)}{\sum_{h=1}^{\mathcal{H}}p(\mathbf{y}_m|\mathbf{x}_h)P(\mathbf{x}_h)},$$

and  $\mathbb{E}_{\mathbf{x}|\mathbf{y}_m} [\text{diag}\{\mathbf{G}_m \mathbf{x}\}]$  is similarly calculated.

For numerical programming, (C.31) can be written as follows.

$$\begin{aligned} & \frac{\partial TSI}{\partial \widetilde{\mathbf{W}}} \\ &= \frac{\mathbf{W}}{N_{\text{total}}\sigma^2} \sum_{\mathbf{x}_0, \mathbf{y}_{m_0}} \left\{ \sum_{m=1}^{N_{\text{Rx}}} \left\{ \text{diag}\{\mathbf{G}_m \mathbf{x}_0\} \cdot \{\mathbf{G}_m \mathbf{x}_0\} - \mathbb{E}_{\mathbf{x}|\mathbf{y}_{m_0}} [\text{diag}\{\mathbf{G}_m \mathbf{x}\}] \cdot \mathbb{E}_{\mathbf{x}|\mathbf{y}_{m_0}} [\mathbf{G}_m \mathbf{x}] \right\} \right\} \end{aligned} \quad (\text{C.32})$$

where  $\hat{\mathbf{y}}_0 = [\mathbf{y}_{1_0}, \mathbf{y}_{2_0}, \dots, \mathbf{y}_{0N_{\text{Rx}}}]^T$  and

$$\mathbb{E}_{\mathbf{x}|\mathbf{y}_{m_0}} [\mathbf{G}_m \mathbf{x}] = \frac{\sum_{h=1}^{\mathcal{H}} \{\mathbf{G}_m \mathbf{x}_h\} p(\mathbf{y}_{m_0}|\mathbf{x}_h) P(\mathbf{x}_h)}{\sum_{h=1}^{\mathcal{H}} p(\mathbf{y}_{m_0}|\mathbf{x}_h) P(\mathbf{x}_h)}.$$

## REFERENCES

- [1] 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, September 2008.
- [2] M.R. Bell, “Information theory and radar waveform design,” *Information Theory, IEEE Transactions on*, vol. 39, no. 5, pp. 1578–1597, Sep 1993.
- [3] S.U. Pillai, H.S. Oh, D.C. Youla, and J.R. Guerci, “Optimal transmit-receiver design in the presence of signal-dependent interference and channel noise,” *Information Theory, IEEE Transactions on*, vol. 46, no. 2, pp. 577 –584, mar 2000.
- [4] J.R. Guerci and S.U. Pillai, “Theory and application of optimum transmit-receive radar,” in *Radar Conference, 2000. The Record of the IEEE 2000 International*, 2000, pp. 705 –710.
- [5] D.A. Garren, M.K. Osborn, A.C. Odom, J.S. Goldstein, S.U. Pillai, and J.R. Guerci, “Enhanced target detection and identification via optimised radar transmission pulse shape,” *Radar, Sonar and Navigation, IEE Proceedings -*, vol. 148, no. 3, pp. 130 –138, jun 2001.
- [6] D.A. Garren, A.C. Odom, M.K. Osborn, J.S. Goldstein, S.U. Pillai, and J.R. Guerci, “Full-polarization matched-illumination for target detection and identification,” *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 38, no. 3, pp. 824 – 837, jul 2002.

- [7] A. Leshem, O. Naparstek, and A. Nehorai, “Information theoretic adaptive radar waveform design for multiple extended targets,” *Selected Topics in Signal Processing, IEEE Journal of*, vol. 1, no. 1, pp. 42–55, June 2007.
- [8] N.A. Goodman, P.R. Venkata, and M.A. Neifeld, “Adaptive waveform design and sequential hypothesis testing for target recognition with active sensors,” *Selected Topics in Signal Processing, IEEE Journal of*, vol. 1, no. 1, pp. 105–113, June 2007.
- [9] R.A. Romero and N.A. Goodman, “Waveform design in signal-dependent interference and application to target recognition with multiple transmissions,” *Radar, Sonar Navigation, IET*, vol. 3, no. 4, pp. 328–340, August 2009.
- [10] J.-J. Xiao, S. Cui, Z.-Q. Luo, and A. J. Goldsmith, “Power scheduling of universal decentralized estimation in sensor networks,” *IEEE Transactions on Signal Processing*, vol. 54, no. 2, pp. 413–422, February 2006.
- [11] D.P. Palomar and S. Verdu, “Gradient of mutual information in linear vector gaussian channels,” *Information Theory, IEEE Transactions on*, vol. 52, no. 1, pp. 141–154, Jan. 2006.
- [12] Daniel P. Palomar and Sergio Verdu, “Representation of mutual information via input estimates,” *Information Theory, IEEE Transactions on*, vol. 53, no. 2, pp. 453–470, Feb. 2007.
- [13] R. R. Tenney and N. R. Sandell Jr., “Detection with distributed sensors,” *IEEE Transactions on Aerospace Electronic Systems*, vol. AES-17, no. 4, pp. 501–510, July 1981.

- [14] A. R. Reibman and L. W. Nolte, "Optimal detection and performance of distributed sensor systems," *IEEE Transactions on Aerospace Electronic Systems*, vol. AES-23, no. 1, pp. 24–30, January 1987.
- [15] I. Y. Hoballah and P. K. Varshney, "Distributed bayesian signal detection," *IEEE Transactions on Information Theory*, vol. 35, no. 6, pp. 995–1000, September 1989.
- [16] R. S. Blum, S. A. Kassam, and H. V. Poor, "Distributed detection with multiple sensors: Part II - advanced topics," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 64–79, January 1997.
- [17] T. M. Duman and M. Salehi, "Decentralized detection over multiple-access channels," *IEEE Transactions on Aerospace Electronic Systems*, vol. 34, no. 2, pp. 469–476, April 1998.
- [18] Chen and P. Willett, "On the optimality of the likelihood-ratio test for local sensor decision rules in the presence of nonideal channels," *IEEE Transactions on Information Theory*, vol. 51, no. 2, pp. 693–699, February 2005.
- [19] R. Niu, B. Chen, and P. 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.
- [20] J. Chamberland and V. V. Veeravalli, "Asymptotic results for decentralized detection in power constrained wireless sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 6, pp. 1007–1015, August 2004.
- [21] J.-J. Xiao, S. Cui, Z.-Q. Luo, and A. J. Goldsmith, "Power-efficient analog forwarding transmission in an inhomogeneous gaussian sensor network," in *IEEE Workshop*

- on Signal Processing Advances in Wireless Communications*, New York, NY, January 2005, pp. 121–125.
- [22] X. Zhang, H. V. Poor, and M. Chiang, “Power allocation in distributed detection with wireless sensor networks,” in *Proc. 44th Allerton Conference on Communication, Control, and Computing*, September 2006, Invited Paper.
- [23] S.C.A. Thomopoulos, R. Viswanathan, and D. K. Bougoulas, “Optimal distributed decision fusion,” *IEEE Transactions on Aerospace Electronic Systems*, vol. IT-25, no. 5, pp. 761–765, September 1989.
- [24] S.C.A. Thomopoulos and L. Zhang, “Distributed decision fusion with networking delays and channel errors,” *Information Science*, vol. 66, pp. 91–118, December 1992.
- [25] F.A. Sadjadi, “Hypotheses testing in a distributed environment,” *Aerospace and Electronic Systems, IEEE Transactions on*, vol. AES-22, no. 2, pp. 134–137, March 1986.
- [26] J. J. Chao, E. Drakopoulos, and C. C. Lee, “An evidential reasoning approach to distributed multiple-hypothesis detection,” in *Decision and Control, 1987. 26th IEEE Conference on*, Dec. 1987, vol. 26, pp. 1826–1831.
- [27] X.G. Wang and H.C. Shen, “Multiple hypotheses testing strategy for distributed multisensor systems,” in *Intelligent Robots and Systems, 2000. (IROS 2000). Proceedings. 2000 IEEE/RSJ International Conference on*, 2000, vol. 2, pp. 1440–1445 vol.2.
- [28] X.G. Wang, M. Moallem, and R.V. Patel, “Distributed multiple hypotheses testing with serial distributed decision fusion,” in *Computational Intelligence in Robotics*

- and Automation, 2001. Proceedings 2001 IEEE International Symposium on, 2001, pp. 549–554.*
- [29] S. Oh, S. Sastry, and L. Schenato, “A hierarchical multiple-target tracking algorithm for sensor networks,” in *Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on, April 2005*, pp. 2197–2202.
- [30] S.M. Sowelam and A.H. Tewfik, “Waveform selection in radar target classification,” *Information Theory, IEEE Transactions on*, vol. 46, no. 3, pp. 1014 –1029, may 2000.
- [31] Y. Yang and R.S. Blum, “Mimo radar waveform design based on mutual information and minimum mean-square error estimation,” *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 43, no. 1, pp. 330 –343, january 2007.
- [32] S. Kay, “Waveform design for multistatic radar detection,” *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 45, no. 3, pp. 1153 –1166, july 2009.
- [33] S. Kay, “Optimal signal design for detection of gaussian point targets in stationary gaussian clutter/reverberation,” *Selected Topics in Signal Processing, IEEE Journal of*, vol. 1, no. 1, pp. 31 –41, june 2007.
- [34] Mark A. Neifeld, Amit Ashok, and Pawan K. Baheti, “Task-specific information for imaging system analysis,” *Journal of the Optical Society of America*, vol. 24, no. 12, pp. B25–B164, 2007.
- [35] Pawan K. Baheti and Mark A. Neifeld, “Recognition using information-optimal adaptive feature-specific imaging,” *Journal of the Optical Society of America*, vol. 26, pp. 1055–1070, 2009.

- [36] T. Cover and J. Thomas, *Elements of Information Theory*, John Wiley & Sons, 2nd edition, 2006.
- [37] H. Kobayashi and J. B. Thomas, “Distance measures and related criteria,” in *in Proceedings of the Fifth Annual Allerton Conference Circuit and System Theory*, October 1967, pp. 491–500.
- [38] H. Kobayashi, “Distance measures and asymptotic relative efficiency,” *IEEE Transactions on Information Theory*, vol. 16, no. 3, pp. 288–291, May 1970.
- [39] T. Kailath, “The divergence and bhattacharyya distance measures in signal selection,” *IEEE Transactions on Communication Technology*, vol. 15, no. 2, pp. 52–60, February 1967.
- [40] S. M. Ali and S. D. Silvey, “A general class of coefficients of divergence of one distribution from another,” in *Journal of the Royal Statistical Society*, 1966, vol. 28 of *B*, pp. 131–142.
- [41] Shuguang Cui, Jin-Jun Xiao, A.J. Goldsmith, Zhi-Quan Luo, and H.V. Poor, “Estimation diversity and energy efficiency in distributed sensing,” *Signal Processing, IEEE Transactions on*, vol. 55, no. 9, pp. 4683–4695, sept. 2007.
- [42] H. Kim, J. Wang, P. Cai, and S. Cui, “Detection outage and detection diversity in a homogeneous distributed sensor network,” *IEEE Transactions on Signal Processing*, vol. 57, no. 7, pp. 2875–2881, July 2009.
- [43] E. Mosca, “Probing signal design for linear channel identification,” *IEEE Transactions on Information Theory*, vol. IT-18, no. 4, pp. 481–487, July 1972.
- [44] T. L. Grettenberg, “Signal selection in communication and radar systems,” *IEEE Transactions on Information Theory*, vol. IT-9, pp. 265–275, October 1963.

- [45] A. Weiss, “An introduction to large deviations for communication networks,” *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 6, pp. 938–952, August 1995.
- [46] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, Jones and Bartlett, Boston, 1993.
- [47] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, 2nd edition, Sep 1999.
- [48] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Mar 2004.
- [49] M.F. Huber, T. Bailey, H. Durrant-Whyte, and U.D. Hanebeck, “On entropy approximation for gaussian mixture random vectors,” in *Multisensor Fusion and Integration for Intelligent Systems, 2008. MFI 2008. IEEE International Conference on*, 20-22 2008, pp. 181 –188.
- [50] J.R. Hershey and P.A. Olsen, “Approximating the kullback leibler divergence between gaussian mixture models,” in *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, April 2007, vol. 4, pp. IV–317 – IV–320.
- [51] Y. Yang and R.S. Blum, “Mimo radar waveform design based on mutual information and minimum mean-square error estimation,” *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 43, no. 1, pp. 330 –343, january 2007.
- [52] T.B. Butler and N.A. Goodman, “Multistatic target classification with adaptive waveforms,” in *Radar Conference, 2008. RADAR '08. IEEE*, 26-30 2008, pp. 1 –6.

- [53] Jr. Campbell, J.P., “Speaker recognition: a tutorial,” *Proceedings of the IEEE*, vol. 85, no. 9, pp. 1437–1462, sep 1997.