

# **MEDIUM-AWARE INFERENCE FOR WIRELESS SENSOR NETWORKS**

**A Thesis Submitted to  
the Graduate School of  
İzmir Institute of Technology  
in Partial Fulfillment of the Requirements for the Degree of**

**DOCTOR OF PHILOSOPHY**

**in Electrical and Electronics Engineering**

**by  
Muath Abed Alrauf WAHDAN**

**December 2020  
İZMİR**

## ACKNOWLEDGMENTS

All praises to Allah and His blessing for the completion of this thesis. I thank God for all the opportunities, trials and strength that he gave me to finish writing this thesis. My humblest gratitude to the Prophet Muhammad whose way of life has been continuous guidance for me.

First and foremost, I would like to thank sincerely my supervisor Assoc. Prof. Dr. Mustafa A. Altinkaya, for his guidance, patience and most importantly answering all my questions and giving me excellent advice, which helped me in achieving my goals. It has been a great pleasure and honor to have him as my supervisor.

I wish to express my special thanks to my thesis progress committee members Prof. Dr. Olcay Akay and Assoc. Prof. Dr. Barış Atakan for their excellent support and guidance which helped me get results of better quality. I am also grateful to the committee members Prof. Dr. Aydın Akan and Assoc. Prof. Dr. Berna Özbek for their insightful comments and encouragement.

I would also like to thank Türkiye Bursları (YTB) for their support with the scholarship they provided me during part of my Ph.D. studies.

I also want to extend my thanks to Dr. Ahmed K. Sultan Salem, for the useful course he provided for free on YouTube channel, namely "Detection theory".

I would sincerely like to thank all my beloved friends who were with me and support me through thick and thin.

Last but not least, I want to send my warm sincere to all my family members, for their support and encouragement. I would like to thank my dearest mother Khawla, my father Abed Alrauf, my sisters; Israa, Zainab, Shayma and Maymona, my brother Baraa.

# ABSTRACT

## MEDIUM-AWARE INFERENCE FOR WIRELESS SENSOR NETWORKS

In wireless sensor networks, multilevel quantization is necessary in order to find a compromise between small power consumption of sensors and good detection performance at the fusion center (FC) and generally, distance measures like J-divergence (JD) are used for quantization. This thesis proposes an approach based on maximizing the average output entropy of the sensors under both hypotheses named as maximum average entropy (MAE) method which is used in a Neyman-Pearson criterion based distributed detection scheme in order to detect a point source.

Firstly, a deterministic signal with isotropic propagation model was considered. The receiver operating characteristics with multilevel MAE quantization of sensor outputs was evaluated both when the sensor outputs are available error-free at the FC and when they are transmitted using non-coherent communication via Rayleigh fading channels. Also sequential probability ratio tests of Wald were performed. Then, the case of spatially correlated sensors with a Gaussian isotropic event source was investigated. The computational requirements in evaluating multidimensional cumulative densities necessitated a rectangular grid model of sensor deployment and block-diagonal approximations of covariance matrix related to the event signal at the sensors without losing generality.

The simulation studies show the success of MAE in the deterministic signal model and at six-level quantization its performance approaches that of non-quantized data transmission. In the sequential tests, again MAE was more successful compared to MJD resulting in smaller average sample numbers. It was observed that spatial correlation degrades system performance and MJD performs better in hypothesis tests based on change in variance.

# ÖZET

## TELSİZ ALGIÇ AĞLARI İÇİN ORTAM GÖZETEREK ÇIKARSAMA

Telsiz algıç ağlarında algıçların düşük güç harcaması ile füzyon merkezindeki iyi sezim başarımı arasında bir ödünleşme bulmak için çok seviyeli nicemleme gereklidir ve genellikle J-iraksaması gibi uzaklık ölçütleri nicemlemede kullanılmaktadır. Bu tez ikili hipotez testindeki her bir hipotezin geçerli olduğu durumda algıçların çıkış entropisini enbüyük yapmaya dayalı, enbüyük ortalama entropi (MAE) yöntemi olarak adlandırılan ve Neyman-Pearson kriterine dayalı bir dağıtık sezim projesinde noktasal bir kaynağı sezme için kullanılan bir yaklaşım önermektedir.

İlk olarak, yönbağımsız yayılım modeli ile gerekirci bir sinyal ele alınmıştır. Algıç çıktılarının MAE nicemlemesi ile alıcı çalışma karakteristiği, algıç çıktılarının hatasız olarak füzyon merkezinde bulunduğu ve evre-uyumsuz iletişim kullanılarak Rayleigh sönümlenmeli kanal üzerinden yollandığı her iki durum için elde edilmiştir. Wald'ın sıralı olasılık oranı testi de uygulanmıştır. Sonra Gauss bir yönbağımsız olay kaynağı ile uzamsal ilintili algıçların olduğu durum incelenmiştir. Çok boyutlu birikimli yoğunlukların hesaplanmasındaki işlemsel gereksinimler, algıç yayılımı için dikedörtgensel bir ağ modeli ve genel geçerlilik kaybedilmeksizin algıçlardaki olay sinyaliyle ilgili kovaryans matrisinin blok-köşegenel yaklaşıklıklarını gerektirmiştir.

Benzetim çalışmaları MAE'nin gerekirci sinyal modelindeki başarısını göstermektedir ve altı-seviyeli nicemleme ile başarımı nicemlenmemiş veri iletimininkine yaklaşmaktadır. Sıralı testlerde yine MAE, MJD'ye kıyasla daha başarılı olmuş ve daha küçük ortalama örnek oranı sonucuna ulaşmıştır. Uzamsal ilintinin sistem başarımını kötüleştirdiği ve MJD'nin varyansta değişime dayalı hipotez testlerinde daha başarılı olduğu gözlemlenmiştir.

# TABLE OF CONTENTS

LIST OF FIGURES .....	viii
LIST OF TABLES .....	xi
LIST OF SYMBOLS .....	xii
ABBREVIATIONS .....	xiv
CHAPTER 1. INTRODUCTION .....	1
1.1. Motivation .....	1
1.2. Objectives of the Thesis .....	3
1.3. Organization of the Thesis .....	3
CHAPTER 2. PRELIMINARIES .....	5
2.1. Distributed Detection .....	5
2.1.1. Distributed Neyman-Pearson Detection.....	5
2.1.2. Distributed Bayesian Detection.....	8
2.1.2.1. Derivation of the Bayesian Local Decision Rules when All Decision Rules are Inter-Dependent .....	9
2.1.2.2. Conditionally Independent Local Observations .....	12
2.1.2.3. Fusion Rule at the FC .....	12
2.1.2.4. Incorporation of Binary Symmetric Channels Between Sensors and FC .....	13
2.1.2.5. Identical Local Detectors .....	16
2.2. Sequential Detection .....	19
2.2.1. Bayesian Sequential Test .....	19
2.2.2. Sequential Probability Ratio Test.....	20
2.3. Composite Hypothesis Testing .....	23
2.3.1. Bayesian Approach.....	25

2.3.2. Generalized Likelihood Ratio Test .....	28
2.4. Energy Detector .....	28
CHAPTER 3. MAXIMUM AVERAGE ENTROPY-BASED QUANTIZATION OF LOCAL OBSERVATIONS FOR DISTRIBUTED DETECTION.....	32
3.1. Introduction.....	32
3.2. System Model .....	33
3.2.1. Point Source .....	34
3.2.2. Fusion System: Channel Between Sensors and FC.....	36
3.2.2.1. Fading Channel .....	36
3.3. Quantizer Design .....	40
3.3.1. Maximum Average Entropy Method .....	41
3.3.2. Maximum J-Divergence Method .....	43
3.3.3. Relation of MAE and MJD Methods .....	45
3.4. Simulation Results .....	47
3.4.1. Binary Direct Data Transmission.....	48
3.4.2. Performance of MAE and MJD with Multilevel Quantization..	48
3.4.3. Multiple Level Data Transmission over Rayleigh Fading Chan- nel .....	51
3.5. Conclusions on MAE Based Quantization Method.....	52
CHAPTER 4. SEQUENTIAL DETECTION FOR MULTIPLE LOCAL SENSORS	55
4.1. Introduction.....	55
4.2. Centralized SPRT for Multiple Local Detectors .....	59
4.3. Distributed SPRT for Multiple Local Detectors .....	61
4.4. Simulations for Sequential Probability Ratio Test.....	63
4.4.1. The Relation Between ASN and the Required $p_d$ for MAE Method .....	63
4.4.2. Comparison Between SPRT and FSS .....	64
4.4.3. Comparison Between MAE and MJD Methods .....	65
4.4.4. SPRT for Multiple Level Data Transmission over Rayleigh Fading Channel .....	66

4.5. Conclusions on Sequential Detection .....	67
CHAPTER 5. SPATIAL CORRELATION MODELLING FOR SENSOR OBSER-	
VATIONS .....	75
5.1. Introduction.....	75
5.2. System Model .....	76
5.3. Sensor Deployment Model.....	79
5.4. Quantizer Design .....	80
5.5. Simulations for the Spatial Correlation Model.....	82
5.6. Conclusions on Spatially Correlated Observations .....	84
CHAPTER 6. CONCLUSIONS AND SUGGESTIONS FOR FUTURE RESEARCH	91
6.1. Conclusions.....	91
6.2. Suggestions for Future Research .....	92
REFERENCES .....	94
APPENDIX A. DERIVATION OF THE LOCAL DECISION RULE .....	101
APPENDIX B. DESIGN EXAMPLE OF A BAYESIAN PARALLEL FUSION	
NETWORK WITH TWO SENSORS .....	104
APPENDIX C. TOY EXAMPLES FOR APPLYING MAE AND MJD .....	112

# LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
Figure 2.1. Distributed detection with two sensors. ....	6
Figure 2.2. Probability that the LR is greater than the threshold, $\eta$ , under $H_0$ in the DD problem consisting of 2 sensors. ....	8
Figure 2.3. ROC curve for DD problem consisting of 2 sensors. ....	9
Figure 2.4. Parallel topology of DD with fusion center. ....	10
Figure 2.5. DD problem with BSC ....	14
Figure 2.6. Binary symmetric channel ....	15
Figure 2.7. The thresholds for the Bayesian sequential test with $c=0.03$ . ....	21
Figure 2.8. The thresholds for the Bayesian sequential test with $c=0.3$ . ....	22
Figure 2.9. ROC curve for the energy detector ....	31
Figure 3.1. Parallel distributed detection system ....	34
Figure 3.2. Positions of the event location and uniformly distributed sensors in a scenario for detecting a point source. ....	35
Figure 3.3. The pdf of the signal amplitude observed at the sensors, $p(A_n)$ . ....	37
Figure 3.4. A partitioning of the pdf for the observations at each sensor for 4-level quantization. ....	40
Figure 3.5. The entropy functions $\hat{F}_{H_0}$ , $\hat{F}_{H_1}$ and $\hat{F}_{av}$ for binary quantization. ....	43
Figure 3.6. The entropy functions $\hat{F}_{H_0}$ , $\hat{F}_{H_1}$ and $\hat{F}_{av}$ for three level quantization. ...	44
Figure 3.7. The JD for binary quantization. ....	46
Figure 3.8. The J-divergence for three level quantization. ....	47
Figure 3.9. Comparison between the ROC curves obtained using MAE, MJD and Kth root methods for binary and Gaussian DDT. ....	49
Figure 3.10. Comparison between the ROC curves obtained using MAE and MJD methods for three-level and Gaussian DDT. ....	50
Figure 3.11. Comparison between the ROC curves obtained using MAE method for six-level and Gaussian DDT. ....	51
Figure 3.12. ROC curves in the case of fading channel by using MAE based quantization and optimum fusion rule. ....	53



<u>Figure</u>	<u>Page</u>
Figure 3.13. A comparison between the ROCs of the optimal and sub-optimal fusion rule for binary and six level quantizations and the corresponding non quantized data transmissions in the case of fading channel. ....	54
Figure 4.1. Ratio of the ASN in sequential detection and the FSS in block detection versus $p_{FA}$ . ....	59
Figure 4.2. Simulated $ASN_m$ versus logarithm of false alarm or miss probability for different levels of MAE or MJD quantization and DDT. ....	67
Figure 4.3. Theoretical $ASN_m$ versus logarithm of false alarm or miss probability for different levels of MAE or MJD quantization and DDT. ....	68
Figure 4.4. Theoretical and simulated $ASN_m$ versus logarithm of false alarm or miss probability for different levels of MAE or MJD quantization and DDT. ....	69
Figure 4.5. Simulated $ASN_m$ versus logarithm of false alarm or miss probability for different levels of MAE quantization and non-coherent M-FSK communication over fading channels. ....	71
Figure 4.6. Simulated $ASN_m$ versus logarithm of false alarm or miss probability for different levels of MJD quantization and non-coherent M-FSK communication over fading channels. ....	72
Figure 4.7. Comparison of DDT and non-coherent M-FSK communication over fading channel cases using simulated $ASN_m$ versus logarithm of false alarm or miss probability for different levels of MAE quantization. ....	73
Figure 5.1. The model architecture. ....	77
Figure 5.2. Pdfs of the observations when there is a target at $T_3$ position and binary decision regions for hypothesis test based on change in variance ....	81
Figure 5.3. The entropy functions $\hat{F}_{H_0}$ , $\hat{F}_{H_1}$ and $\hat{F}_{av}$ for binary quantization of spatially correlated signals. ....	82
Figure 5.4. The J-divergence for binary quantization of spatially correlated signals. ....	83
Figure 5.5. Covariance matrix $\times 10^4$ of sensor observations for the grid sensor network with event location $T_3$ as shown in Figure 5.7. ....	85
Figure 5.6. Visualization for the covariance matrix shown in Figure 5.5 ....	86
Figure 5.7. Spatially correlated sensors in the sensitivity circle of target $T_3$ ....	87

<u>Figure</u>	<u>Page</u>
Figure 5.8. ROC curve for the independent Gaussian signals. ....	88
Figure 5.9. ROC curve for the correlated Gaussian signals. ....	89
Figure 5.10. Comparison between i.i.d. and correlated observations for MJD method ( $\theta_1 = 21.83$ SNR = 12, dB (w.r.t. a sensor at 5 unit distance)). ....	90

## LIST OF TABLES

<u>Table</u>	<u>Page</u>
Table 2.1. The possible outputs for the two sensors and their corresponding probabilities. ....	9
Table 2.2. Monotonic global fusion rules at the FC according to binary decisions from two sensors. ....	10
Table 3.1. The relation between $p_D$ and $p_{FA}$ for different levels of quantization obtained with MAE and MJD methods. ....	50
Table 3.2. Achieved gain in $p_D$ by using the MAE method in quantization instead of MJD. ....	52
Table 4.1. The relation between the ASN and probability of detection for the different quantization level at $p_{fa} = 0.1$ and SNR= 0 dB. ....	64
Table 4.2. The relation between $p_d$ and $p_{fa}$ for different levels of quantization obtained with MAE and MJD methods. ....	65
Table 4.3. The ASN using MAE quantization method for different types of data transmissions. ....	70
Table 4.4. The ASN using MJD quantization method for different types of data transmissions. ....	70
Table 4.5. Achieved gain in $ASN_m$ using the MAE method in quantization w.r.t MJD for DDT. ....	71
Table 4.6. Achieved gain in $ASN_m$ by using the MAE method in quantization w.r.t MJD in the case of fading channel. ....	72

## LIST OF SYMBOLS

$\log(\cdot)$	Natural Logarithm
$\Lambda(\cdot)$	Likelihood Ratio
$A$	Signal Amplitude
$D_{\text{KL}}$	Kullback–Leibler Divergence
$H_i$	Hypothesis $i$
$Q_{\text{func}}(\cdot)$	Complimentary Cumulative Density Function for Standard Gaussian Random Variable
$\Gamma_{\text{inc}}(\cdot)$	Incomplete Gamma Function
$\mathfrak{R}$	Bayesian Risk
$\beta_M$	Vector of Local Thresholds for $M$ -Level Quantization
$\epsilon_k$	Additive White Gaussian Noise at the $k^{\text{th}}$ Sensor
$\eta$	Global Threshold
$\eta_0$	Lower Threshold for SPRT
$\eta_1$	Upper Threshold for SPRT
$\hat{F}$	Entropy Function
$\mathbf{Y}$	Matrix
$\sigma_N^2$	Variance of the AWGN at Sensor Observations
$\sigma_n^2$	Variance of the AWGN at Fading Channels
$\sigma_s^2$	Signal Variance
$\sim$	Distributed According to
$\mathbf{I}$	Identity Matrix
$c_{ij}$	Cost to Decide $H_i$ when $H_j$ is True

$h_k$	Complex Channel Coefficient at the $k^{\text{th}}$ Sensor
$n_k$	The Number of Ones at the $k^{\text{th}}$ Sensor
$p(H_j)$	Prior Probability for Hypothesis $j$
$p_{d_i}$	Local Probability of Detection at Sensor $i$
$p_{fa_i}$	Local Probability of False Alarm at Sensor $i$
$E(\cdot)$	Expectation Operation
$\mu$	Mean
$u_0$	Global Binary Decision at Fusion Center
$u_i$	Local Binary Decision at Sensor $i$
$y_i$	Observation at Sensor $i$
$K$	Number of Sensors
$\bar{\pi}$	Upper Threshold for Bayesian Sequential Test
$\underline{\pi}$	Lower Threshold for Bayesian Sequential Test

## ABBREVIATIONS

ASN	Average Sample Number
AWGN	Additive White Gaussian Noise
BSC	Binary Symmetric Channel
CI	Chernoff Information
CRLB	Cramer-Rao Lower Bound
DC	Deflection Criterion
DD	Distributed Detection
DDT	Direct Data Transmission
FC	Fusion Center
FSS	Fixed Sample Size
G-LOD	Generalized-Locally Optimum Detector
GLRT	Generalized Likelihood Ratio Test
i.i.d.	independently and identically distributed
IoT	Internet of Things
JD	J-Divergence
LR	Likelihood Ratio
LRT	Likelihood Ratio Test
MAE	Maximum Average Entropy
MEMS	Micro-Electro-Mechanical Systems
M-FSK	M-ary Frequency Shift Keying
MJD	Maximum JD
ML	Maximum Likelihood

MLEs	ML Estimates
MOE	Maximum Output Entropy
MVNCD	Multivariate Normal Cumulative Density
NP	Neyman-Pearson
OC	Operating Characteristic
OVBS	One Variate Bivariate Screening
pdf	probability density function
pmf	probability mass function
ROC	Receiver Operating Characteristics
SNR	Signal-to-Noise Ratio
SP	Saving Percentage
SPRT	Sequential Probability Ratio Rest
TVBS	Two Variate Bivariate Screening
UMP	Uniformly Most Powerful
w.p.	with probability
w.r.t.	with respect to
WSN	Wireless Sensor Network

# CHAPTER 1

## INTRODUCTION

### 1.1. Motivation

Wireless sensor networks (WSNs) have come into the spotlight recently due to a major development in the Micro-Electro-Mechanical Systems (MEMS) (Hill et al., 2004; Ragam and Sahebraoji, 2019). The recent development of WSNs has made this field a research focus on intensive researches. The researchers are widely using it in monitoring and characterizing large physical environments and for tracing various environmental or physical conditions such as temperature, pressure, wind, and humidity. Apart from these, WSNs have vast fields to be applied in, such as harmful environmental exploration, wildlife monitoring, target tracking and smart cities established based on Internet of Things (IoT) (Rajput and Kumaravelu, 2019; Muduli et al., 2018; Wahdan et al., 2015; Zanella et al., 2014). Typically a WSN uses a huge number of comparatively inexpensive and low-energy sensors to collect observations and pre-process the observations. These sensors are generally deployed in the environment. Owing to strict energy and bandwidth restrictions, observations of the sensors are frequently needed to be quantized before transmitting them to a fusion center (FC) where a global decision is made (Al-Jarrah et al., 2018; Ciunozzo et al., 2012). A very important aspect of the detection procedures in WSNs that discriminates them from classical detection problems is their distributive nature which necessitates solving additional coupled optimization problems.

The pioneering research on distributed detection (DD) was made by Tenney and Sandell (Tenney and Sandell, 1981) and Chair and Varshney (Chair and Varshney, 1986a). In (Tenney and Sandell, 1981), a detection problem consisting of two sensors and one FC with a fixed fusion rule was considered to show that the optimum local decision rule is the likelihood ratio test (LRT) under the Bayesian criterion. However, individual thresholds are coupled. Later, in (Chair and Varshney, 1986a), it was shown that the optimum fusion rule at the FC for multiple observations is also an LRT both under the Neyman-



Pearson (NP) and the Bayesian criteria. Determining local decision rules is significantly more complicated. The optimality of LRT for each local decision rule was considered in (Tsitsiklis, 1993) and (Viswanathan and Varshney, 1997), by assuming conditional independence of the observations under each hypothesis. But because of the coupling between the LRT thresholds at the local detectors among themselves and with the one at the FC solving the global optimization problem is mathematically complex though not intractable (Tsitsiklis and Athans, 1985). This complexity suggested determining the thresholds of the local detectors independently, that is, the threshold of each sensor is optimized for fixed decision rules at the other detectors and the FC (Veeravalli and Varshney, 2012). The adopted conditional independence assumption in those works produces only locally optimal decisions, but even they become prohibitively complex for large sensor networks and simpler solutions are needed. Additionally, the gain obtained by having more sensor nodes outperforms the gain of getting more information from each sensor in WSNs (Veeravalli and Varshney, 2012).

Most of the previously mentioned WSN applications considered the field source where Signal-to-Noise Ratio (SNR) is known in which the probability of detection and probability of false alarm, is known to the FC. On the other hand, in many scenarios, the point source has been considered where the emitted signal is assumed to be unknown and decays as a function of the distance from the target. These works deal with targets that emit an isotropic signal (Niu and Varshney, 2007; Niu and Varshney, 2006; Ciuonzo and Rossi, 2017). In (Niu and Varshney, 2007), the detection performances, namely either probability of false alarm and probability of detection have been derived by several methods in order to simplify its computations. Differently in (Niu and Varshney, 2006) a generalized LRT (GLRT) detector and the Cramer-Rao Lower Bound (CRLB) are derived; however, GLRT needs a grid search on both the emitted power from the target and the location domains. Therefore, simpler solutions are needed. In (Ciuonzo and Rossi, 2017), one-bit DD of an uncooperative target is considered which assumes that both the target location and the emitted signal are unknown at FC. In that work, a generalized- locally optimum detector (G-LOD) test with nuisance parameters has been proposed for a fixed value of local sensor thresholds, moreover different types of fusion rules are developed according to the detection framework. However, these works have considered binary data transmissions and the effect of Rayleigh fading was not considered.

Traditional DD systems were developed by assuming an error-free communication between the local detectors and the FC (Viswanathan and Varshney, 1997). Applying this theory to WSNs leads to a detection performance loss in case of erroneous channels. Fusion rules for DD systems considering fading channels were first discussed in (Chen et al., 2002) and later in (Chen et al., 2004; Chen and Willett, 2005; Niu et al., 2006) mainly for a binary hypothesis testing whereas M-ary hypothesis testing was considered in (Kotecha et al., 2005; Liu et al., 2011; Maleki and Vosoughi, 2012; Hajibabaei and Vosoughi, 2014) for binary data transmissions. Many times the sensor observations are spatially correlated, however, the works considering local and global decision rules have not considered this until now, to the best of our knowledge.

## **1.2. Objectives of the Thesis**

In this thesis, we will propose a novel method for quantizing the sensor outputs in order to suggest a compromise between the low power consumption of the wireless sensors and better global detection performance. Specifically, we aim

- to give new insights for DD of point sources using WSNs,
- to propose information theoretic method(s) in quantization of intermediate results (local observations or logarithms of the corresponding local likelihood ratios),
- to apply this in both of the cases when the sensor outputs are available error-free at the FC or sent over a wireless channel,
- comparison with rival information theoretic methods,
- applying the proposed information theoretic method(s) for sequential detection of point sources in WSN,
- accounting for sensor correlations, i.e., introducing location-awareness to the developed method.

### **1.3. Organization of the Thesis**

We start our discussion in Chapter 2 by introducing preliminaries of the DD, sequential detection, composite hypothesis testing and energy detector.

In Chapter 3, we formulate the parallel DD problem of a point source including sensor to FC transmission over a Rayleigh fading channel and various fusion rules. Also, the proposed average entropy based quantization method, the Jd based method and their relation are considered.

In Chapter 4, we have generalized the classical sequential probability ratio test (SPRT) to different types of data transmissions, using MAE and MJD quantization method.

In Chapter 5, a spatial correlation model for the sensor observations is proposed using a stochastic signal propagation.

We give the conclusions of the thesis and suggestions for future research in Chapter 6.

# CHAPTER 2

## PRELIMINARIES

WSNs are in a form of collective sensor nodes that must provide cooperation in performing particular functions. Specifically, one of the prominent applications of WSNs is performing event detection by utilizing the ability of these nodes to sense, process data, and communicate with each other. In several WSN applications such as fire detection and target detection, the aim is to evaluate and estimate the generated signal from a point source by collecting the observations from the individual sensor nodes. Differently, in some applications such as seismic monitoring and temperature monitoring where the physical phenomenon is dispersed out over the sensor field, the source can be modeled as a field source. In this thesis, we are interested in the DD of point sources with the aid of a WSN where we possibly exploit the spatial correlation among the sensors. So, understanding the basics of DD is crucial and in this chapter, we firstly give the basics of DD. Afterwards, we give the fundamentals of composite hypothesis testing since in the considered DD problem composite hypothesis tests are required to make at local detectors. Subsequently, we mention the basics of sequential detection. Then, in the last part of the chapter, we shortly give the essentials of energy detection which is the applied method of detection in the thesis when spatial correlation of the sensor observations are taken into account.

### 2.1. Distributed Detection

For the following part of the thesis, we will consider the DD problem for binary hypothesis testing using NP detection formulation and Bayesian formulation.

### 2.1.1. Distributed Neyman-Pearson Detection

The DD problem includes the design of the detection rule at the local detectors and at the FC. Firstly, we will consider a detection problem by assuming that the local detector performances are known and they are given by the variables  $(p_{fa_1}, p_{d_1})$  and  $(p_{fa_2}, p_{d_2})$ . Moreover, we consider that the local detectors transmit their binary decisions to the FC over perfect reporting channels as shown in Figure 2.1. In the following, we will design the detection rules for given local binary decisions  $u_1$  and  $u_2$ , for the first and second detectors, respectively. We can write the likelihood functions under hypothesis  $H_0$  and  $H_1$  for the local decision vector  $\mathbf{u} = [u_1 \ u_2]$ , which represent the observations at FC, respectively as follows:

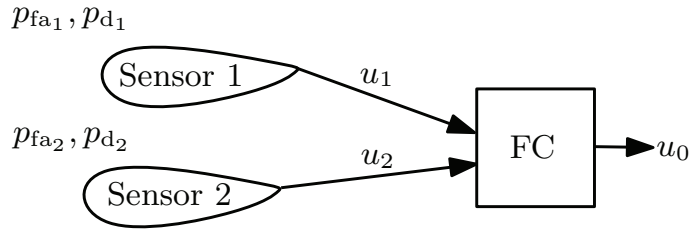


Figure 2.1. Distributed detection with two sensors.

$$p(\mathbf{u}|H_0) = p_{fa_1}^{u_1} (1 - p_{fa_1})^{1-u_1} p_{fa_2}^{u_2} (1 - p_{fa_2})^{1-u_2} \quad (2.1)$$

and

$$p(\mathbf{u}|H_1) = p_{d_1}^{u_1} (1 - p_{d_1})^{1-u_1} p_{d_2}^{u_2} (1 - p_{d_2})^{1-u_2}, \quad (2.2)$$

assuming  $p_{d_i} > 0.5$  and  $p_{fa_i} < 0.5$  for  $i = 1, 2$ .

Note that for the two sensor we have the vector  $[u_1 \ u_2]^T$  which gives 4 possible outputs with their probabilities under  $H_0$  as follows:  $[0 \ 0]^T$  with  $(1 - p_{fa_1})(1 - p_{fa_2})$ ,  $[0 \ 1]^T$  with  $(1 - p_{fa_1})p_{fa_2}$ ,  $[1 \ 0]^T$  with  $p_{fa_1}(1 - p_{fa_2})$  and  $[1 \ 1]^T$  with  $p_{fa_1}p_{fa_2}$ . These 4 cases construct  $p(\mathbf{u}|H_0)$ . In the same way, we can write  $p(\mathbf{u}|H_1)$ . We give them in Table 2.1 also. From (2.1) and (2.2) we can obtain the likelihood ratio (LR),  $\Lambda(\mathbf{u})$ , as:

$$\Lambda(\mathbf{u}) = \frac{p(\mathbf{u}|H_1)}{p(\mathbf{u}|H_0)} = \frac{(1 - p_{d_1})(1 - p_{d_2})}{(1 - p_{fa_1})(1 - p_{fa_2})} \left[ \frac{p_{d_1}(1 - p_{fa_1})}{p_{fa_1}(1 - p_{d_1})} \right]^{u_1} \left[ \frac{p_{d_2}(1 - p_{fa_2})}{p_{fa_2}(1 - p_{d_2})} \right]^{u_2}. \quad (2.3)$$

Taking the natural logarithm of (2.3) gives:

$$\log(\Lambda(\mathbf{u})) = u_1 \log\left(\frac{p_{d1}(1-p_{fa1})}{p_{fa1}(1-p_{d1})}\right) + u_2 \log\left(\frac{p_{d2}(1-p_{fa2})}{p_{fa2}(1-p_{d2})}\right) + \frac{(1-p_{d1})(1-p_{d2})}{(1-p_{fa1})(1-p_{fa2})} \quad (2.4)$$

where  $\log(\cdot)$  denotes the natural logarithm. With more simplification of (2.4) we can re-write it as :

$$\log(\Lambda(\mathbf{u})) = w_1 u_1 + w_2 u_2 + w_0, \quad (2.5)$$

where  $w_1 = \log\left(\frac{p_{d1}(1-p_{fa1})}{p_{fa1}(1-p_{d1})}\right) > 0$ ,  $w_2 = \log\left(\frac{p_{d2}(1-p_{fa2})}{p_{fa2}(1-p_{d2})}\right) > 0$  and  $w_0 = \frac{(1-p_{d1})(1-p_{d2})}{(1-p_{fa1})(1-p_{fa2})}$ .

For this case, we consider an NP type detection at the FC using binary decisions from the two sensors ( $u_1$  and  $u_2$ ) in order to make a global decision  $u_0 \equiv \{0(H_0), 1(H_1)\}$ :

$$u_{0_{NP}} = \begin{cases} H_1 & \text{if } \log(\Lambda(\mathbf{u})) > \eta \\ H_1 \text{ w.p. } \alpha & \text{if } \log(\Lambda(\mathbf{u})) = \eta \\ H_0 & \text{if } \log(\Lambda(\mathbf{u})) < \eta \end{cases} \quad (2.6)$$

where w.p. stands for "with probability". For this case, we can see that the  $\log(\Lambda(\mathbf{y}))$  takes only 4-values as shown in Table 2.1 which means that a randomized test is needed to achieve all points on the receiver operating characteristics (ROC) curve. Referring to Figure 2.2 and (2.6) the resulting NP test can be given as:

$$u_{0_{NP}} = \begin{cases} 1 & \text{if } \log(\Lambda(\mathbf{u})) > w_0 \\ 1 \text{ w.p. } \alpha = \frac{p_{fa}^{FC} - \alpha_1}{1 - \alpha_1} & \text{if } \log(\Lambda(\mathbf{u})) = w_0 \end{cases} \quad (2.7)$$

for  $\alpha_1 \leq p_{fa}^{FC} < 1$ ,

$$u_{0_{NP}} = \begin{cases} 1 & \text{if } \log(\Lambda(\mathbf{u})) > w_1 + w_0 \\ 1 \text{ w.p. } \alpha = \frac{p_{fa}^{FC} - \alpha_2}{\alpha_1 - \alpha_2} & \text{if } \log(\Lambda(\mathbf{u})) = w_1 + w_0 \\ 0 & \text{if } \log(\Lambda(\mathbf{u})) = w_0 \end{cases} \quad (2.8)$$

for  $\alpha_2 \leq p_{fa}^{FC} < \alpha_1$ ,

$$u_{0_{NP}} = \begin{cases} 1 & \text{if } \log(\Lambda(\mathbf{u})) > w_2 + w_0 \\ 1 \text{ w.p. } \alpha = \frac{p_{fa}^{FC} - \alpha_3}{\alpha_2 - \alpha_3} & \text{if } \log(\Lambda(\mathbf{u})) = w_2 + w_0 \\ 0 & \text{if } \log(\Lambda(\mathbf{u})) < w_2 + w_0 \end{cases} \quad (2.9)$$

for  $\alpha_3 \leq p_{fa}^{FC} < \alpha_2$  and

$$u_{0_{NP}} = \begin{cases} 0 & \text{if } \log(\Lambda(\mathbf{u})) < w_1 + w_2 + w_0 \\ 1 \text{ w.p. } \alpha = \frac{p_{fa}^{FC}}{\alpha_3} & \text{if } \log(\Lambda(\mathbf{u})) = w_1 + w_2 + w_0 \end{cases} \quad (2.10)$$

for  $0 \leq p_{fa}^{FC} < \alpha_3$ ,  $\alpha_1$ ,  $\alpha_2$  and  $\alpha_3$  can be given as:

$$\alpha_1 = 1 - (1 - p_{fa_1})(1 - p_{fa_2}),$$

$$\alpha_2 = 1 - (1 - p_{fa_1})(1 - p_{fa_2}) - p_{fa_1}(1 - p_{fa_2}),$$

$$\alpha_3 = 1 - (1 - p_{fa_1})(1 - p_{fa_2}) - p_{fa_1}(1 - p_{fa_2}) - (1 - p_{fa_2})p_{fa_1}.$$

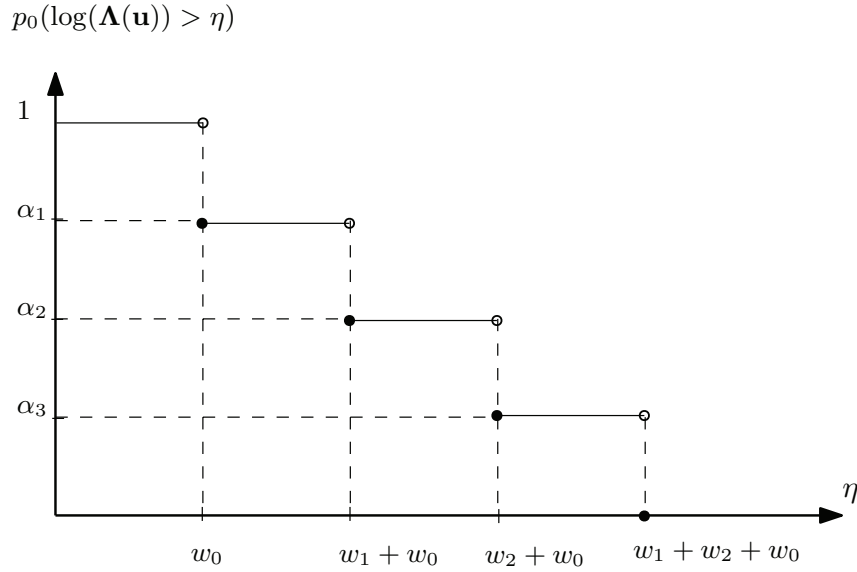


Figure 2.2. Probability that the LR is greater than the threshold,  $\eta$ , under  $H_0$  in the DD problem consisting of 2 sensors.

Table 2.1. The possible outputs for the two sensors and their corresponding probabilities.

$u_1$	$u_2$	$\log(\Lambda(\mathbf{u}))$	$p(u H_0)$	$p(u H_1)$
0	0	$w_0$	$(1 - p_{fa1})(1 - p_{fa2})$	$(1 - p_{d1})(1 - p_{d2})$
0	1	$w_2 + w_0$	$(1 - p_{fa1})p_{fa2}$	$(1 - p_{d1})p_{d2}$
1	0	$w_1 + w_0$	$p_{fa1}(1 - p_{fa2})$	$p_{d1}(1 - p_{d2})$
1	1	$w_1 + w_2 + w_0$	$p_{fa1}p_{fa2}$	$p_{d1}p_{d2}$

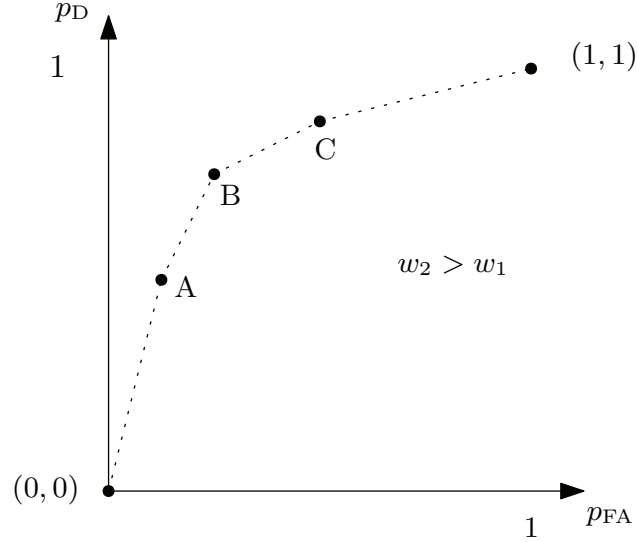


Figure 2.3. ROC curve for DD problem consisting of 2 sensors.

## 2.1.2. Distributed Bayesian Detection

In the subsection 2.1.1, we considered a DD problem which supposes that the local detectors are already designed and we made a design for the FC whereas in this subsection we will design the whole system including the decision rules at each sensor and the fusion rule at the FC.

### 2.1.2.1. Derivation of the Bayesian Local Decision Rules when All Decision Rules are Inter-Dependent

In this subsection, we present a Bayesian approach for a DD problem consisting of  $K$ -sensors connected to the FC through perfect reporting channels to perform a binary



Table 2.2. Monotonic global fusion rules at the FC according to binary decisions from two sensors.

$u_1 \ u_2$		$u_0$				
		(0,0)	A	B	C	(1,1)
0	0	0	0	0	0	1
0	1	0	0	1	1	1
1	0	0	0	0	1	1
1	1	0	1	1	1	1

global decision  $u_0$ , as shown in Figure 2.4. Let  $\mathbf{y} = [y_1, y_2, \dots, y_K]^T$  represent the observations vector and  $\mathbf{u} = [u_1, u_2, \dots, u_K]^T$  represent the local decision vector. We aim to find the fusion rule which minimizes the Bayesian risk  $\mathfrak{R}$  given by (2.11). The prior probabilities  $p(H_j)$  and the decision costs  $c_{ij}$  are needed to solve this problem;  $c_{ij}$  denotes the cost to decide  $H_i$  when  $H_j$  is true where  $i, j = 0, 1$ .

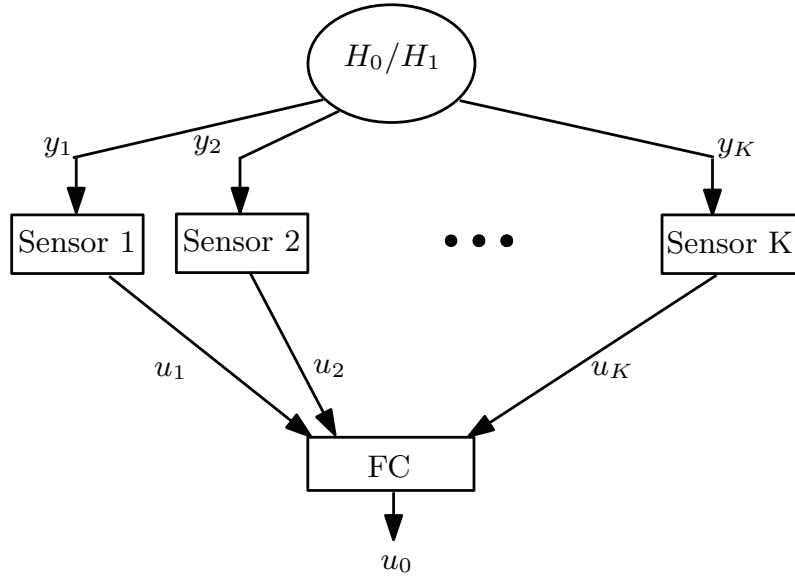


Figure 2.4. Parallel topology of DD with fusion center.

$$\begin{aligned}
 \mathfrak{R} &= c_{00}p(H_0) (1 - p_{\text{FA}}) + c_{10} p(H_0) p_{\text{FA}} + c_{01} p(H_1) (1 - p_{\text{D}}) + c_{11} p(H_1) p_{\text{D}} \\
 &= c_{00} p(H_0) + c_{01} p(H_1) + (c_{10} - c_{00}) p(H_0) p_{\text{FA}} - (c_{01} - c_{11}) p(H_1) p_{\text{D}}
 \end{aligned} \tag{2.11}$$

where  $p_{\text{FA}}$  and  $p_{\text{D}}$  represent the probability of false alarm and probability of detection at the FC which are given respectively by:

$$p_{\text{FA}} = \sum_{\mathbf{u}} p(u_0 = 1|\mathbf{u}) p(\mathbf{u}|H_0) \quad (2.12)$$

and

$$p_{\text{D}} = \sum_{\mathbf{u}} p(u_0 = 1|\mathbf{u}) p(\mathbf{u}|H_1). \quad (2.13)$$

We can simplify the Bayesian risk in (2.11) as

$$\mathfrak{R} = c + c_{\text{FA}} p_{\text{FA}} - c_{\text{D}} p_{\text{D}}, \quad (2.14)$$

where

$$c = c_{00} p(H_0) + c_{01} p(H_1),$$

$$c_{\text{FA}} = (c_{10} - c_{00}) p(H_0) > 0,$$

$$c_{\text{D}} = (c_{01} - c_{11}) p(H_1) > 0,$$

$$c_{10} > c_{00}, c_{01} > c_{11}.$$

By substituting the global probability expression of false alarm  $p_{\text{FA}}$  from (2.12) and the global probability of detection  $p_{\text{D}}$  from (2.13) in (2.14) we can rewrite it as in [7, sec. 2], (Hoballah and Varshney, 1989)

$$\mathfrak{R} = c + c_{\text{FA}} \sum_{\mathbf{u}} p(u_0 = 1|\mathbf{u}) p(\mathbf{u}|H_0) - c_{\text{D}} \sum_{\mathbf{u}} p(u_0 = 1|\mathbf{u}) p(\mathbf{u}|H_1), \quad (2.15)$$

where  $\sum_{\mathbf{u}}$  represents the summation over all possible values of  $u$  which corresponds to  $2^K$  possibilities. The end result for the local decision rule at sensor  $m$  is obtained as in (2.16) and the derivations are given in Appendix A.

$$\lambda(y_m) \underset{H_0}{\overset{H_1}{\gtrless}} \frac{c_{\text{FA}} \sum_{\mathbf{u}^m} \int_{\mathbf{y}^m} A(\mathbf{u}^m) \left[ \prod_{k=1, k \neq m}^K p(u_k|y_k) \right] p(\mathbf{y}^m|y_m, H_0) dy}{c_{\text{D}} \sum_{\mathbf{u}^m} \int_{\mathbf{y}^m} A(\mathbf{u}^m) \left[ \prod_{k=1, k \neq m}^K p(u_k|y_k) \right] p(\mathbf{y}^m|y_m, H_1) dy}. \quad (2.16)$$

where

$\mathbf{u}^m = [u_1, u_2, \dots, u_{m-1}, u_{m+1}, \dots, u_K]^T$ , is the local decision vector excluding the  $m^{\text{th}}$  decision with  $K - 1$  elements,

$$\mathbf{u}^{mj} = [u_1, u_2, \dots, u_{m-1}, u_m = j, u_{m+1}, \dots, u_K]^T,$$

$\mathbf{y}^m = [y_1, y_2, \dots, y_{m-1}, y_{m+1}, \dots, y_K]^T$ , observation vector excluding the  $m^{\text{th}}$  observation

with  $K - 1$  elements,  $j = 0, 1$ , and

$$A(\mathbf{u}^m) = (p(u_0 = 1|\mathbf{u}^{m1}) - p(u_0 = 1|\mathbf{u}^{m0})).$$

The LRT in (2.16), depends on the global decision  $A(u^m)$ , and also on the decisions from the other sensors  $p(u_k|y_k)$  and even it's a function of the observations for the same sensor  $p(\mathbf{y}^m|y_m, H_j)$ ,  $j = 0, 1$ , which makes it data dependent. For this reason, it is called as non-conventional or data dependent LRT (Viswanathan and Varshney, 1997).

### 2.1.2.2. Conditionally Independent Local Observations

If the observations of each detector are independent of other detectors then  $p(\mathbf{y}^m|y_m, H_j)$  can be written as  $p(\mathbf{y}^m|H_j)$  which is an independent decision rule; the decision rule will not be data-dependent as in the previous section 2.1.2.1 and it will be given similar to (Varshney, 2012) as:

$$\lambda(y_m) \underset{H_0}{\overset{H_1}{\gtrless}} \frac{c_{FA} \sum_{\mathbf{u}^m} A(\mathbf{u}^m) \prod_{k=1, k \neq m}^K \int_{\mathbf{y}^m} p(u_k|y_k)p(y_K|H_0)dy_k}{c_D \sum_{\mathbf{u}^m} A(\mathbf{u}^m) \prod_{k=1, k \neq m}^K \int_{\mathbf{y}^m} p(u_k|y_k)p(y_K|H_1)dy_k}, \quad (2.17)$$

the right-hand side which is a constant, and thus the local decision rule becomes a standard LRT with a constant threshold.

### 2.1.2.3. Fusion Rule at the FC

In a similar manner to what we have described in section 2.1.2.1 we can drive the fusion rule at the FC. Let  $\mathbf{u}^*$  be a one out of  $2^K$  possible values of the local decision  $\mathbf{u}$ , i.e. if we have two sensors  $\mathbf{u}^*$  could be one possibility from the four possible values like  $[1 \ 0]$ . Then, we can express (2.15) as follows:

$$\begin{aligned} \mathfrak{R} = & c + c_{FA} p(u_0 = 1|\mathbf{u}^*) p(\mathbf{u}^*|H_0) - c_D p(u_0 = 1|\mathbf{u}^*) p(\mathbf{u}^*|H_1) \\ & + c_{FA} \sum_{\mathbf{u} \neq \mathbf{u}^*} p(u_0 = 1|\mathbf{u}) p(\mathbf{u}|H_0) - c_D \sum_{\mathbf{u} \neq \mathbf{u}^*} p(u_0 = 1|\mathbf{u}) p(\mathbf{u}|H_1). \end{aligned} \quad (2.18)$$

The term  $p(u_0 = 1|\mathbf{u}^*)$  in (2.18) represents the probability of global decision at the FC in favor of  $H_1$  for a particular local decision  $\mathbf{u}^*$ .

Rearranging the previous equation and taking the global decision  $p(u_0 = 1|\mathbf{u}^*)$  as a common factor gives:

$$\begin{aligned} \mathfrak{R} = c + c_{\text{FA}} \sum_{\mathbf{u} \neq \mathbf{u}^*} p(u_0 = 1|\mathbf{u}) p(\mathbf{u}|H_0) - c_{\text{D}} \sum_{\mathbf{u} \neq \mathbf{u}^*} p(u_0 = 1|\mathbf{u}) p(\mathbf{u}|H_1) \\ + p(u_0 = 1|\mathbf{u}^*) [c_{\text{FA}} p(\mathbf{u}^*|H_0) - c_{\text{D}} p(\mathbf{u}^*|H_1)]. \end{aligned} \quad (2.19)$$

In a similar manner to what we have performed for minimizing the average risk at the local decisions, we will apply it here for minimizing the average risk at the global decision.

From (2.19), we can minimize the average risk by assigning the value 1 to  $p(u_0 = 1|u^*)$  if the term  $c_{\text{FA}}p(\mathbf{u}^*|H_0) - c_{\text{D}}p(\mathbf{u}^*|H_1) < 0$  and assigning the value 0 if  $c_{\text{FA}}p(\mathbf{u}^*|H_0) - c_{\text{D}}p(\mathbf{u}^*|H_1) > 0$ , (Viswanathan and Varshney, 1997), which yields:

$$c_{\text{FA}} p(\mathbf{u}^*|H_0) dy \underset{H_0}{\overset{H_1}{\gtrless}} c_{\text{D}} p(\mathbf{u}^*|H_1) dy. \quad (2.20)$$

We have  $2^K$  inequalities, which represent the number of possibilities for  $\mathbf{u}^*$ . Where

$$p(\mathbf{u}^*|H_0) = \int_{\mathbf{y}} p(\mathbf{u}^*|\mathbf{y}) p(\mathbf{y}|H_0) dy \quad (2.21)$$

for the continuous observations. As we mentioned previously, if the local detectors are independent of each other then we can rewrite (2.20) as

$$c_{\text{FA}} \int_{\mathbf{y}} \left( \prod_{k=1}^K p(\mathbf{u}_k^*|y_k) \right) p(\mathbf{y}|H_0) dy \underset{H_0}{\overset{H_1}{\gtrless}} c_{\text{D}} \int_{\mathbf{y}} \left( \prod_{k=1}^K p(\mathbf{u}_k^*|y_k) \right) p(\mathbf{y}|H_1) dy. \quad (2.22)$$

In Appendix B, we consider a simple example consisting of two sensors to show how we can design a DD problem.

#### 2.1.2.4. Incorporation of Binary Symmetric Channels Between Sensors and FC

In this section, we consider accounting of a simplified realistic channel model for transmission of sensor data to the FC, namely binary symmetric channel (BSC). In this model, a bit transmitted from sensor  $k$  arrives at the FC without error with probability  $(1 - b_k)$  and correspondingly inverted with probability  $b_k$ . As shown in Figure 2.5, we will

consider a parallel topology consisting of two sensors and one FC. Assume the observations are independently and identically distributed (i.i.d.) and the BSCs are characterized by  $p(\hat{u}_1 = 0|u_1 = 0) = 1 - b_k$ ,  $p(\hat{u}_1 = 1|u_1 = 0) = b_k$ ,  $p(\hat{u}_1 = 1|u_1 = 1) = 1 - b_k$  and  $p(\hat{u}_1 = 0|u_1 = 1) = b_k$  as shown in Figure 2.6.

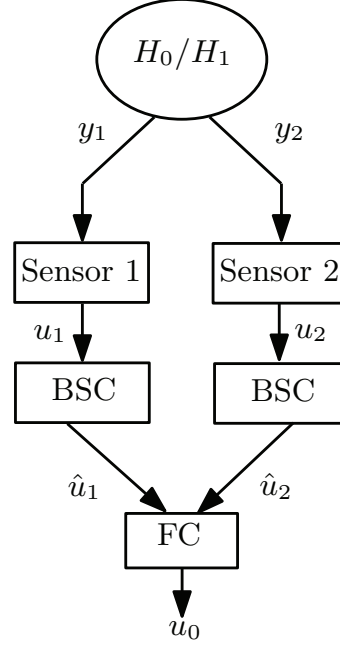


Figure 2.5. DD problem with BSC

From (2.17) the likelihood ratio at the first sensor is

$$\lambda(y_1) \underset{H_0}{\overset{H_1}{\gtrless}} \frac{c_{FA} \sum_{u_2} \int_{y_2} A(u_2) p(u_2|y_2) p(y_2|H_0) dy_2}{c_D \sum_{u_2} \int_{y_2} A(u_2) p(u_2|y_2) p(y_2|H_1) dy_2} \quad (2.23)$$

where  $A(u_2)$  for the case of an ideal channel depends directly on the local decisions  $u_1$  and  $u_2$  as follows

$$A(u_2) = p(u_0 = 1|u_1 = 1, u_2) - p(u_0 = 1|u_1 = 0, u_2). \quad (2.24)$$

On the other hand, using the imperfect channels BSCs affects the decision at the FC which depends on the values of  $\hat{u}_1$  and  $\hat{u}_2$  instead of  $u_1$  and  $u_2$ . Using the law of total probability,  $A(u_2)$  is given as

$$A(u_2) = \sum_{\hat{u}_1, \hat{u}_2} p(u_0 = 1|\hat{u}_1, \hat{u}_2) p(\hat{u}_1, \hat{u}_2|u_1 = 1, u_2) - p(u_0 = 1|\hat{u}_1, \hat{u}_2) p(\hat{u}_1, \hat{u}_2|u_1 = 0, u_2). \quad (2.25)$$

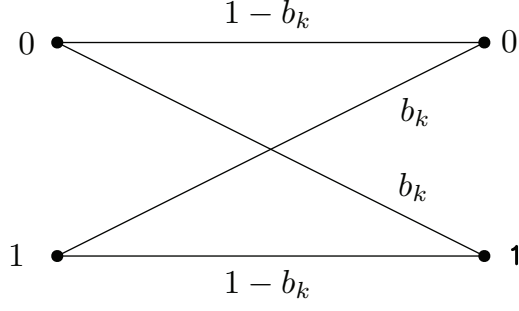


Figure 2.6. Binary symmetric channel

Let  $\eta_{ijk} = p(u_0 = i | \hat{u}_1 = j, \hat{u}_2 = k)$  where  $i, j, k = 0, 1$ , then

$$A(u_2) = \sum_{\hat{u}_1, \hat{u}_2} \eta_{1\hat{u}_1, \hat{u}_2} p(\hat{u}_2 | u_2) [p(\hat{u}_1 | u_1 = 1) - p(\hat{u}_1 | u_1 = 0)] \quad (2.26)$$

where  $i, j, k = 0, 1$ .

We can obtain the LRT at the second sensor in a similar fashion as follows:

$$\lambda(y_m) \underset{H_0}{\overset{H_1}{\gtrless}} \frac{c_{FA} \sum_{u_1} \int_{y_1} A(u_1) p(u_1 | y_1) p(y_1 | H_0) dy_1}{c_D \sum_{u_1} \int_{y_1} A(u_1) p(u_1 | y_1) p(y_1 | H_1) dy_1} \quad (2.27)$$

where

$$\begin{aligned} A(u_1) &= p(u_0 = 1 | u_2 = 1, u_1) - p(u_0 = 1 | u_2 = 0, u_1) \\ &= \sum_{\hat{u}_1, \hat{u}_2} p(u_0 = 1 | \hat{u}_1, \hat{u}_2) p(\hat{u}_1, \hat{u}_2 | u_2 = 1, u_1) - p(u_0 = 1 | \hat{u}_1, \hat{u}_2) p(\hat{u}_1, \hat{u}_2 | u_2 = 0, u_1). \end{aligned} \quad (2.28)$$

Substituting  $\eta_{ijk}$  in (2.28) gives:

$$A(u_1) = \sum_{\hat{u}_1, \hat{u}_2} \eta_{1\hat{u}_1, \hat{u}_2} p(\hat{u}_1 | u_1) [p(\hat{u}_2 | u_2 = 1) - p(\hat{u}_2 | u_2 = 0)]. \quad (2.29)$$

The fusion rule at the FC for the BSCs can be obtained in a similar manner to what is derived in section 2.1.2.3. Let  $\mathbf{u}^*$  be a particular outcome for  $\hat{u}_1$  and  $\hat{u}_2$  values. Then the decision rule will be given as

$$c_{FA} p(\mathbf{u}^* | H_0) \underset{H_0}{\overset{H_1}{\gtrless}} c_D p(\mathbf{u}^* | H_1), \quad (2.30)$$

where

$$\begin{aligned}
p(\mathbf{u}^*|H_0) &= \sum_{\mathbf{u}} p(\mathbf{u}^*|\mathbf{u}) p(\mathbf{u}|H_0) d\mathbf{y} \\
&= \sum_{\mathbf{u}} \int_{\mathbf{y}} p(\mathbf{u}^*|\mathbf{u}) p(\mathbf{u}|\mathbf{y}) p(\mathbf{y}|H_0) d\mathbf{y} \\
&= \sum_{u_1, u_2} \int_{y_1} \int_{y_2} \underbrace{p(\hat{u}_1|u_1) p(\hat{u}_2|u_2)} p(u_1|y_1) p(u_2|y_2) p(y_1|H_0) p(y_2|H_0) dy_1 dy_2.
\end{aligned} \tag{2.31}$$

and

$$p(\mathbf{u}^*|H_1) = \sum_{u_1, u_2} \int_{y_1} \int_{y_2} \underbrace{p(\hat{u}_1|u_1) p(\hat{u}_2|u_2)} p(u_1|y_1) p(u_2|y_2) p(y_1|H_1) p(y_2|H_1) dy_1 dy_2. \tag{2.32}$$

The underbraced terms in the previous equations correspond to the channel between the local sensors and the FC.

### 2.1.2.5. Identical Local Detectors

In this section, we will consider a special case of interest that happens when the local detectors are identical. The probability of false alarm and the probability of detection at each sensor are the same,  $\tilde{p}_{fa}$  and  $\tilde{p}_d$  respectively, which is only possible in the case of field sources (Yu and Ephremides, 2006; Tsitsiklis, 1988). We assume that each sensor performs  $T$  binary observations,  $y_k(t) = 0$  or  $1$ ,  $k = 1, \dots, K$  and  $t = 1, \dots, T$ , additionally, the observations at each sensor and across sensors follow an i.i.d. distribution. The number of ones,  $n_k$  out of  $T$  observations at each sensor is a sufficient statistic to make its local decision, which obeys a binomial pdf under both hypotheses  $H_0$  and  $H_1$  respectively which given as:

$$p[n_k|H_0] = \binom{T}{n_k} (\tilde{p}_{fa})^{n_k} (1 - \tilde{p}_{fa})^{T-n_k} \tag{2.33}$$

and

$$p[n_k|H_1] = \binom{T}{n_k} (\tilde{p}_d)^{n_k} (1 - \tilde{p}_d)^{T-n_k}. \tag{2.34}$$

From Bayes rule

$$\frac{p[n_k|H_1]}{p[n_k|H_0]} \underset{H_0}{\overset{H_1}{\geq}} \frac{p(H_0)}{p(H_1)}. \tag{2.35}$$

Substituting (2.33) and (2.34) in (2.36) gives:

$$\frac{\binom{T}{n_k} (\tilde{p}_d)^{n_k} (1 - \tilde{p}_d)^{T-n_k}}{\binom{T}{n_k} (\tilde{p}_{fa})^{n_k} (1 - \tilde{p}_{fa})^{T-n_k}} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{p[H_0]}{p[H_1]}. \quad (2.36)$$

Taking the natural logarithm and rearranging the terms in (2.36) yields the local threshold  $\beta$  at each sensor as follows:

$$n_k \geq \frac{\log \frac{p(H_0)}{p(H_1)} + T \log \frac{1-\tilde{p}_{fa}}{1-\tilde{p}_d}}{\log \frac{\tilde{p}_d(1-\tilde{p}_{fa})}{\tilde{p}_{fa}(1-\tilde{p}_d)}} = \beta. \quad (2.37)$$

The local optimal fusion rule at each sensor is given as

$$\ell_k = \begin{cases} H_1 & \text{if } n_k \geq \beta \\ H_0 & \text{if } n_k < \beta. \end{cases} \quad (2.38)$$

The final decision at the FC is performed depending on the binary local decisions  $[\ell_1, \ell_2, \dots, \ell_K]$  at  $K$  sensors, which also follows an i.i.d. distribution. Thus the total number of ones at the FC,  $L = \sum_{k=1}^K \ell_k$ , obey also a binomial pdf under both hypothesis  $H_0$  and  $H_1$  respectively as follows:

$$p[L|H_0] = \binom{K}{L} p_{fa}^L (1 - p_{fa})^{K-L} \quad (2.39)$$

and

$$p[L|H_1] = \binom{K}{L} p_d^L (1 - p_d)^{K-L}, \quad (2.40)$$

where  $p_{fa}$ ,  $p_d$  represent the probability of false alarm and probability of detection at each local detector, which is given respectively as follows:

$$p_{fa} = p[\ell_k = 1|H_0] = p[n_k \geq \beta|H_0] = \sum_{n_k=\lceil\beta\rceil}^T \binom{T}{n_k} (\tilde{p}_{fa})^{n_k} (1 - \tilde{p}_{fa})^{T-n_k}, \quad (2.41)$$

and



$$p_d = p[\ell_k = 1|H_1] = p[n_k \geq \beta|H_1] = \sum_{n_k=\lceil\beta\rceil}^T \binom{T}{n_k} (\tilde{p}_d)^{n_k} (1 - \tilde{p}_d)^{T-n_k} \quad (2.42)$$

$\lceil\beta\rceil$  denotes the ceiling function which maps  $\beta$  to the least integer greater than or equal to  $\beta$ . Again, using Bayes rule

$$\frac{p[L|H_1]}{p[L|H_0]} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{p(H_0)}{p(H_1)} \quad (2.43)$$

we will write the threshold at the global detector as

$$\eta = \frac{\log \frac{p(H_0)}{p(H_1)} + K \log \frac{1-p_{fa}}{1-p_d}}{\log \frac{p_d(1-p_{fa})}{p_{fa}(1-p_d)}}. \quad (2.44)$$

The final decision rule will be as follows:

$$\hat{H} = \begin{cases} H_1 & \text{if } b \geq \eta \\ H_0 & \text{if } b < \eta. \end{cases} \quad (2.45)$$

The overall false alarm and detection probabilities are given by

$$p_{FA} = p[L \geq \eta|H_0] = \sum_{L=\lceil\eta\rceil}^K \binom{K}{L} p_{fa}^L (1 - p_{fa})^{K-L}, \quad (2.46)$$

$$p_D = p[L \geq \eta|H_1] = \sum_{L=\lceil\eta\rceil}^K \binom{K}{L} p_d^L (1 - p_d)^{K-L}. \quad (2.47)$$

In our previous discussions we considered the case of one-bit decision but also it is possible to perform quantized  $M$ -bit quantity ( $q_k$  for  $s_k, q_k \in \{0, 1, \dots, 2^M - 1\}$ ,  $1 \leq M \leq T$ ) at each sensor. Then, the FC makes the last decision based on the quantized information from the  $K$ -sensors  $\{q_1, q_2, \dots, q_K\}$ . In (Zhang et al., 2002) some optimal quantization algorithm was developed where they supposed that the number of 1's out of  $T$  observations  $\{n_1, n_2, \dots, n_T\}$  is a sufficient statistic, so it is sufficient to quantize  $n_k$  into an  $M$ -bit quantity  $q_k$  at  $s_k$  and send it to the FC. Therefore, the quantization algorithm is a mapping of  $n_k \in \{0, 1, \dots, T\}$  to  $q_k \in \{0, 1, \dots, 2^M - 1\}$  for  $k = 1, \dots, K$ .

All sensor nodes apply identical quantization algorithms because the system parameters are the same for all the nodes. For binary hypothesis testing based on the  $k$  quantized

quantities  $\{q_1, q_2, \dots, q_k\}$ , the decision has been taken as follows:

$\hat{H} = H_1$  if  $p[H_1|q_1, \dots, q_k] \geq p[H_0|q_1, \dots, q_k]$ . The optimal decision rule at the control center is given by

$$\hat{H} = \begin{cases} H_1 & \text{if } \frac{p[q_1, \dots, q_k|H_1]}{p[q_1, \dots, q_k|H_0]} \geq \frac{p[H_0]}{p[H_1]} \\ H_0 & \text{if } \frac{p[q_1, \dots, q_k|H_1]}{p[q_1, \dots, q_k|H_0]} < \frac{p[H_0]}{p[H_1]}. \end{cases} \quad (2.48)$$

## 2.2. Sequential Detection

In the classical theory of hypothesis testing, the number of observations is treated as constant. Many approaches such as the NP approach and Bayesian approach can be employed in a fixed sample size (FSS) detection. On the other hand, the sequential test depends on the outcome of the observation process and the test thresholds, therefore, the number of observations is not predetermined, but a random value. For some experiments, a decision can be made by taking a small number of observations, while for others, the process of making observations is extended before making a decision. The main two approaches of designing the sequential tests are the Bayesian sequential test and SPRT which are presented next.

### 2.2.1. Bayesian Sequential Test

Bayesian sequential test expresses the hypothesis testing problem as an optimal stopping problem where the prior probabilities are assumed to be known and there is a cost of  $c$  units for each observation. The cost of deciding  $H_i$  is true when  $H_j$  is present is denoted as  $c_{ij}$  for  $i, j = 0, 1$ . The Bayes sequential test takes samples  $y_1, \dots, y_K$  until  $\Lambda(\mathbf{y})$  falls outside the interval  $(\underline{\pi}, \bar{\pi})$  and then it decides on  $H_0$  or  $H_1$ , depending on whether  $\Lambda(\mathbf{y})$  falls below  $\underline{\pi}$  or above  $\bar{\pi}$ .

Sequential decision rule  $(\phi, \delta)$  for the likelihood ratio can be given in terms of two decision statistics,  $\phi$  and  $\delta$ , as follows:

$$\phi_n(\mathbf{y}) = \begin{cases} 0 & \text{if } \underline{\pi} < \Lambda(\mathbf{y}) < \bar{\pi} \\ 1 & \text{otherwise,} \end{cases} \quad (2.49)$$

and

$$\delta_n(\mathbf{y}) = \begin{cases} 1 & \text{if } \Lambda(\mathbf{y}) \geq \bar{\pi} \\ 0 & \text{if } \Lambda(\mathbf{y}) \geq \underline{\pi}, \end{cases} \quad (2.50)$$

where

$$\bar{\pi} = \frac{\gamma_k^u p(H_1)}{1 - \gamma_k^u p(H_0)}, \quad (2.51)$$

$$\underline{\pi} = \frac{\gamma_k^l p(H_1)}{1 - \gamma_k^l p(H_0)}, \quad (2.52)$$

and  $\gamma_k^u, \gamma_k^l$  are the upper and lower thresholds defined for the posterior probability (Levy, 2008, Poor; 2013). The thresholds can be found by using the backward induction method, which proceeds backward in time, from the end of a situation or problem, to decide a series of optimal actions, i.e. by finding the threshold at step 20 in Figure 2.7 we can proceed to find the optimal thresholds at any step (DeGroot, 2005; Ferguson, 2014). Figure 2.8 is an example to illustrate the effect of the sample cost on the upper and lower thresholds for the Bayesian sequential test. The simulation has been performed by considering Bayesian costs  $c_{01} = 1, c_{10} = 2$  and  $c_{11} = c_{00} = 0$ , for two different value of the sample cost  $c = 0.3$  and  $c = 0.03$  as shown in Figure 2.7 and 2.8 respectively, whereas the likelihood functions are given below:

$$p(y|H_0) = \left( \frac{1}{\sqrt{2\pi}} \right) \exp\left( -\frac{(y_m + 1)^2}{2} \right), \quad (2.53)$$

$$p(y|H_1) = \left( \frac{1}{\sqrt{2\pi}} \right) \exp\left( -\frac{(y_m - 1)^2}{2} \right). \quad (2.54)$$

We can notice in Figures 2.7 and 2.8 that the distance between the upper and lower thresholds decreases by increasing the sample cost which gives the sequential test a higher opportunity to exceed one of the two thresholds and take the final decision with less number of observations.

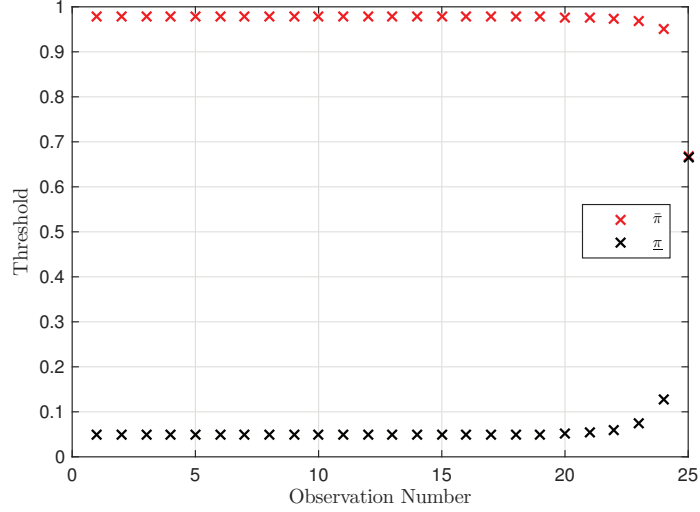


Figure 2.7. The thresholds for the Bayesian sequential test with a sample cost  $c=0.03$ .

### 2.2.2. Sequential Probability Ratio Test

This approach aims to minimize the expected number of observations, as well as to satisfy the required probability of detection and probability of false alarm under the two hypotheses  $H_j$ , for  $j = 0, 1$ . The optimal solution for this problem is known as Wald's SPRT (Wald and Wolfowitz, 1948). For developing SPRT, consider the problem of the binary hypothesis testing  $H_0$  versus  $H_1$  and let  $p(y_i|H_j)$  denote the probability density function (pdf) of the random variable  $y_i$ ,  $i = 1, 2, \dots, K$ , which represents the observation at each local sensor. The observations at each sensor node and across sensor nodes are assumed to be i.i.d. conditioned on hypothesis  $H_j$  for  $j = 0, 1$ . Each local detector transmits its observation to the FC where the final decision is made.  $SPRT(\eta_0, \eta_1)$  for testing  $H_1$  versus  $H_0$  is defined by two positive constants  $\eta_0$  and  $\eta_1$  ( $\eta_0 < 1 < \eta_1$ ), which are chosen to achieve specific value of detection probability, where

$$p_d \geq \eta_1 p_{fa} \quad (2.55)$$

and

$$1 - p_{fa} \geq \frac{1}{\eta_0} (p_m). \quad (2.56)$$

From (2.55) and (2.69) we can prove that by using the sequential detection we can achieve a very high  $p_d$  near to 1 and a very low  $p_{fa}$  near to 0 by increasing the number

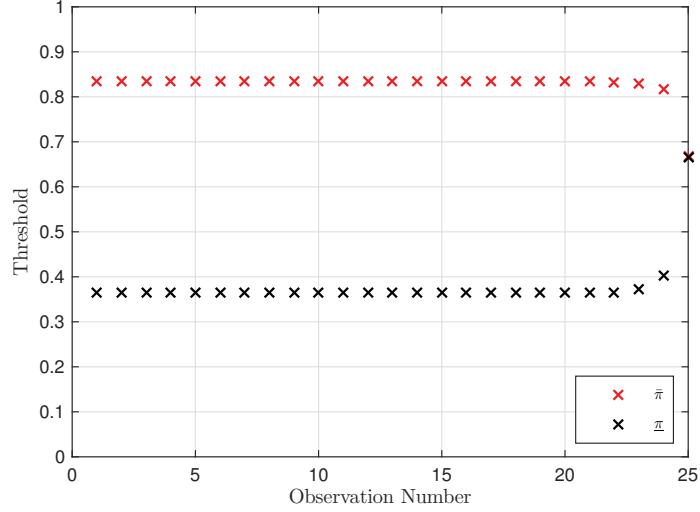


Figure 2.8. The thresholds for the Bayesian sequential test with a sample cost  $c=0.3$ .

of observations as follows:

let  $\eta_1 = \frac{A}{B}$  and  $\eta_0 = \frac{1-A}{1-B}$  where  $A, B \in [0, 1]$ . Then we can re-write (2.55) as

$$p_{fa} \leq \frac{B}{A} p_d \leq \frac{B}{A} \quad (2.57)$$

and (2.69) as

$$1 - p_d \leq \frac{1-A}{1-B} (1 - p_{fa}) \leq \frac{1-A}{1-B}, \quad (2.58)$$

by adding (2.57) and (2.59) we can find

$$p_{fa} + (1 - p_d) \leq 1 - A + B, \quad (2.59)$$

from this equation, we can see that it is possible to achieve a specific probability of error by choosing the specific value for  $A$  and  $B$ .

At each phase of testing, the likelihood ratio function  $\Lambda(\mathbf{y})$  is computed and compared to the thresholds  $\eta_0$  and  $\eta_1$ . We can characterize the performance of an SPRT in terms of the ASN and operating characteristic (OC) functions; they are discussed below (Kowalski, 1971; Levy, 2008).

#### A. Average Sample Number of the SPRT

Assume that  $K$  represents the number of observations needed to terminate the sequential test, and  $E(K|H_j)$  represents the expected value of  $K$  under hypothesis  $H_j$ , for  $j = 0, 1$ , and it is defined as the ASN which is given as follows:

$$E(K|H_1) \approx \frac{p_m \log\left(\frac{p_m}{1-p_{fa}}\right) + (1-p_m) \log\left(\frac{1-p_m}{p_{fa}}\right)}{D_{KL}(p(y|H_1)||p(y|H_0))}, \quad (2.60)$$

$$E(K|H_0) \approx \frac{(1-p_{fa}) \log\left(\frac{p_m}{1-p_{fa}}\right) + p_{fa} \log\left(\frac{1-p_m}{p_{fa}}\right)}{-D_{KL}(p(y|H_0)||p(y|H_1))}, \quad (2.61)$$

where

$$D_{KL}(f(y)||g(y)) = \int f(y) \log\left(\frac{f(y)}{g(y)}\right) dy \quad (2.62)$$

represents the Kullback-Leibler divergence between two pdfs  $f(\cdot)$  and  $g(\cdot)$ .

#### B. The OC function of the SPRT

The OC function,  $L(\theta)$ , for  $\text{SPRT}(\eta_0, \eta_1)$  can be defined as the probability of terminating the sequential test by accepting hypothesis  $H_0$  when  $\theta$  is the actual value of the parameter. OC function is important in cases where the exact knowledge of the parameter(s) of the distribution function is unknown. An approximate formula for OC is given by

$$L(\theta) = \frac{B^{h(\theta)} - 1}{\eta_1^{h(\theta)} - \eta_0^{h(\theta)}} \quad (2.63)$$

where  $h(\theta)$  is non-zero and unique solution for the following equation,

$$\int_{-\infty}^{\infty} \left[ \frac{p(\mathbf{y}|\theta_1)}{p(\mathbf{y}|\theta_0)} \right]^{h(\theta)} p(\mathbf{y}|\theta) d\mathbf{y} = 1, \quad (2.64)$$

when  $y$  is a continuous random variable and

$$\sum_{\mathbf{y}} \left[ \frac{p(\mathbf{y}|\theta_1)}{p(\mathbf{y}|\theta_0)} \right]^{h(\theta)} p(\mathbf{y}|\theta) = 1 \quad (2.65)$$

when  $y$  is a discrete random variable. From (2.64) and (2.65), it is clear that  $h(\theta_0) = 1$  and  $h(\theta_1) = -1$ . Hence, it follows from (2.63) that for  $\eta_0$  and  $\eta_1$ , we have  $L(\theta_0) = 1 - p_{fa}$  and  $L(\theta_1) = p_m$ .

### 2.3. Composite Hypothesis Testing

The simple hypothesis testing assumes full knowledge of the likelihoods under hypotheses but in practical problems, the pdf may not be known completely. For example, the strength of the received signal in a mobile communication system may not be known exactly due to shadowing effects and channel fading. The variance of noise in a radar receiver may fluctuate due to interference or jamming. The received signal amplitudes in target detection may not be known in advance. We will express the unknown parameter vector  $\theta_j$  under hypothesis  $H_j$  as  $p(\mathbf{y}|H_j; \theta_j)$ ,  $j = 0, 1$ . We will start our discussion by considering a one-sided parameter test with unknown signal amplitude  $A > 0$  as follows.

**Case 1:** Consider a composite binary hypothesis testing problem with unknown signal amplitude,  $A > 0$  in additive white Gaussian noise (AWGN)

$$\begin{aligned} H_0 : y_k &= \epsilon_k, \\ \text{versus} \\ H_1 : y_k &= A + \epsilon_k, \end{aligned} \tag{2.66}$$

where  $y_k$  denotes an i.i.d. observation under hypothesis  $H_j$ ,  $j = 0, 1$ .  $\epsilon_k$  denotes AWGN with variance  $\sigma_N^2$  and mean zero where  $k = 1, \dots, K$  represents the number of observations. We will design a NP detector that yields the maximum probability of detection  $p_d$  for any value of the probability of false alarm  $p_{fa}$  and for any value of  $A$ . This test is called a uniformly most powerful (UMP) test; it refers to the test which gives the maximum detection probability (in statistics  $p_d$  is called as the power of a test) for any value of  $A > 0$  (Kay, 1993). The NP detector will be designed as if we know the value of  $A$  and then we will manipulate it in a way making the test not dependent on the unknown parameter  $A$  as follows

$$\frac{p(\mathbf{y}|H_1, A)}{p(\mathbf{y}|H_0)} = \frac{\left(\frac{1}{\sqrt{2\pi\sigma_N^2}}\right)^K \exp\left(\frac{-\sum_{k=1}^K (y_k - A)^2}{2\sigma_N^2}\right)}{\left(\frac{1}{\sqrt{2\pi\sigma_N^2}}\right)^K \exp\left(\frac{-\sum_{k=1}^K y_k^2}{2\sigma_N^2}\right)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta. \tag{2.67}$$

Simplifying of (2.67) results in

$$\sum_{k=1}^K y_k \underset{H_0}{\overset{H_1}{\geq}} \tau, \quad (2.68)$$

where  $\tau = \frac{K}{2}A + \frac{\sigma_N^2}{A} \log \eta$  and  $A > 0$ . The summation of the observations  $y_k$  is the sufficient statistic for this test.

The observations under  $H_0$  follow a Gaussian distribution with zero mean and  $\sigma_N^2/K$  variance. We can determine  $\tau$  from the false alarm constraint

$$p_{\text{fa}} = Q_{\text{func}} \left( \frac{\tau}{\sqrt{\sigma_N^2/K}} \right), \quad (2.69)$$

then

$$\tau = \sqrt{K} \sigma_N Q_{\text{func}}^{-1}(p_{\text{fa}}), \quad (2.70)$$

which is independent of  $A$ ; where  $Q_{\text{func}}(\cdot)$  represents the complimentary cumulative distribution function for standard Gaussian random variable. Maximum  $p_d$  for any value of  $A > 0$  is given as:

$$p_d = 1 - Q_{\text{func}} \left( \frac{KA - \tau}{\sqrt{K} \sigma_N} \right). \quad (2.71)$$

The inequality in (2.68) will be inverted if the  $A < 0$ . Unfortunately, UMP tests rarely exist. If, for instance,  $A$  may take any value,  $-\infty < A < \infty$ , then we will obtain different tests for the positive  $A$  and the negative  $A$  which makes the test fail and consequently, we will not be able to decide whether to choose  $H_0$  or  $H_1$ . For a UMP to exist the parameter test must be one-sided. Thus, two-sided testing problems never produce UMP tests but the one-sided problem may.

Based on the previous discussion, we can manipulate the test and solve it without dependence on the unknown parameter as a special case. In the next section, we will consider two major approaches to composite hypothesis testing; the Bayesian approach and the generalized LRT .



### 2.3.1. Bayesian Approach

The Bayesian approach considers the unknown parameters  $\theta_j$  as a random variable and assigns a prior pdf to it, in order to obtain likelihood functions that are independent of the unknown parameters.

The optimal NP detector with the Bayesian approach is given (Kay, 1993) as

$$\Lambda(\mathbf{y}) = \frac{\int p(\mathbf{y}|H_1; \boldsymbol{\theta}_1)p(\boldsymbol{\theta}_1)d\boldsymbol{\theta}_1}{\int p(\mathbf{y}|H_0; \boldsymbol{\theta}_0)p(\boldsymbol{\theta}_0)d\boldsymbol{\theta}_0} \underset{H_0}{\overset{H_1}{\gtrless}} \eta, \quad (2.72)$$

where  $\boldsymbol{\theta}_j$  is a vector of the unknown parameters and the integrals are multidimensional with dimension corresponding to the ones of  $\boldsymbol{\theta}_j, j = 0, 1$ . The resulting integrals in (2.72) give an LRT which does not include unknown parameters. Even though the Bayesian approach gives an optimal decision rule, in general it suffers from some weakness such as it is difficult to assume the prior probability in many applications. Moreover the averaging concerning the prior probabilities may not be possible to find in a closed-form.

From our discussions in Case 1, we conclude that a one-sided test may produce a UMP test whereas a two-sided test never produces a UMP test. In the following example, we will use the Bayesian approach for a two-sided test.

**Case 2:** In this example, we will consider the binary hypothesis testing in (2.66) where the unknown signal amplitude,  $A$ , has a Gaussian pdf with mean zero and variance  $\sigma_A^2$  denoted as  $A \sim \mathcal{N}(0, \sigma_A^2)$  and it is independent of the noise  $\epsilon_k$ ; the symbol " $\sim$ " stands for "distributed according to". So, the conditional pdf under  $H_0$  is known whereas the conditional pdf under  $H_1$  is given by

$$p(\mathbf{y}|H_1, A) = \left( \frac{1}{\sqrt{2\pi\sigma^2}} \right)^K \exp \left( - \frac{\sum_{k=1}^K (y_k - A)^2}{2\sigma_N^2} \right). \quad (2.73)$$

But

$$\begin{aligned}
p(\mathbf{y}|H_1) &= \int_{-\infty}^{\infty} \left( \frac{1}{\sqrt{2\pi\sigma_N^2}} \right)^K \exp\left( -\frac{\sum_{k=1}^K (y_k - A)^2}{2\sigma_N^2} \right) \frac{1}{\sqrt{2\pi\sigma_A^2}} \exp\left( -\frac{A^2}{2\sigma_A^2} \right) dA \\
&= \left( \frac{1}{\sqrt{2\pi\sigma_N^2}} \right)^K \left( \frac{1}{\sqrt{1 + \frac{K\sigma_A^2}{\sigma_N^2}}} \right) \exp\left( -\frac{\sum_{k=1}^K y_k^2}{2\sigma_N^2} \right) \exp\left( -\frac{\left( \sum_{k=1}^K y_k \right)^2}{2\sigma_N^2 \left( K + \frac{\sigma_N^2}{\sigma_A^2} \right)} \right)
\end{aligned} \tag{2.74}$$

and

$$p(\mathbf{y}|H_0) = \left( \frac{1}{\sqrt{2\pi\sigma_N^2}} \right)^K \exp\left( -\frac{\sum_{k=1}^K y_k^2}{2\sigma_N^2} \right). \tag{2.75}$$

Substituting  $\theta_1 = A$  in (2.72) we can write the optimal NP detector with Bayesian approach as

$$\Lambda(\mathbf{y}) = \frac{\int_{-\infty}^{\infty} \left( \frac{1}{2\pi\sigma_N^2} \right)^{\frac{-K}{2}} \exp\left( -\frac{\sum_{k=1}^K (y_k - A)^2}{2\sigma_N^2} \right) \frac{1}{2\pi\sigma_A^2} \exp\left( -\frac{A^2}{2\sigma_A^2} \right) dA}{\left( \frac{1}{\sqrt{2\pi\sigma_N^2}} \right)^K \exp\left( -\frac{\sum_{k=1}^K y_k^2}{2\sigma_N^2} \right)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta. \tag{2.76}$$

Evaluating the integral in the numerator and performing further simplifications we can write (2.76) as

$$\Lambda(\mathbf{y}) = \frac{1}{\sqrt{1 + \frac{K\sigma_A^2}{\sigma_N^2}}} \exp\left[ \frac{\left( \sum_{k=1}^K y_k \right)^2}{2\sigma_N^2 \left( K + \frac{\sigma_N^2}{\sigma_A^2} \right)} \right] \underset{H_0}{\overset{H_1}{\gtrless}} \eta. \tag{2.77}$$

The likelihood ratio in (2.77) depends only on the observations and known parameters.

Taking the natural logarithm of (2.77) gives

$$\left( \sum_{k=1}^K y_k \right)^2 \underset{H_0}{\overset{H_1}{\gtrless}} 2\sigma_N^2 \left[ K + \frac{\sigma_N^2}{\sigma_A^2} \right] \ln \left( \eta \sqrt{1 + \frac{K\sigma_A^2}{\sigma_N^2}} \right) \tag{2.78}$$

or

$$\left| \sum_{k=1}^K y_k \right| \underset{H_0}{\overset{H_1}{\gtrless}} \tau \tag{2.79}$$

where  $\tau = \left( 2\sigma_N^2 \left[ K + \frac{\sigma_N^2}{\sigma_A^2} \right] \log \left( \eta \sqrt{1 + \frac{K\sigma_A^2}{\sigma_N^2}} \right) \right)^{1/2}$ . The decision statistics  $p_{fa}$  and  $p_d$  are found respectively as

$$p_{fa} = p \left( \left| \sum_{k=1}^K y_k \right| > \tau | H_0 \right) = 2Q_{\text{func}} \left( \frac{\tau}{\sqrt{K} \sigma_N^2} \right) \quad (2.80)$$

and

$$p_d = Q_{\text{func}} \left( \frac{KA + \tau}{\sigma_N \sqrt{K}} \right) + Q_{\text{func}} \left( \frac{\tau - KA}{\sigma_N \sqrt{K}} \right), \quad (2.81)$$

where  $\tau = \sqrt{K} \sigma_N Q_{\text{func}}^{-1} \left( \frac{p_{fa}}{2} \right)$ .

### 2.3.2. Generalized Likelihood Ratio Test

The GLRT uses the ML estimates (MLEs) to replace the unknown parameters. GLRT does not consider an optimal decision rule but it is used widely because it has a less restrictive assumption and is easy to implement. GLRT does not treat unknown parameters as random variables but it estimates them using MLE. The LRT is given as:

$$\frac{p(\mathbf{y}|H_1, \hat{\boldsymbol{\theta}}_1)}{p(\mathbf{y}|H_0, \hat{\boldsymbol{\theta}}_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta \quad (2.82)$$

where  $\hat{\boldsymbol{\theta}}_j$  represents the MLE of  $\boldsymbol{\theta}_j$  and then  $p(\mathbf{y}|H_j, \hat{\boldsymbol{\theta}}_j) = \underset{\boldsymbol{\theta}_j}{\text{argmax}} p(\mathbf{y}|H_j, \boldsymbol{\theta}_j)$ ,  $j = 0, 1$ .

We can write the GLRT also in the following form:

$$\frac{\underset{\boldsymbol{\theta}_1}{\text{argmax}} p(\mathbf{y}|H_1, \boldsymbol{\theta}_1)}{\underset{\boldsymbol{\theta}_0}{\text{argmax}} p(\mathbf{y}|H_0, \boldsymbol{\theta}_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta. \quad (2.83)$$

This detector is not used in the thesis.

### 2.4. Energy Detector

In this section, we will illustrate the energy detector for a binary hypothesis testing problem consisting of a group of  $K$  sensors and one FC which cooperate to detect the existence of a point source. The hypothesis testing at each sensor node can be described as

$$H_0 : y_k = \epsilon_k,$$

versus (2.84)

$$H_1 : y_k = s_k + \epsilon_k,$$

where  $y_k$  denotes the observation at the  $k$ th sensor and  $\epsilon_k$  denotes AWGN with variance  $\sigma_N^2$  and zero mean.  $s_k$  denotes the received signal which obeys a Gaussian distribution with zero mean and variance  $\sigma_s^2$ , i.e.  $y_k \sim \mathcal{N}(0, \sigma_N^2)$  under  $H_0$  and  $y_k \sim \mathcal{N}(0, \sigma_s^2 + \sigma_N^2)$  under  $H_1$ , where  $\{y_k, k = 1, 2, \dots, K\}$  are i.i.d.. Using the NP detector the decision will be in favor of hypothesis  $H_1$  if

$$\Lambda(\mathbf{y}) = \frac{\frac{1}{(2\pi(\sigma_s^2 + \sigma_N^2))^{\frac{K}{2}}} \exp\left(-\frac{\sum_{k=1}^K y_k}{2(\sigma_s^2 + \sigma_N^2)}\right)}{\frac{1}{(2\pi\sigma_N^2)^{\frac{K}{2}}} \exp\left(-\frac{\sum_{k=1}^K y_k}{2\sigma_N^2}\right)} \underset{H_0}{\overset{H_1}{\geq}} \eta, \quad (2.85)$$

or equivalently

$$\Lambda(\mathbf{y}) = \left(\frac{\sigma_N^2}{\sigma_s^2 + \sigma_N^2}\right)^{\frac{K}{2}} \exp\left(\frac{1}{2} \left(\sum_{k=1}^K y_k^2\right) \left[\frac{1}{\sigma_N^2} - \frac{1}{\sigma_s^2 + \sigma_N^2}\right]\right) \underset{H_0}{\overset{H_1}{\geq}} \eta, \quad (2.86)$$

which simplifies to

$$\exp\left(\frac{1}{2} \left(\sum_{k=1}^K y_k^2\right) \left[\frac{1}{\sigma_N^2} - \frac{1}{\sigma_s^2 + \sigma_N^2}\right]\right) \underset{H_0}{\overset{H_1}{\geq}} \eta \left(\frac{\sigma_s^2 + \sigma_N^2}{\sigma_N^2}\right)^{\frac{K}{2}}. \quad (2.87)$$

By taking the natural logarithm of (2.87)

$$\sum_{k=1}^K y_k^2 \underset{H_0}{\overset{H_1}{\geq}} \tilde{\eta}, \quad (2.88)$$

where

$$\tilde{\eta} = 2\sigma_N^2 \left[1 + \frac{\sigma_N^2}{\sigma_s^2}\right] \log\left(\eta \left[1 + \frac{\sigma_s^2}{\sigma_N^2}\right]^{\frac{K}{2}}\right). \quad (2.89)$$

Then, we find the probability of false alarm  $p_{fa}$  as

$$p_{fa} = p\left\{\sum_{k=1}^K y_k^2 > \tilde{\eta} | H_0\right\}. \quad (2.90)$$

We can normalize the Gaussian random variable  $y_k$  in (2.90) by dividing both sides of the inequality over  $\sigma_N^2$  to get a sum of squared normalized Gaussian random variable,  $\sum_{k=1}^K \left(\frac{y_k}{\sigma_N}\right)^2$  which represents a chi-square distribution with  $K$  degrees of freedom.

$$p_{\text{fa}} = p \left\{ \sum_{k=1}^K \left(\frac{y_k}{\sigma_N}\right)^2 > \frac{\tilde{\eta}}{\sigma_N^2} | H_0 \right\}. \quad (2.91)$$

Then  $p_{\text{fa}}$  and  $p_d$  can be written in terms of the incomplete Gamma function ( $\Gamma_{\text{inc}}(\cdot)$ ) respectively as

$$p_{\text{fa}} = 1 - \Gamma_{\text{inc}} \left( \frac{\tilde{\eta}}{2\sigma^2}, \frac{K}{2} \right) \quad (2.92)$$

and

$$p_d = 1 - \Gamma_{\text{inc}} \left( \frac{\tilde{\eta}}{2(\sigma_N^2 + \sigma_s^2)}, \frac{K}{2} \right). \quad (2.93)$$

The relationship between (2.92) and (2.93) can be represented as a ROC curve, as shown in Figure 2.9. The black curve is obtained at  $\text{SNR} = \frac{\sigma_s^2}{\sigma_N^2} = 1$  with  $K = 3$  observations. For the same number of observations, we can extend the region of achievable  $(p_{\text{fa}}, p_d)$  pairs by increasing the value of SNR to  $\frac{\sigma_s^2}{\sigma_N^2} = 4$  as shown in the red curve. We can see also that the ROC performance can be increased by increasing the number of observations to  $K = 6$  as shown by the blue curve.

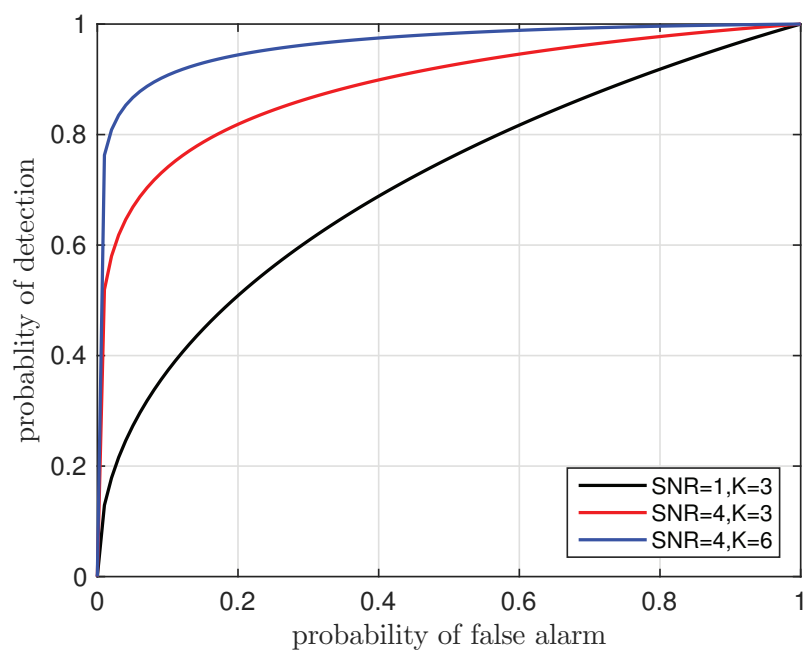


Figure 2.9. ROC curve for the energy detector

## CHAPTER 3

# MAXIMUM AVERAGE ENTROPY-BASED QUANTIZATION OF LOCAL OBSERVATIONS FOR DISTRIBUTED DETECTION

### 3.1. Introduction

In WSNs, there are strict energy and bandwidth restrictions. Because of this, observations of the sensors are frequently needed to be quantized before transmitting them to a FC in an effort to find a compromise between low power consumption of the wireless sensors and good global detection performance. So, in this chapter, we will consider the problem of how the sensor outputs should be quantized.

Optimum quantization levels in the sense of information theoretic criteria for DD systems were presented in (Poor and Thomas, 1977; Poor, 1988; Lee and Chao, 1989; Warren and Willett, 1999). In (Poor and Thomas, 1977), the quantization based on Ali-Silvey distances between two simple hypotheses were investigated. After that, in (Poor, 1983; Poor, 1988), the divergence was proposed as a distortion measure by considering a class of  $f$ -divergence measures which shows that the loss in divergence is quadratic with the quantization step size. In (Lee and Chao, 1989; Warren and Willett, 1999), the authors considered that each local detector transmits a multiple-bit decision to the FC. The solution for partitioning the local decision space was derived by maximizing the distance between the mean values of the quantized hypotheses. It was shown that the global decision performance increases monotonically by increasing the number of partitions at the individual detector. This method is locally optimum in the sense of JD, but it does not necessarily yield a globally optimum solution. Even when only four quantization levels are considered, the solution is given by complicated analytic expressions explaining the functional relationships between the detection probability and the false alarm probability of all detectors and their derivatives. In those works, it was assumed that all local sensors

are identical NP detectors observing the same SNR.

In (Altay and Delic, 2016), in order to perform optimum quantization in the sense of mean-error, deflection criterion (DC) and Chernoff information (CI) was defined for DD systems consisting of one FC and multiple sensors by using Bayesian detection criterion. DC and CI pose a nonlinear and non-convex problem, which mostly has more than one extreme. These kinds of optimization criteria are suitable for the case of known SNR where the probability of detection and probability of false alarm is known for each local detector.

Inspiring from quantization of signals using the Maximum Output Entropy (MOE) in (Messerschmitt, 1971), we propose an entropy based method by maximizing the average entropy of observations under both hypotheses to determine the quantization intervals at distributed sensors in order to optimize the global binary decision at the FC about the existence of a point source under the NP criterion where sensors observe different signal levels which they do not know. Although maximizing the entropy is a well-known approach, it has not been used in decision problems until now to the best of our knowledge. The most probable reason for this is the widespread acceptance that an information theoretic criterion for decision problems should concentrate on the distance of rival hypotheses. We consider scenarios with non-equally important hypotheses, that is why NP criterion is considered to be more suitable compared to the probability of error criterion, in this thesis.

## 3.2. System Model

In this chapter, a binary hypothesis testing problem is considered, where a group of  $K$  sensors and one FC cooperate to detect the existence of a point source as shown in Figure 3.1. The hypothesis testing at each sensor node can be described as

$$\begin{aligned}
 &H_0 : y_k = \epsilon_k, \\
 &\text{versus} \\
 &H_1 : y_k = A_k + \epsilon_k,
 \end{aligned} \tag{3.1}$$

where  $y_k$  denotes the observation at the  $k$ th sensor and  $\epsilon_k$  denotes AWGN with variance  $\sigma_N^2$  and zero mean.  $A_k$  denotes the received signal amplitude which is equal to  $\alpha_k A_{\max}$ . Each



sensor in the range of the point source detects a signal attenuated with a factor of  $\alpha_k$  and makes a local decision  $u_k$ . The local decision is transmitted through the multiplicative channel  $h_k$  to the FC where the final decision  $u_0$  is made. In Figure 3.1, the sensor outputs  $\{\bar{u}_k, k = 1, 2, \dots, K\}$ , the AWGNs in the channel from the sensors to the FC  $\{\xi_k, k = 1, 2, \dots, K\}$  and the received signals  $\{\bar{y}_k, k = 1, 2, \dots, K\}$  are shown as vectors in accordance with the  $M$ -dimensional signal model of FSK related modulated signal model, explained in detail in section 3.2.2.1.

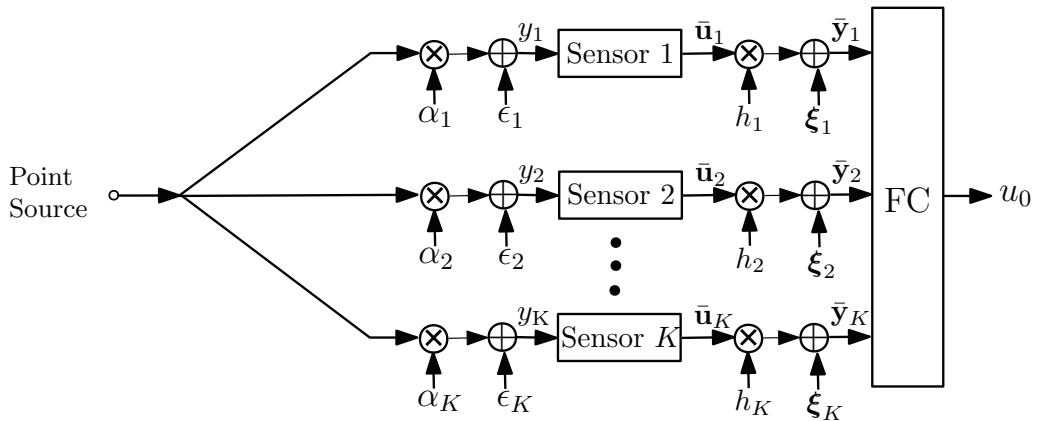


Figure 3.1. Parallel distributed detection system

Based on the dispersion pattern over the surveillance zone and the physical characteristics, the phenomenon to be detected can be modeled either as a field source or a point source. A field source is dispersed over the sensor field such as in temperature monitoring. On the other hand, the event is generated by a single point source such as in target detection and fire detection.

### 3.2.1. Point Source

In this chapter, we consider a point event source emitting constant power uniformly in all directions. For such a source the signal amplitude received by a sensor will be inversely proportional to the distance from the source. Considering uniformly deployed sensors, only those sensors which are within a circle the radius of which is determined by the sensitivity of the sensors will receive a signal.

Let  $A_{\max}$  denote the signal amplitude on a circle with radius  $r_{\min}$  centered by the event location as shown in Figure 3.2. We assume that  $A_{\max}$  corresponds to the maximum detectable signal level or the saturation level of the sensors and  $A_{\min}$  denotes the minimum value of the detectable signal observed at a distance of  $r_{\max}$  from the event location. This yields a different and unknown amplitude value for each individual sensor. Assuming there are no sensors in the small circle, the pdf of the normalized signal amplitude,  $A_n = A/A_{\max}$ , at a sensor will have the form shown in Figure 3.3 and will be given as:

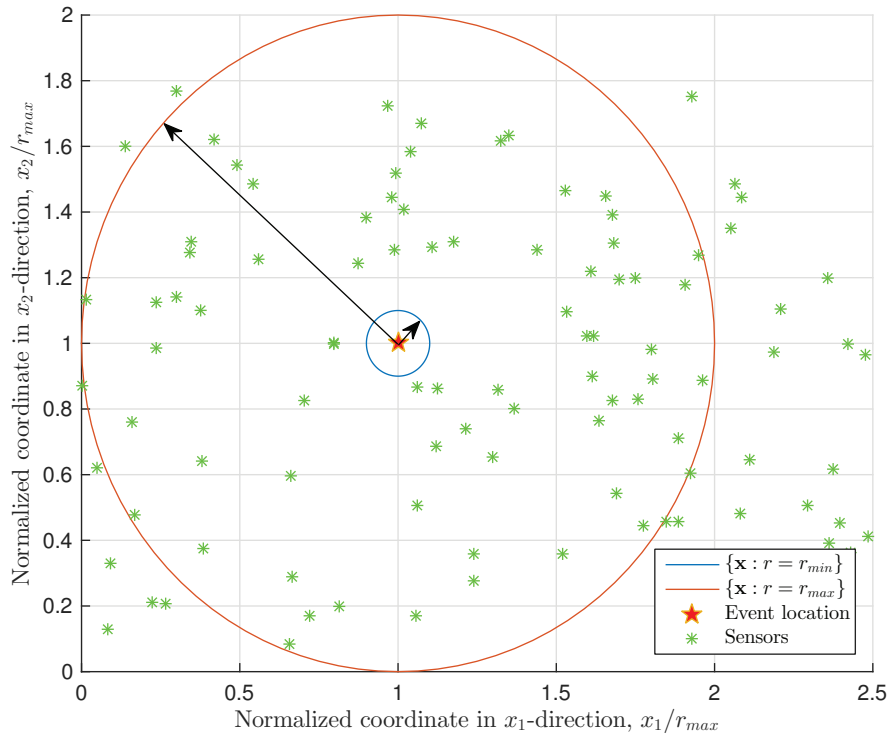


Figure 3.2. Positions of the event location and uniformly distributed sensors in a scenario for detecting a point source.

$$p(A_n) = \frac{1}{A_n \log(L)} \quad (3.2)$$

where  $L = A_{\max}/A_{\min}$ . We define the SNR as the ratio between the maximum signal power,  $A_{\max}^2$ , and the noise power,  $\sigma^2$ . Let us assume that  $K$  of the sensors uniformly deployed in the area will be in the fat ring (or punctured disk) described by the radii  $r_{\min}$

and  $r_{\max}$ . Then, the signal amplitudes at these sensors will be independent and come from the pdf given in (3.2) in the case of an event. Assuming that the sensor observations are available distortion-free at the FC, i.e. without transmission over a wireless channel, the optimal Bayesian NP detector can be written as:

$$\Lambda(\mathbf{y}) = \frac{\prod_{k=1}^K \int_{A_{\max}/L}^{A_{\max}} p(y_k|H_1; A_k)p(A_k)dA_k}{p(\mathbf{y}|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta. \quad (3.3)$$

Since each  $A_k$  comes from the independent and identical pdf given in (3.2), we eliminate the index,  $k$ , and express the likelihood ratio as

$$\Lambda(\mathbf{y}) = \frac{\prod_{k=1}^K \int_{A_{\max}/L}^{A_{\max}} \frac{1}{\sqrt{2\pi\sigma_N^2}} \exp\left(\frac{-(y_k - A)^2}{2\sigma_N^2}\right) \frac{1}{A \log(L)} dA}{\left(\frac{1}{\sqrt{2\pi\sigma_N^2}}\right)^K \exp\left(\frac{-\sum_{k=1}^K (y_k)^2}{2\sigma_N^2}\right)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta, \quad (3.4)$$

where  $\mathbf{y} = [y_1, y_2, \dots, y_K]^T$  denotes vector of observations from  $K$  sensors.

### 3.2.2. Fusion System: Channel Between Sensors and FC

In this section, we will investigate the complete model of the sensor to FC communication by using a Rayleigh fading channel model and an  $M$ -ary FSK modulation scheme where each one of the  $M$  different symbols is transmitted by a carrier wave of a different frequency.  $M$ -FSK is a suitable modulation scheme for low-power low data rate transmission as preferred by the majority of the sensor device equipment. In (Hajibabaei and Vosoughi, 2014), the error probability when the detectors perform FSK modulation was minimized when training symbol transmit power is zero. Accordingly, the non-coherent demodulation of  $M$ -FSK was adopted in this thesis. Additionally, in order to concentrate on the fusion of sensor data with non-identical signal levels, we considered the case of no channel, i.e. when error-free sensor outputs are available at the FC, which we called as direct data transmission (DDT). Once data from the sensors are at the FC, an equal gain fusion rule is applied for every different type of sensor transmissions to FC since the relative reliability of sensor outputs are not evaluated.

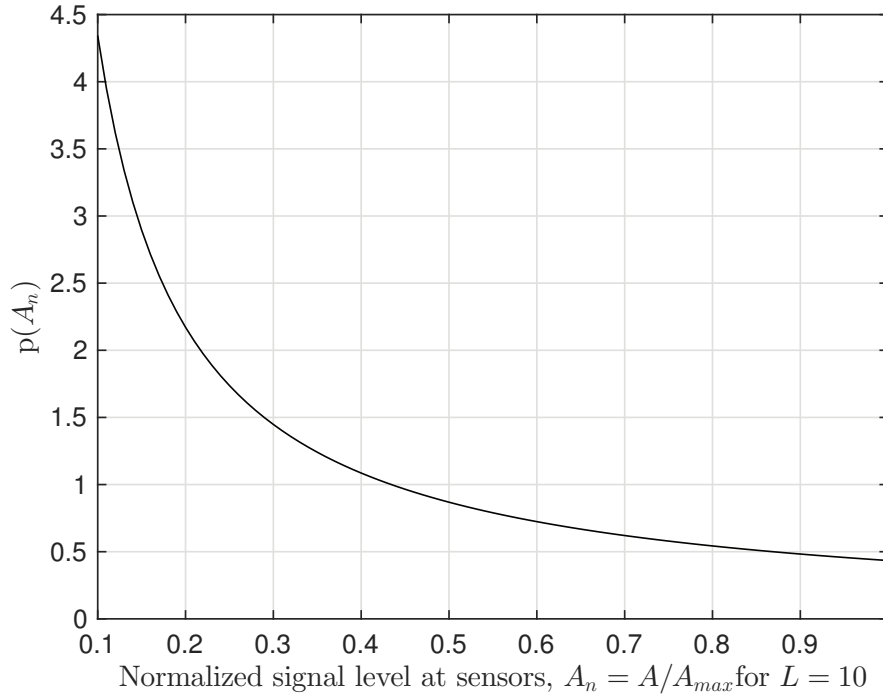


Figure 3.3. The pdf of the signal amplitude observed at the sensors,  $p(A_n)$ .

### 3.2.2.1. Fading Channel

In this subsection, the problem of fusing the data transmitted over a fading channel is considered, as shown in Figure 3.1. The FC has only information on the channel statistics. non-coherent  $M$ -FSK modulation is employed for transmitting data to the FC. Let  $\mathbf{u}_k$  denote the  $M$ -FSK modulated symbol at sensor  $k$ , where  $\mathbf{u}_k \in \{\mathbf{e}_m, m = 1, \dots, M\}$  and  $\mathbf{e}_m$  is an  $M \times 1$  column vector, all elements of which except the  $m$ th one are zero. We refer to the transmit power of the data symbol as  $P_d$ . Assuming  $M$ -dimensional signal model for representing the orthogonal channels of  $M$ -FSK modulation scheme between the detectors and the FC (Hajibabaei and Vosoughi, 2014) simplifies the analysis. Then, the output of the channel which is corresponding to detector  $k$  at the FC can be given as:

$$\begin{aligned} \bar{\mathbf{y}}_k &= \sqrt{P_k} h_k \mathbf{u}_k + \boldsymbol{\xi}_k \\ &= h_k \bar{\mathbf{u}}_k + \boldsymbol{\xi}_k \end{aligned} \quad (3.5)$$

where  $P_k$  represents the received power which is a function of  $P_d$ , the wavelength, the path

loss exponent and the distance between detector  $k$  and the FC (Hajibabaei and Vosoughi, 2014) and it describes the effect of large scale fading. The channel noise is denoted as  $\mathbf{n}_k$  which is a zero mean complex Gaussian vector  $\boldsymbol{\xi}_k \sim \mathcal{CN}(\mathbf{0}, \sigma_n^2 \mathbf{I}_M)$ , where  $\mathbf{I}_M$  is an  $M \times M$  identity matrix. The complex channel coefficient  $h_k$  in (3.5) is modeled as  $h_k \sim \mathcal{CN}(0, 1)$  which can be also represented as  $h_k = \alpha_k e^{j\phi_k}$ , where  $\alpha_k$  represents the amplitude with Rayleigh distribution and  $\phi_k$  represents the phase with uniform distribution. We adopt NP criterion to find the optimal and a sub-optimal fusion rule at the FC in order to obtain a global decision  $u_0 \in \{H_0, H_1\}$  as follows.

I. The optimal fusion rule for the i.i.d. vectors,  $\bar{\mathbf{y}}_k, k = 1, 2, \dots, K$ , is defined as follows:

$$\log(\Lambda(\mathbf{Y})) = \log\left(\frac{p(\mathbf{y}|H_1)}{p(\mathbf{y}|H_0)}\right) = \log\left(\prod_{k=1}^K \frac{p(\bar{\mathbf{y}}_k|H_1)}{p(\bar{\mathbf{y}}_k|H_0)}\right) \underset{H_0}{\overset{H_1}{\geq}} \eta, \quad (3.6)$$

where  $\mathbf{Y}$  is the matrix composed by row-wise stacking column vectors  $\bar{\mathbf{y}}_k, k = 1, 2, \dots, K$ .

Expanding  $p(\bar{\mathbf{y}}_k|H_1)$  and  $p(\bar{\mathbf{y}}_k|H_0)$  in (3.6) over the  $M$ -level sensor decisions we obtain

$$\log(\Lambda(\mathbf{Y})) = \sum_{k=1}^K \log\left(\frac{\sum_{m=1}^M p(\bar{\mathbf{y}}_k|\mathbf{u}_k(m))p(\mathbf{u}_k(m)|H_1)}{\sum_{m=1}^M p(\bar{\mathbf{y}}_k|\mathbf{u}_k(m))p(\mathbf{u}_k(m)|H_0)}\right) \underset{H_0}{\overset{H_1}{\geq}} \eta, \quad (3.7)$$

where  $K$  represents the number of sensors and  $M$  represents the number of quantization levels at each local sensor. The conditional density  $p(\bar{\mathbf{y}}_k|\mathbf{u}_k(m))$  in (3.7) is a complex multi-variate Gaussian density,  $\bar{\mathbf{y}}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{C}_{\bar{\mathbf{y}}})$ ,  $\mathbf{C}_{\bar{\mathbf{y}}}$  represents the diagonal matrix with entries  $\mathbf{C}_{\bar{\mathbf{y}}}(j, j) = \sigma_n^2$  for  $j \neq m$  and  $\mathbf{C}_{\bar{\mathbf{y}}}(j, j) = P_k \sigma_h^2 + \sigma_n^2$  for  $j = m$ , where  $j = 1, \dots, M$ . We can prove that  $p(\bar{\mathbf{y}}_k|\mathbf{u}_k(m))$  equals to

$$\frac{1}{\sqrt{\pi^M \det |\mathbf{C}_{\bar{\mathbf{y}}_m}|}} \exp\left\{-\left(\bar{\mathbf{y}}_k - \boldsymbol{\mu}\right)^H \mathbf{C}_{\bar{\mathbf{y}}_m}^{-1} \left(\bar{\mathbf{y}}_k - \boldsymbol{\mu}\right)\right\}. \quad (3.8)$$

The values of  $p(\mathbf{u}_k(m)|H_1)$  represent the probability masses under hypothesis  $H_1$ , which are estimated as:

$$\overline{p_m^{H_1}} = \int_{A_{\max/L}}^{A_{\max}} p_m^{H_1}(A_n) p(A_n) dA_n, \quad (3.9)$$

where  $p_m^{H_1}(A_n)$  represents the probability mass under  $H_1$  as shown in Figure 3.4 for an observed signal level  $A_n$  which is the mean of the Gaussian signal.

The values of  $p(\mathbf{u}_k(m)|H_0)$  represent the probability masses under hypothesis  $H_0$ :

$$p(\mathbf{u}_k(m)|H_0) = p_m^{H_0}. \quad (3.10)$$

Figure 3.4 shows a possible partitioning of a pdf under hypothesis  $H_i$ ,  $i = 0, 1$  and the probability masses for  $M = 4$  corresponding to the areas under the pdf between successive thresholds.

- II. A sub-optimal fusion rule can be derived as follows: In (3.7) we see both the effects of fading channel and the local detection outputs in order to achieve the optimal performance. A direct alternate could be used as a sub-optimal fusion rule by separating this into two-steps. First,  $\bar{y}_k$  is used to infer about the local detector by applying the maximum likelihood (ML) estimate as an intermediate decision,  $\hat{u}_k$ , and then, the optimum fusion rule based on  $\hat{u}_k$  is applied:

$$\hat{u}_k = \arg \max_m \boldsymbol{\theta}_m, \quad (3.11)$$

where  $\boldsymbol{\theta}_m$  is given as

$$\boldsymbol{\theta}_m = p(\bar{\mathbf{y}}_k|\mathbf{u}_k(m)). \quad (3.12)$$

We can re-write (3.8) as in (Hajibabaei and Vosoughi, 2014)

$$p(\bar{\mathbf{y}}_k|\mathbf{u}_k(m)) = \frac{1}{\sqrt{\pi^M \det |\mathbf{C}_{\bar{\mathbf{y}}_m}|}} \exp\left(\frac{P_k \sigma_h^2 |\bar{y}_k(m)|^2}{\sigma_n^2 (\sigma_h^2 + \sigma_n^2)}\right) \prod_{j=1}^M \exp\left(\frac{|\bar{y}_k(j)|^2}{\sigma_n^2}\right). \quad (3.13)$$

By substituting (3.13) in (3.11) after eliminating the terms which are irrelevant to  $m$ , we can re-write (3.11) as

$$\hat{u}_k = \arg \max_m \exp\left(\frac{P_k \sigma_h^2 |\bar{y}_k(m)|^2}{\sigma_n^2 (\sigma_h^2 + \sigma_n^2)}\right), \quad (3.14)$$

where  $m = 1, \dots, M$ . Note that  $|\bar{y}_k(m)|^2$  in (3.14) denotes the squared envelopes of  $M$  cross-correlators corresponding to non-coherent FSK detection.

The final decision rule is given as

$$u_0 = \sum_{k=1}^K \hat{u}_k \underset{H_0}{\overset{H_1}{\gtrless}} \eta. \quad (3.15)$$

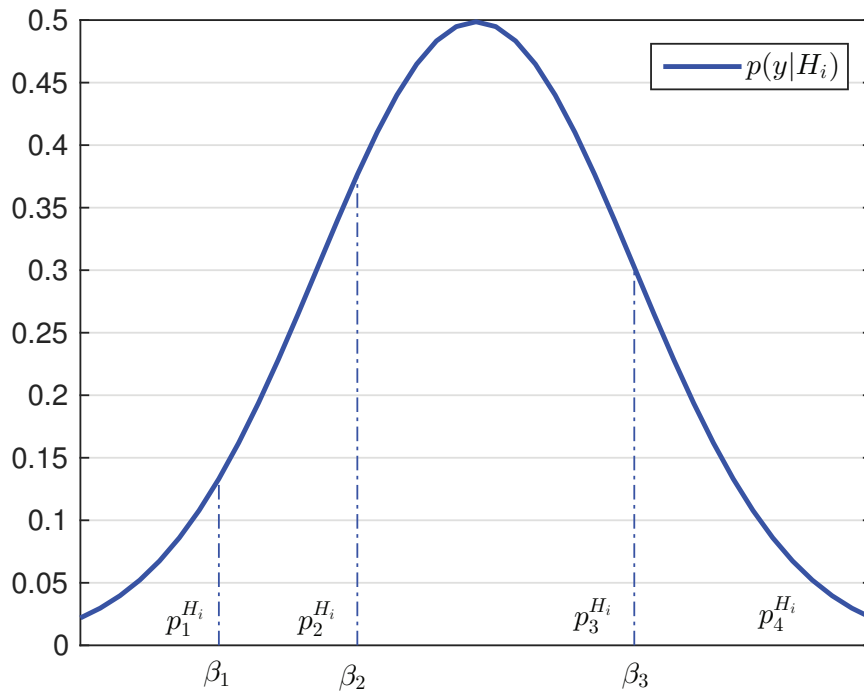


Figure 3.4. A partitioning of the pdf for the observations at each sensor for 4-level quantization.

### 3.3. Quantizer Design

It is aimed to make a global decision at the FC under the NP criterion. Let us assume that each sensor will only make a single observation and will transmit this observation to the FC. Then, sensors will make i.i.d. observations under  $H_0$  and none of the sensors can estimate the signal level under  $H_1$ . Consequently, there is no information at

the sensors in order to use different quantization thresholds under  $H_1$ . So, it is reasonable to use identical quantization thresholds at each sensor irrespective of their distance to the event location since it cannot be estimated. Definitely, the choice of the quantization thresholds affects the performance which makes it desirable to choose the quantization thresholds which maximize the system performance. This chapter proposes the maximum average entropy (MAE) method, that is, determining the quantization thresholds at the sensors in a way to maximize the average entropy of the discrete information collected at the FC under both hypotheses without considering the effects of the wireless channels between sensors and FC. To the best of our knowledge, all of the entropy based quantizers for detection problems are some distance measures (Poor and Thomas, 1977; Altay and Delic, 2016). The proposed MAE method differs from them in that it maximizes the transmitted information corresponding to both of the underlying probability mass functions (pmfs) jointly.

The optimum detector at the FC is based on likelihood ratios as given in (3.4). Equivalently, one can use log-likelihood (logarithm of likelihood) ratios (LLRs) and the LLR for the  $k$ th sensor with an unknown signal amplitude can be calculated by using the expected value of the signal amplitude,  $\bar{A}$ , as follows:

$$\Lambda(y_k) = -\frac{\bar{A}^2}{2\sigma_N^2} + \frac{\bar{A}}{\sigma_N^2}y_k. \quad (3.16)$$

The linear (or more appropriately affine) transformation of observations in (3.16) to LLRs is irrelevant in entropy based quantization because that kind of transformation only results in translation and scaling of the underlying pdfs and will preserve the resulting probability masses corresponding to a vector of thresholds (such as  $\beta_1$ ,  $\beta_2$  and  $\beta_3$  in Figure 3.4). Consequently, the sensors will transmit a quantized observation signal to the FC. A common information based criterion for determining the quantization thresholds is the maximum JD (MJD) method which was used in the case of the constant signal level at sensors formerly (Lee and Chao, 1989). We will first explain these criteria and subsequently the relation between them. Two toy examples in Appendix C are utilized for a better understanding of the proposed method MAE and the corresponding MJD from the previous work. The quantization thresholds are found using these methods and the corresponding ROC curves are obtained.



### 3.3.1. Maximum Average Entropy Method

An intuitive idea to have optimum performance at the FC is to maximize the entropy under both hypotheses which we call as MAE method. So, we propose to determine the quantization intervals at the sensors as resulting in MAE under both hypotheses. The entropy of a quantized sensor output can be calculated based on the partitioning of the observation pdf at each sensor as shown in Figure 3.4. In this figure, the number of quantization intervals is 4. For a general number of  $M$  quantization intervals, there will be  $M - 1$  thresholds,  $\{\beta_1, \beta_2, \dots, \beta_{M-1}\}$ , and  $M$  partitions with corresponding probability masses of observations  $\{p_1^{H_i}, p_2^{H_i}, \dots, p_M^{H_i}\}$  where  $i = 0, 1$ . Under  $H_i$ , the entropy of the observation can be estimated as

$$\hat{F}_{H_i} = E \left( - \sum_{m=1}^M p_m^{H_i} \log_2(p_m^{H_i}) \right) \text{ bit.} \quad (3.17)$$

The expectation,  $E(\cdot)$ , is with respect to (w.r.t.) the distribution of the  $K$  sensors and in the special case of the scenario described in Figure 3.2, this distribution is uniform in the sensing range of the sensors defined by a circle within a radius of  $r_{\max}$  from the event location.  $\mathbf{p}_M^{H_i} = [p_1^{H_i}, p_2^{H_i}, \dots, p_M^{H_i}]$  denotes the vector of these probability masses, i.e. the probabilities of the partitions. In practice, an estimate of this expectation is obtained by averaging the information of the sensors over the distribution of the sensor locations and AWGN realizations which is called a histogram method (Messerschmitt, 1971). Figure 3.5 shows the entropy function  $\hat{F}_{H_0}$ ,  $\hat{F}_{H_1}$  and  $\hat{F}_{\text{av}} = \frac{1}{2} (\hat{F}_{H_0} + \hat{F}_{H_1})$  for binary quantization. For  $M$ -ary quantization,  $\boldsymbol{\beta}_M^* = [\beta_1^*, \beta_2^*, \dots, \beta_{M-1}^*]$  denotes the vector of optimum quantization thresholds in the sense of MAE which is found as

$$\boldsymbol{\beta}_M^* = \arg \max_{\boldsymbol{\beta}_M} \hat{F}_{\text{av}}. \quad (3.18)$$

The main part of the proposed MAE algorithm is given in Algorithm 1. Optimal quantization thresholds for binary quantization corresponds to the maximum of  $\hat{F}_{\text{av}}$  as shown in Figure 3.5 which is  $\beta_2^* = 0.093$ . In a similar way, we can estimate the optimal thresholds for 3-level quantization to be  $\boldsymbol{\beta}_3^* = [-0.341 \quad 0.528]$  as shown in Figure 3.6 in terms of equal level contours. Similarly, the optimum thresholds are  $\boldsymbol{\beta}_4^* = [-0.367 \quad 0.195 \quad 0.835]$

in the case of 4-level quantization and  $\beta_6^* = [-1.08 \quad -0.572 \quad -0.060 \quad 0.4513 \quad 0.963]$  for 6-level quantization. The given optimal quantization thresholds are found for SNR = 0 dB.

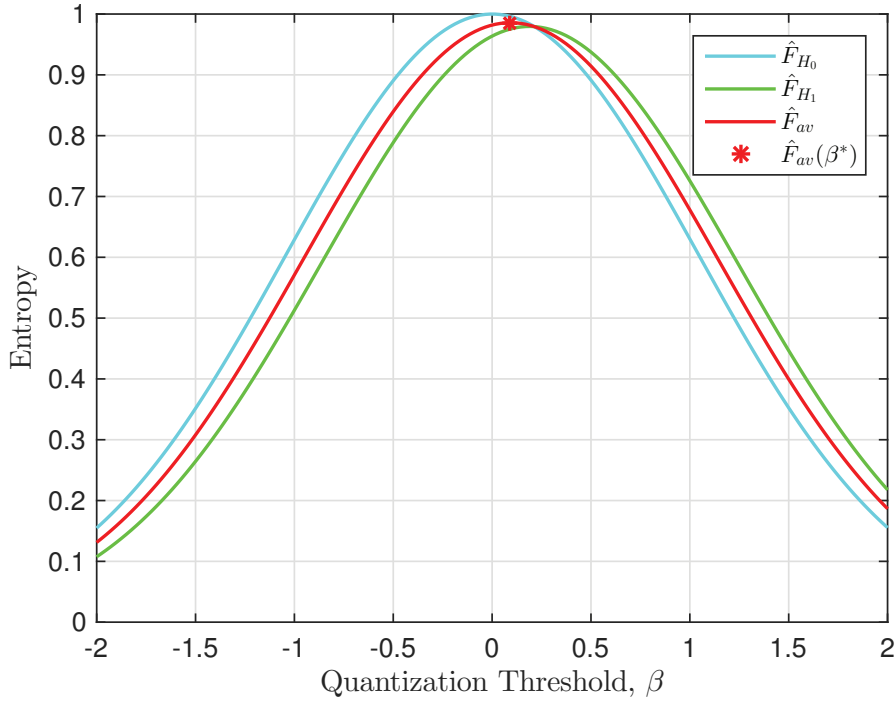


Figure 3.5. The entropy functions  $\hat{F}_{H_0}$ ,  $\hat{F}_{H_1}$  and  $\hat{F}_{av}$  for binary quantization.

### 3.3.2. Maximum J-Divergence Method

JD can be written in terms of the relative entropy for discrete probability distributions  $P$  and  $Q$  observed under the two hypotheses  $H_0$  and  $H_1$ , respectively, as follows:

$$J = D_{\text{KL}}(P||Q) + D_{\text{KL}}(Q||P), \quad (3.19)$$

where the relative entropy between two pmfs  $P(x)$  and  $Q(x)$  is given as follows:

$$D_{\text{KL}}(P||Q) = \sum_{x \in \mathcal{X}} P(x) \log_2 \left( \frac{P(x)}{Q(x)} \right), \quad (3.20)$$

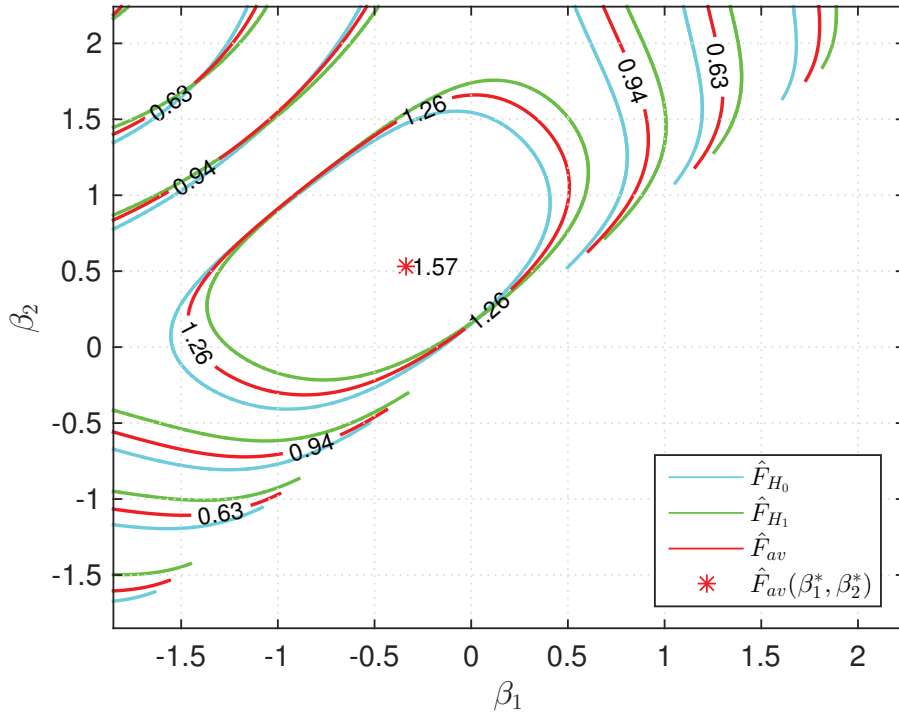


Figure 3.6. The entropy functions  $\hat{F}_{H_0}$ ,  $\hat{F}_{H_1}$  and  $\hat{F}_{av}$  for three level quantization.

where  $\chi$  denotes the alphabet of the pmfs for  $P$  and  $Q$ . In our context, JD measures the distributional distance or dissimilarity between the distributions of the observations under two hypotheses  $H_0, H_1$  and this can be used to find the local thresholds. The choice of local thresholds facilitates the design of local detectors which in turn determines the performance of the whole system. An estimate of the expected value for the JD can be obtained by averaging the contribution to the JD over the distribution of sensor locations and noise realizations as performed for entropy of the observations in (3.17) and can be written as:

$$\hat{j} = E \left( \sum_{m=1}^M \left[ p_m^{H_1} \log_2 \left( \frac{p_m^{H_1}}{p_m^{H_0}} \right) - p_m^{H_0} \log_2 \left( \frac{p_m^{H_1}}{p_m^{H_0}} \right) \right] \right), \quad (3.21)$$

where we substitute  $\mathbf{p}_m^{H_0}$  and  $\mathbf{p}_m^{H_1}$  for pmfs  $P(x)$  and  $Q(x)$  in (3.20) that is the probability masses corresponding to the partitions of the pdf. For  $M$ -ary quantization,  $\beta_M^\diamond = [\beta_1^\diamond, \beta_2^\diamond, \dots, \beta_{M-1}^\diamond]$  denotes the JD optimized vector of quantization thresholds which can

---

**Algorithm 1** The procedure of performing MAE quantization.
 

---

```

1: Input  $A, \sigma_N^2$  ▷ Where  $A$ - signal amplitude and  $\sigma_N^2$ - noise variance
2:  $\beta_m = \beta_{low} : \delta_\beta : \beta_{high}$ 
3: Initialize the iteration  $m = 0, z = 0, x = 0$ 
4: exhausted search for  $M - 1$  thresholds such as  $\beta_{M-1} \geq \beta_{M-2} \geq \dots \geq \beta_1$ 
5: for  $z = 1$  to  $\text{length}(\beta_m)$  do ▷ estimate the entropies for all possible thresholds
6:   Step size of  $\delta_\beta = 0.01$ 
7:    $A_j = A_{low} : \delta_A : A_{high}$ 
8:   calculate the weight of each  $A_j$ 
9:    $w_j = \frac{1}{A_j \log(L)}$ 
10:  for  $x = 1$  to  $\text{length } A$  do ▷ Averaging over the histogram of A
11:    Step size of  $\delta_A = 0.01$ 
12:    calculate the probability mass functions under hypothesis  $H_1$ 
13:     $p_m^{H_1} = \int_{Area_m} p(y|H_1) dy, m = 1 : M$ 
14:    calculate the entropy under hypothesis  $H_1$ 
15:     $H_1 = w_j \times (-\sum_{m=1}^M p_m^{H_1} \log_2(p_m^{H_1}))$  bit.
16:  end for
17:   $\hat{F}_{H_1} = \sum_x(H_1)$ 
18:  calculate the probability mass functions under hypothesis  $H_0$ 
19:   $p_m^{H_0} = \int_{Area_m} p(y|H_0) dy$ 
20:  calculate the entropy under hypothesis  $H_0$ 
21:   $\hat{F}_{H_0} = -\sum_{m=1}^M p_m^{H_0} \log_2(p_m^{H_0})$  bit.
22:   $\hat{F}_{av} = \frac{1}{2}(\hat{F}_{H_0} + \hat{F}_{H_1})$ 
23: end for
24: find the optimal thresholds
25:  $\beta_M^* = \arg \max_{\{\beta_M\}} \hat{F}_{av}$ 

```

---

be given as

$$\beta_M^\circ = \arg \max_{\beta_M} \hat{J}. \quad (3.22)$$

Again, based on the pdfs at SNR = 0dB, optimization of JD is performed. Optimal quantization thresholds correspond to the maximum of  $\hat{J}$  which is found to be  $\beta_2^\circ = 0.17$  for binary quantization as shown in Figure 3.7. In a similar way, we can estimate the optimal thresholds for 3-level quantization to be  $\beta_3^\circ = [-0.444 \quad 0.784]$  as shown in Figure 3.8. Similarly, the optimal thresholds are  $\beta_4^\circ = [-0.725 \quad -0.699 \quad 0.6559]$  and  $\beta_6^\circ = [-6.19 \quad -0.572 \quad -0.0603 \quad 0.9628 \quad 6.59]$  in the cases of 4-level and 6-level quantizations, respectively.

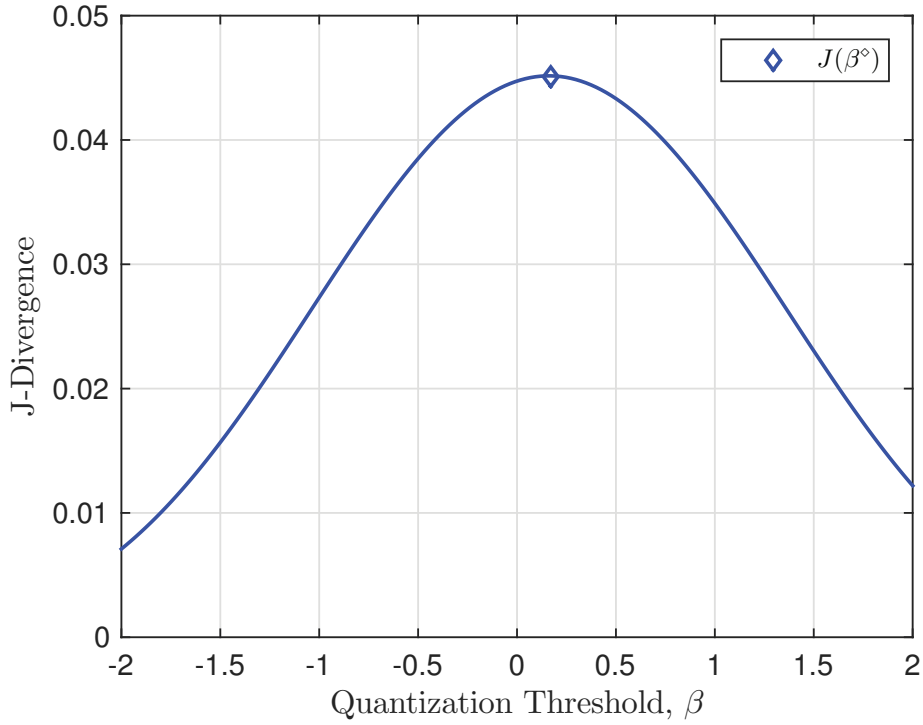


Figure 3.7. The JD for binary quantization.

### 3.3.3. Relation of MAE and MJD Methods

In this subsection, we will demonstrate that both of the information based criteria, namely, MAE and MJD, maximize similar quantities in showing that they are positively proportional. Let us first express  $D_{\text{KL}}(P||Q)$  given in (3.20) as follows:

$$D_{\text{KL}}(P||Q) = \underbrace{\sum_{x \in \mathcal{X}} P(x) \log_2 \left( \frac{1}{Q(x)} \right)}_{R_1} + \underbrace{\sum_{x \in \mathcal{X}} P(x) \log_2(P(x))}_{-F_{H_0}} \geq 0. \quad (3.23)$$

The equality holds only when  $P = Q$ . Similarly,

$$D_{\text{KL}}(Q||P) = R_2 - F_{H_1} \geq 0, \quad (3.24)$$

where  $R_2 = \sum_{x \in \mathcal{X}} Q(x) \log_2 \left( \frac{1}{P(x)} \right)$ . Substituting (3.23) and (3.24) into (3.19)

$$J = R_1 + R_2 - (F_{H_0} + F_{H_1}) \geq 0. \quad (3.25)$$

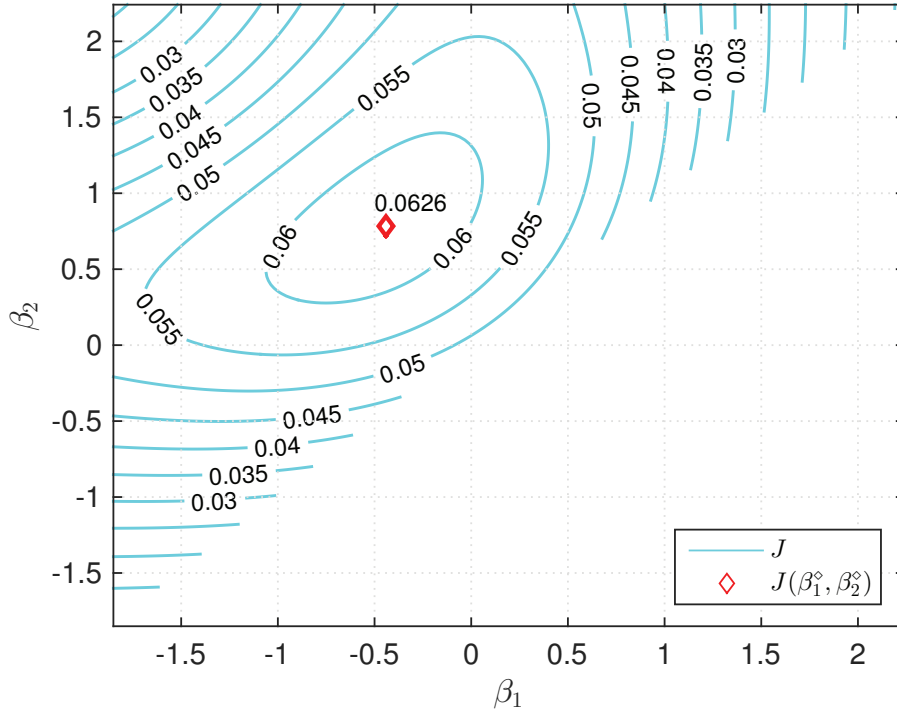


Figure 3.8. The J-divergence for three level quantization.

Then, defining  $D_{\text{KL}}(P||Q) = c_1 F_{H_0}$  and  $D_{\text{KL}}(Q||P) = c_2 F_{H_1}$ , we can re-write the JD in (3.25) to show that there is a proportionality relation between the JD and the average entropy (AE):

$$\begin{aligned}
 J &= c_1 F_{H_0} + c_2 F_{H_1} \\
 &= \min\{c_1, c_2\} \underbrace{(F_{H_0} + F_{H_1})}_{2F_{\text{av}}} + c_3
 \end{aligned} \tag{3.26}$$

with

$$c_3 = \begin{cases} (c_1 - c_2)F_{H_0} & \text{for } c_1 \geq c_2, \\ (c_2 - c_1)F_{H_1} & \text{for } c_1 \leq c_2 \end{cases}. \tag{3.27}$$

Obviously  $c_i \geq 0$  for  $i = 1, 2$  and 3. This means that AE and JD are positively proportional.

## 3.4. Simulation Results

Monte Carlo simulations have been performed in order to evaluate the detection performance for the proposed method at SNR= 0 dB for  $K = 25$  transmitting sensors and  $L = A_{\max}/A_{\min} = 10$ . First, we have performed simulations using the DDT method, that is assuming the sensor outputs are available error-free at the FC for both MAE and MJD methods. Then, a Rayleigh fading channel is considered to show the channel effect on the performance of our proposed quantization method, MAE.

### 3.4.1. Binary Direct Data Transmission

In Figure 3.9, the ROC, that is probability of detection ( $p_D$ ) versus probability of false alarm ( $p_{FA}$ ), curves are plotted for the cases of using the quantization intervals from MAE and MJD methods for the binary data transmission and the corresponding non-quantized data transmissions. The  $K$ th root quantization, which uses the  $K$ th root of the global probability of false alarm  $p_{FA} = 0.1$  to find probability mass function (pmf)  $p_2^{H_0} = 0.89$  at each sensor, is also provided for the comparison with the proposed method, MAE.  $K$ th root method corresponds to setting the false alarm threshold at the FC to a single "one" coming from any of  $K$  sensors. In this figure, we observe a slightly better performance of the MAE-based method compared to the MJD-based one. Each of them performs much better compared to the trivial  $K$ th root method which is supplied as an obvious lower bound. Additionally, we observe that they are clearly inferior to the non-quantized case which shows that there is quite a large space for gain in using higher levels of quantization.

### 3.4.2. Performance of MAE and MJD with Multilevel Quantization

The simulation performances for the three-level, four-level and six-level quantizations by using the MAE and MJD methods are also obtained for DDT.

ROC curves obtained using MAE and MJD methods for three levels of quantiza-

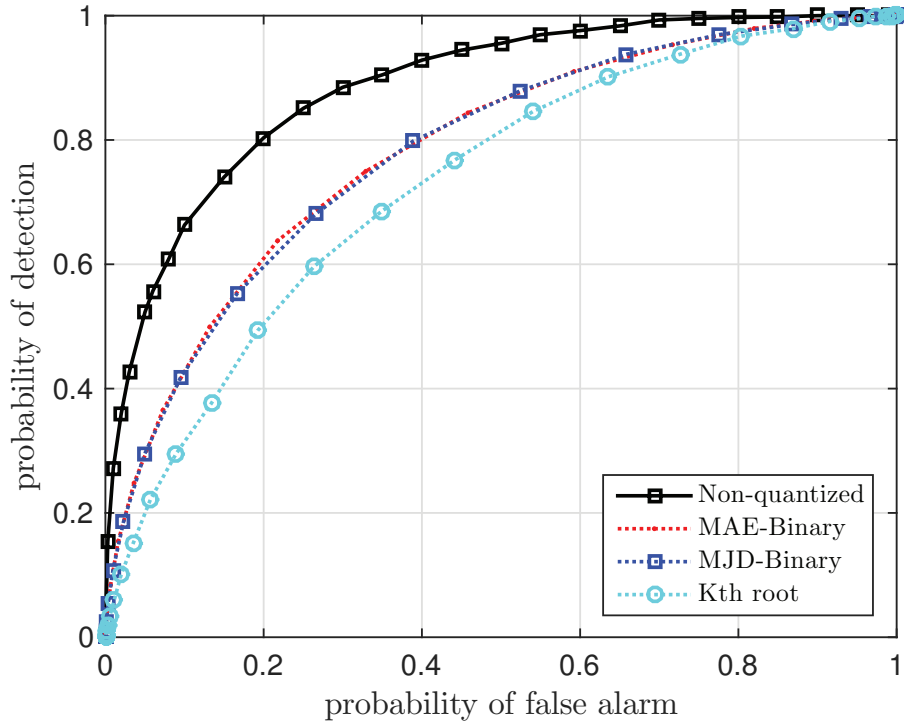


Figure 3.9. Comparison between the ROC curves obtained using MAE, MJD and Kth root methods for binary and Gaussian (non-quantized) DDT.

tion and non-quantized data are shown, in Figure 3.10. This figure depicts that at global false alarm probability  $p_{FA} = 0.2$ , the probability of detection,  $p_D$ , attains the values 0.653, 0.684 and 0.803 for the cases of three-level data transmissions with MJD, MAE and the non-quantized data transmission, respectively. Increasing the quantization level makes the MAE and MJD methods perform more closer to the performance without quantization which is depicted in Tables 3.1 and 3.2.

Table 3.1 shows the  $p_D$  for 2, 3, 4 and 6 level MAE and MJD based quantized and non-quantized data transmissions for the values of  $p_{FA} = 0.1, 0.2, 0.3$  and  $0.4$ . At each quantization level MAE method performs better compared to MJD and the performance increases when the quantization level is increased. At 6-level quantization  $p_D$  obtained by MAE based method is only slightly inferior to the limiting case with no quantization as shown in Figure 3.11. Quantitatively, the difference in  $p_D$  is 0.022, 0.014, 0.018 and 0.002 for  $p_{FA}$  values of 0.1, 0.2, 0.3 and 0.4, respectively. Table 3.2 shows the achieved gain in  $p_D$  by using the MAE method w.r.t. MJD method and is given by  $G = (p_D^{MAE_i} - p_D^{MJD_i})$



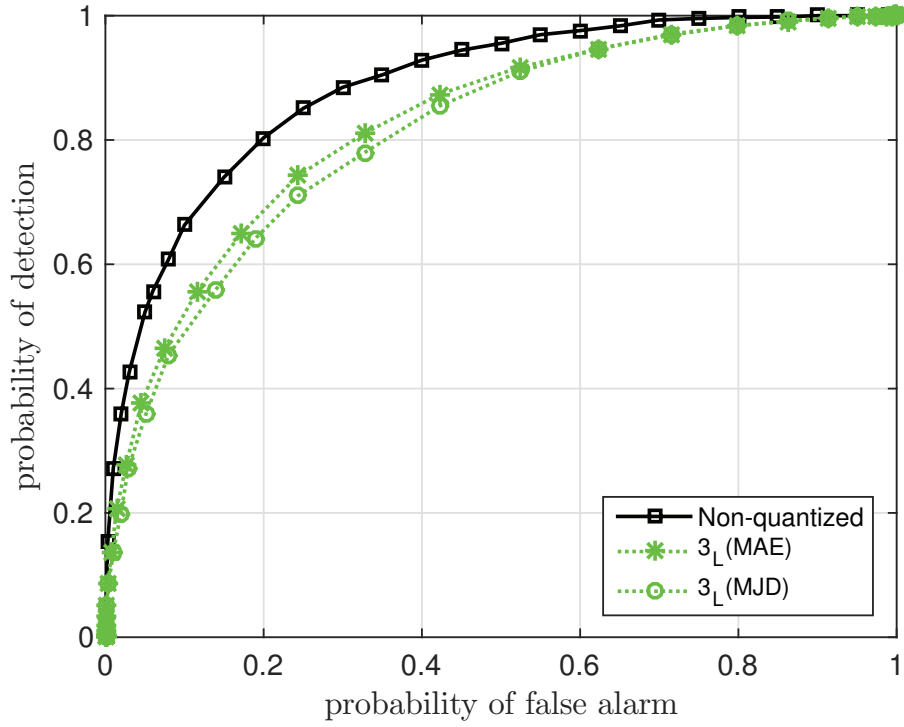


Figure 3.10. Comparison between the ROC curves obtained using MAE and MJD methods for three-level and Gaussian (non-quantized) DDT.

with the resulting percentage gain  $PG = (G \times 100\%) / p_D^{\text{MAE}_i}$ , where  $i = 2, 3, 4, 6$ .

Table 3.1. The relation between  $p_D$  and  $p_{\text{FA}}$  for different levels of quantization obtained with MAE and MJD methods.

$p_{\text{FA}} \backslash p_D$	MJD <sub>2</sub>	MAE <sub>2</sub>	MJD <sub>3</sub>	MAE <sub>3</sub>	MJD <sub>4</sub>	MAE <sub>4</sub>	MJD <sub>6</sub>	MAE <sub>6</sub>	non-quantized
0.1	0.425	0.432	0.497	0.520	0.567	0.592	0.629	0.643	0.665
0.2	0.590	0.610	0.653	0.684	0.728	0.760	0.772	0.789	0.803
0.3	0.710	0.720	0.755	0.790	0.810	0.845	0.850	0.867	0.885
0.4	0.787	0.805	0.825	0.858	0.860	0.895	0.903	0.922	0.924

It is obviously seen from the previous tables that MAE outperforms MJD for  $M \geq 2$  levels. The achieved gain of MAE w.r.t. MJD is on average 0.0138 with a corresponding percentage gain of 2.13% for the binary data transmissions, whereas the average gains are  $= 0.0305, 0.0318$  and  $0.0168$  with corresponding average percentage gains as 4.31%, 4.12% and 2.09% for 3-level, 4-level and 6-level data transmissions, respectively. In

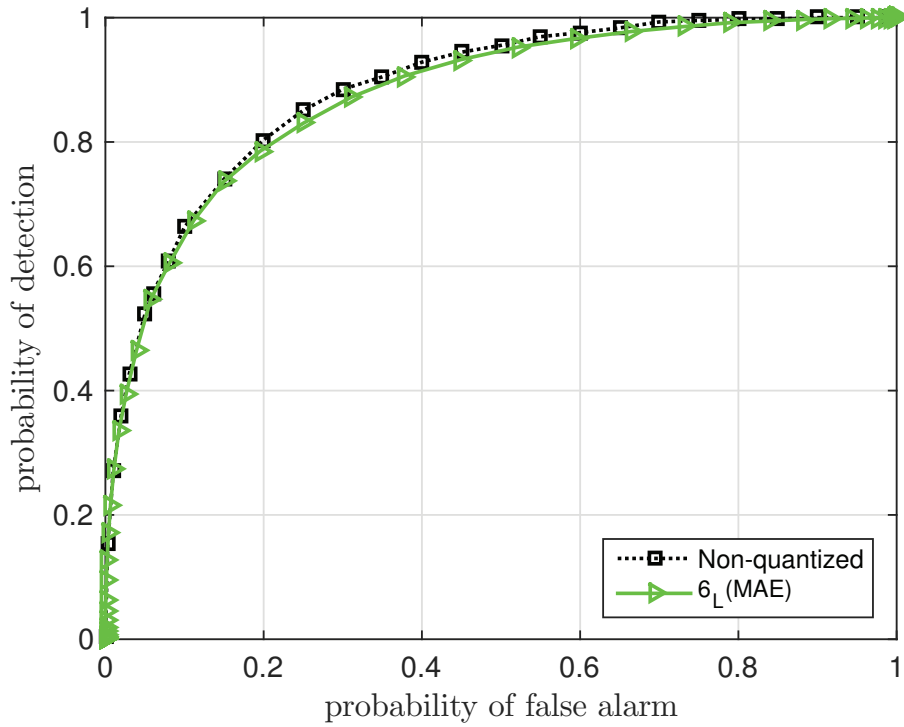


Figure 3.11. Comparison between the ROC curves obtained using MAE method for six-level and Gaussian (non-quantized) DDT.

the same manner, the average difference in  $p_D$ , for  $p_{FA} = 0.1, 0.2, 0.3$  and  $0.4$ , between the 6-level data transmissions achieved by MAE and non-quantized data transmissions equals to 0.014 with 1.8% and it is 0.03 with 3.9% between MJD and non-quantized data transmissions. These results show that 6-level data transmission by using MAE is very close to the non-quantized data transmission and gives better performance than the MJD method.

### 3.4.3. Multiple Level Data Transmission over Rayleigh Fading Channel

Figure 3.12 shows the ROC curves for 2, 3, 4 and 6 level MAE based quantized and non-quantized data transmissions by using  $M$ -FSK modulation scheme with non-coherent demodulation over Rayleigh fading channels and the optimal fusion rule in (3.7). The threshold,  $\eta$ , for each  $p_{FA}$  was estimated by running a Monte Carlo simulation under

Table 3.2. Achieved gain in  $p_D$  by using the MAE method in quantization instead of MJD.

$p_{FA}$ \ $p_D$	2-Level		3-Level		4-Level		6-Level	
	G	PG	G	PG	G	PG	G	PG
0.1	0.01	1.63	0.023	4.42	0.025	4.22	0.014	2.17
0.2	0.02	3.28	0.031	4.53	0.032	4.21	0.017	2.15
0.3	0.01	1.39	0.035	4.43	0.035	4.14	0.017	1.96
0.4	0.018	2.2	0.0330	3.85	0.035	3.914	0.019	2.06

no event case. In this figure we can see that the obtained,  $p_D$  for 6-level quantization falls behind the limiting case of no quantization by 0.09 at  $p_{FA} = 0.1$ . This gain diminishes at  $p_{FA} = 0.7$ . When we compare the performances of different quantization levels, the achieved gain in  $p_D$  by transmitting 6-level quantization instead of 2-level quantization is 0.21 for  $p_{FA} = 0.1$  and this gain diminishes at  $p_{FA} = 0.99$ . Also, the sub-optimal fusion rule in (3.15) have been used to find the ROCs for the different type of data transmissions. Figure 3.13 shows a comparison between the optimal and sub-optimal fusion rule for 2 and 6 levels data transmissions and compare them with the non-quantized data transmissions. The dashed line in ROCs for the sub-optimal fusion rule correspond to randomization in the tests (Poor, 2013). This figure shows that the achieved gain by using the optimum fusion rule w.r.t. the sub-optimal rule is 0.3 and 0.6 at  $p_{FA} = 0.1$  for the 2-level and 6- level data transmissions, respectively.

### 3.5. Conclusions on MAE Based Quantization Method

This chapter, proposed multiple level quantization methods for the sensor outputs in a WSN composed of uniformly deployed sensors and a FC used to detect a static event which can be observed as signals emitted from a point source. The developed method is an NP criterion-based DD scheme depending on MAE and MJD. Obtained ROCs show that the MAE method for quantizing sensor outputs performs significantly better than the MJD and Kth root method or other choices for determining the thresholds in binary quan-

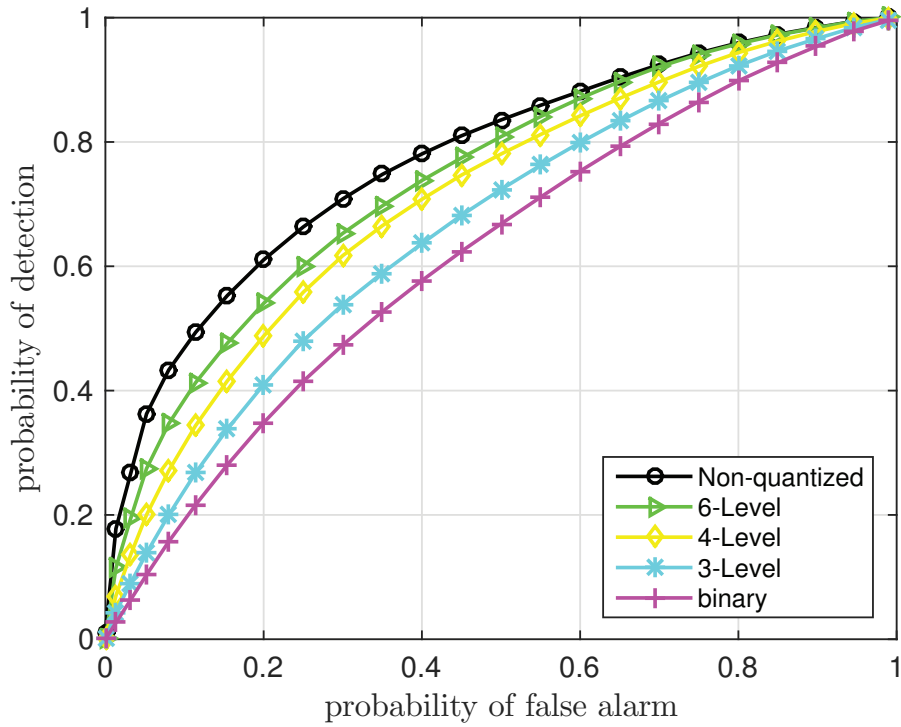


Figure 3.12. ROC curves in the case of fading channel by using MAE based quantization and optimum fusion rule.

tization of sensor outputs. The performance obtained using the MAE method has been also investigated for three-level, four-level, and six-level quantization. The increasing number of levels in quantization has resulted in better performance as expected, the six-level data transmissions approach to the non-quantized data transmissions after the value of  $p_{FA} = 0.2$ . Also, the effects of the Rayleigh fading channel from the sensors to the FC have been investigated incorporating non-coherent M-FSK communication between sensors and FC. Optimal decide and fuse and a suboptimal decide then fuse type fusion rules are applied at the FC. Using the wireless channel model similar results were obtained as in DDT. Again, results with 6-level quantization were comparable to non-quantized data transmission.

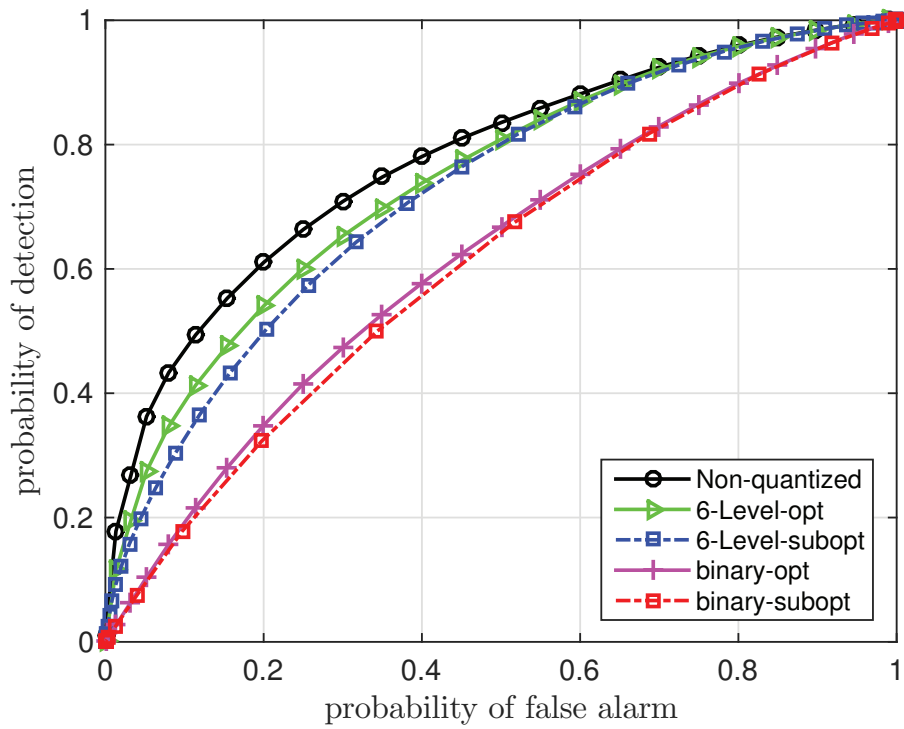


Figure 3.13. A comparison between the ROCs of the optimal and sub-optimal fusion rule for binary and six level quantizations and the corresponding non-quantized data transmissions in the case of fading channel.

## CHAPTER 4

# SEQUENTIAL DETECTION FOR MULTIPLE LOCAL SENSORS

### 4.1. Introduction

Traditional detection methods use only one processor to detect signals such as in sonar or radar systems. Recently, there is a growing interest in using several sensors, that are geographically distributed to accomplish signal detection (Ekchian and Tenney, 1982; Tenney and Sandell, 1981; Chair and Varshney, 1986b). In (Tenney and Sandell, 1981), they consider a detection problem consisting of two sensors and one FC with a fixed fusion rule to show that the optimum local decision rule is the LRT under the Bayesian criterion. Later, in (Chair and Varshney, 1986b), it was shown that the optimum fusion rule at the FC for multiple sensors is also a LRT under the NP and the Bayesian criteria. Then, in (Tsitsiklis, 1988), they considered the case when the number of sensors goes to infinity to show that the optimal decision rule is an identical LRT. In (Blum and Deans, 1998), a distributed random signal detection problem with multiple-bit decisions have been covered.

Many works on decentralized detection, containing the above mentioned approaches, consider a FSS approach where the FC collects a fixed number of observations to make its final decision at a fixed time. There is also a noteworthy volume of literature that considers the sequential detection, i.e. (Tantaratana and Thomas, 1977b; Veeravalli et al., 1993; Fellouris and Moustakides, 2010; Yilmaz et al., 2012a) when the data acquisition is costly and the decision delay is important. The initial research on sequential detection theory was proposed by Wald in (Wald and Wolfowitz, 1948) to decide between two simple hypotheses. The SPRT has been proposed in (Wald and Wolfowitz, 1948) where the cumulative sum of the LLRs is compared with prescribed thresholds to perform the final decision at the FC. SPRT is the optimum among all the sequential test in terms of

the minimum ASN required to obtain specific detection performance. SPRT requires four times less ASN compared to the FSS for specific performance (Poor, 2013).

Tantaratana and Thomas investigate the performance of the sequential detection for binary level data transmissions (the sign test) in (Tantaratana and Thomas, 1977a). Also in (Tantaratana and Thomas, 1977b) they examine the optimal sequential quantizer for multi-level data transmissions of a constant signal. In that work, a considerable saving in the ASN by using sequential detection instead of the optimum FSS detector (Tenney and Sandell, 1981; Chair and Varshney, 1986b) has been demonstrated. After that in (Veeravalli et al., 1993), a binary hypothesis testing problem was investigated for the system with a local memory restricted on the past decisions and full feedback. The sequential detection problem was performed at the FC depending on the received i.i.d. information from each sensor.

Then in (Cheng et al., 2005), a distributed sequential detection problem under communication constraints is studied. A scalar quantizer (independent quantizer design for Bayesian detection) was used to map the received signal at each sensor to a discrete  $M$ -level signal and transmit it to the FC where the sequential data fusion scheme is performed. After that in (Chaudhari et al., 2009), the authors applied a sequential detection problem at FC to detect OFDM signals. The comparison with FSS detection shows a significant saving in the ASN needed to perform the final decision at FC.

In this thesis, we use the regular Rayleigh fading channel model for the wireless channel. Additionally, optimal fusion rules are utilized which are modified from the ones for  $M$ -ary hypothesis testing in (Hajibabaei and Vosoughi, 2014) in order to match the binary hypothesis testing problem with  $M$ -ary modulated data transmission rather than binary modulated data transmission for SPRT in (Yilmaz et al., 2012b).

We notice from the literature that there are two main approaches to solve the sequential problem. The first one solves the sequential test by optimizing the Bayesian problem by using the dynamic programming (Wald and Wolfowitz, 1948), i.e. SPRT, whereas the second approach considers an asymptotic regime which supposes that the cost of the observations is very low (Berger, 2013; Baum and Veeravalli, 1994). The second approach cannot be applied to WSNs because of the limitation on the power of the sensors. Therefore, the dynamic programming approach is used in this thesis.

From the previous discussion, we can notice that the hypothesis testing procedures

can be classified into two main categories according to the number of observations or the test duration. In the classical theory of hypothesis testing, the number of observations is treated as constant. Many approaches such as the NP approach and the Bayesian approach can be employed in a FSS detection. On the other hand, the sequential test depends on the outcome of the observation process and the test thresholds, therefore, the number of observations is not predetermined, but a random value. For some experiments, a decision can be made by taking a small number of observations, while for others, the process of making observations is extended before making a decision.

A comparison between FSS detection and SPRT has been performed for the binary hypothesis testing problem in (3.1), according to the number of observations required to achieve a specific performance. The pdfs of the random variable  $y$  under the two hypotheses  $H_0$  and  $H_1$  are given as  $\mathcal{N}(0, \sigma_N^2)$  and  $\mathcal{N}(\bar{A}, \sigma_N^2)$  respectively. The LLR,  $\log(\lambda(y))$ , can be calculated by using the expected value of the signal amplitude,  $\bar{A}$ , as follows:

$$\log(\lambda(y)) = -\frac{\bar{A}^2}{2\sigma_N^2} + \frac{\bar{A}}{\sigma_N^2}y. \quad (4.1)$$

By using (2.62) we can find the Kullback-Leibler divergence for the conditional pdfs  $p(y|H_0)$  and  $p(y|H_1)$  under the two hypotheses  $H_0$  and  $H_1$  respectively, as follows:

$$D_{\text{KL}}(p(y|H_1)||p(y|H_0)) = \int_{-\infty}^{\infty} \log(\lambda(y))p(y|H_1)dy = \frac{\bar{A}^2}{2\sigma_N^2}, \quad (4.2)$$

$$D_{\text{KL}}(p(y|H_0)||p(y|H_1)) = -\int_{-\infty}^{\infty} \log(\lambda(y))p(y|H_0)dy = -\frac{\bar{A}^2}{2\sigma_N^2}. \quad (4.3)$$

For  $p_m = p_{\text{fa}}$  we can re-write (2.60) and (2.61) as follows:

$$E(K|H_1) = E(K|H_0) \approx \frac{2\sigma_N^2}{\bar{A}^2} \log\left(\frac{1-p_{\text{fa}}}{p_{\text{fa}}}\right) [1-2p_{\text{fa}}]. \quad (4.4)$$

Now, let's find the sample number for FSS. The best block detector is the LR detector, which has probability of false alarm,  $p_{\text{fa}}$  and probability of detection,  $p_d$  for a given sample size  $N$ . For a given  $p_d$  and  $p_{\text{fa}}$ , the sample size needed for FSS detection can be given as:

$$p_{\text{fa}} = Q_{\text{func}}\left(\frac{\lambda}{\sqrt{N\sigma_N^2}}\right). \quad (4.5)$$



By taking the inverse of (4.5)

$$Q_{\text{func}}^{-1}(p_{\text{fa}}) = \left( \frac{\lambda}{\sqrt{N\sigma_N^2}} \right). \quad (4.6)$$

So, the threshold can be written as

$$\lambda = \sqrt{N\sigma_N^2} Q_{\text{func}}^{-1}(p_{\text{fa}}). \quad (4.7)$$

Then the probability of detection for the given  $p_{\text{fa}}$  and  $N$  will be found as

$$p_d = Q_{\text{func}} \left( \frac{\lambda - N\mu}{\sqrt{N\sigma_N^2}} \right). \quad (4.8)$$

By substituting (4.7) into (4.8), we can rewrite (4.8) as

$$p_d = Q_{\text{func}} \left( \frac{\sqrt{N\sigma_N^2} Q_{\text{func}}^{-1}(p_{\text{fa}}) - N\mu}{\sqrt{N\sigma_N^2}} \right), \quad (4.9)$$

which can be further simplified to

$$p_d = Q_{\text{func}} \left( Q_{\text{func}}^{-1}(p_{\text{fa}}) - \frac{\sqrt{N}\mu}{\sigma_N} \right). \quad (4.10)$$

By taking the inverse of (4.9) as

$$Q_{\text{func}}^{-1}(p_d) = \left( Q_{\text{func}}^{-1}(p_{\text{fa}}) - \left( \frac{\sqrt{N}\mu}{\sigma_N} \right) \right) \quad (4.11)$$

we can find the sample size for the FSS detection as

$$N = \frac{\sigma_N^2}{\mu^2} (Q_{\text{func}}^{-1}(p_d) - Q_{\text{func}}^{-1}(p_{\text{fa}}))^2. \quad (4.12)$$

Figure 4.1 shows the ratio of the expected number of the observations for SPRT and the number of observations for FSS detection, for  $p_{\text{fa}} = 0.01$ ,  $p_d = 0.99$  and  $\frac{\sigma_N^2}{\mu^2} = 1$ .

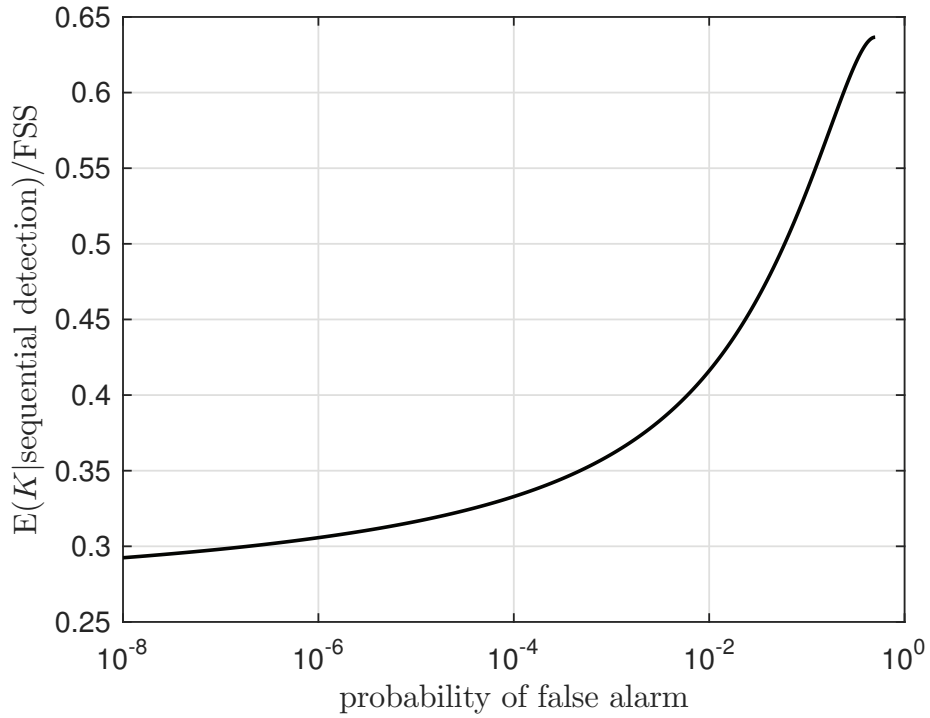


Figure 4.1. Ratio of the expected number of observations in sequential detection and the sample size in block detection versus the global probability of false alarm.

This figure depicts that sequential detection can save around 1/3 of the observations at  $p_{fa} = 10^{-6}$ .

The following subsections will consider two operating modes according to the place(s) at which decisions are made and the contents of the transmissions from the sensor nodes to FC.

## 4.2. Centralized SPRT for Multiple Local Detectors

The central LR ( $\Lambda_K^c(\mathbf{y})$ ) evaluated at the FC for each observation time can be given as

$$\Lambda_K^c(\mathbf{y}) = \prod_{i=1}^K \frac{p(y_i|H_1)}{p(y_i|H_0)}. \quad (4.13)$$

Equation (4.13) can be written as

$$\Lambda_K^c(\mathbf{y}) = \prod_{i=1}^{K-1} \frac{p(y_i|H_1) p(y_K|H_1)}{p(y_i|H_0) p(y_K|H_0)}, \quad (4.14)$$

which is presented as the product of the  $\Lambda^c(\mathbf{y})$  at the  $(K-1)$ th step  $\Lambda_{K-1}^c(\mathbf{y})$  and the  $k$ th increment  $\lambda^c(y_k)$  which is given by

$$\lambda^c(y_k) = \frac{p(y_k|H_1)}{p(y_k|H_0)}. \quad (4.15)$$

Hence,

$$\Lambda_K^c(\mathbf{y}) = \Lambda_{K-1}^c(\mathbf{y}) \lambda^c(y_K). \quad (4.16)$$

The centralized SPRT at step (time)  $k$  compares  $\Lambda_K^c(\mathbf{y})$  with two thresholds (Wald and Wolfowitz, 1948):  $\eta_1$  and  $\eta_0$  as follows

$$\log(\Lambda_K^c(\mathbf{y})) \begin{cases} \geq \log(\eta_1), & \text{stop and decide } H_1 \\ \leq \log(\eta_0), & \text{stop and decide } H_0 \\ \text{otherwise,} & \text{continue,} \end{cases} \quad (4.17)$$

where

$$\eta_1 \approx \frac{1 - p_m}{p_{fa}}, \quad \eta_0 \approx \frac{p_m}{1 - p_{fa}}. \quad (4.18)$$

By taking the logarithm of the  $\Lambda_K^c(\mathbf{y})$  as given in (4.14), we obtain

$$\log(\Lambda_K^c(\mathbf{y})) = \sum_{i=1}^{K-1} \log\left(\frac{p(y_i|H_1)}{p(y_i|H_0)}\right) + \log\left(\frac{p(y_K|H_1)}{p(y_K|H_0)}\right). \quad (4.19)$$

Equation (4.19) shows that at any step of the sequential test every local sensor computes the logarithm of its likelihood function based on its observation and transmits the result to the FC where it uses the received local test to update its statistic  $\log(\Lambda_K^c(\mathbf{y}))$ .

Since the observations at the  $i^{th}$  sensor,  $i = 1, 2, \dots, K$ , are i.i.d. random variables, then,  $\lambda(y_K)$ , is also a series of i.i.d. random variables. Suppose that, there is no excess over the thresholds (Kowalski, 1971), then we have

$$\log(\Lambda_K^c(\mathbf{y})) = \begin{cases} \log(\eta_1), & \text{w.p. } p_j \\ \log(\eta_0), & \text{w.p. } (1 - p_j) \end{cases} \quad (4.20)$$

where  $p_0 = p_{\text{fa}}$  and  $p_1 = (1 - p_m)$  for  $j = 0, 1$ .

By taking the expected value of (4.19) and separating it into two expectations we obtain

$$E(\log(\Lambda_K^c(\mathbf{y})|H_0)) = E_K E \left( \sum_{i=1}^K \log \left( \frac{p(y_i|H_1)}{p(y_i|H_0)} \right) \middle| K, H_0 \right). \quad (4.21)$$

where the outer expectation is w.r.t.  $K$  and the inner one w.r.t. hypothesis  $H_0$ . Now, using the Kullback-Leibler divergence as defined in (2.62) between the conditional pdfs  $p(y|H_0)$  and  $p(y|H_1)$ , we can re-write (4.21) as

$$E(\log(\Lambda_K^c(\mathbf{y})|H_0)) = E(K|H_0)[-D_{\text{KL}}(p(y|H_0)||p(y|H_1))], \quad (4.22)$$

The expected number of the observations for the sequential detection are given as (Levy, 2008; Van et al., 2013):

$$E(K|H_0) = \frac{(1 - p_{\text{fa}}) \log\left(\frac{p_m}{1 - p_{\text{fa}}}\right) + p_{\text{fa}} \log\left(\frac{1 - p_m}{p_{\text{fa}}}\right)}{-D_{\text{KL}}(p(y|H_0)||p(y|H_1))}, \quad (4.23)$$

and

$$E(K|H_1) = \frac{p_m \log\left(\frac{p_m}{1 - p_{\text{fa}}}\right) + (1 - p_m) \log\left(\frac{1 - p_m}{p_{\text{fa}}}\right)}{D_{\text{KL}}(p(y|H_1)||p(y|H_0))}. \quad (4.24)$$

### 4.3. Distributed SPRT for Multiple Local Detectors

In centralized SPRT, we assumed that every local detector transmitted its real observations (Gaussian observations) to the FC. On the contrary, in this subsection, we suppose that each local detector performs its binary decision and transmits the decision bit to the FC where a distributed SPRT is performed. The binary quantization of the local decision reduces the required communication channel bandwidth. Moreover, it simplifies

the complexity of the SPRT at the FC. The system model defined in Section 3.2 is adopted as the sequential detection scenario. The binary hypothesis testing problem at the FC is given by:

$$\begin{aligned} H_0 : \mathbf{y}_k &= \mathbf{B}(p_{\text{fa}}) \\ \text{versus} & \\ H_1 : \mathbf{y}_k &= \mathbf{PB}(\mathbf{p}_d), \end{aligned} \quad (4.25)$$

where  $\mathbf{B}(p_{\text{fa}})$  is a binomial random variable with probability of success  $p_{\text{fa}}$  when  $H_0$  is true and  $\mathbf{PB}(\mathbf{p}_d)$  is Poisson binomial random variable with vector of probabilities  $\mathbf{p}_d$  when  $H_1$  is true, where  $\mathbf{p}_d = [p_{d_1}, p_{d_2}, \dots, p_{d_K}]$ . Because of the complexity in finding a theoretical expression for the ASN in the case of Poisson binomial distribution we assume that there is a single  $p_d$  which is valid for every sensor. In making this approximation, we consider that the observations at the sensors are identically distributed and they come from the same pdf of  $A$  and consequently we assume that they have the same constant signal amplitude which is equal to the mean value of  $A$ . Then, similar to (4.23) we can write the ASN as follows:

$$E(K|H_0) = \frac{(1 - p_{\text{fa}}) \log\left(\frac{p_m}{1 - p_{\text{fa}}}\right) + p_{\text{fa}} \log\left(\frac{1 - p_m}{p_{\text{fa}}}\right)}{p^{H_0} \log(p^{H_1}/p^{H_0}) + (1 - p^{H_0}) \log((1 - p^{H_1})/(1 - p^{H_0}))}, \quad (4.26)$$

and

$$E(K|H_1) = \frac{p_m \log\left(\frac{p_m}{1 - p_{\text{fa}}}\right) + (1 - p_m) \log\left(\frac{1 - p_m}{p_{\text{fa}}}\right)}{p^{H_1} \log(p^{H_1}/p^{H_0}) + (1 - p^{H_1}) \log((1 - p^{H_1})/(1 - p^{H_0}))}. \quad (4.27)$$

Each binary local detector transmits a binary value (0 or 1) to the FC without regard to the signal level its observation, this leads to a degradation in the system performance by increasing the ASN, but we can compensate this degradation by further partitioning of the sample space and adding more quantization levels.

The binary hypothesis testing problem for multilevel local detectors is defined as:

$$\begin{aligned} H_0 : \mathbf{Y}_k &= \mathbf{M}(\mathbf{p}_k^{\text{H}_0}) \\ \text{versus} & \\ H_1 : \mathbf{Y}_k &= \mathbf{M}(\mathbf{p}_k^{\text{H}_1}), \end{aligned} \quad (4.28)$$

$\mathbf{M}(\mathbf{p}_k^{\text{H}_j})$  is a multinomial random variable where the probabilities of the particular outcomes are given by the  $M$ -vector of probability masses  $\mathbf{p}_k^{\text{H}_j}$  when  $H_j$  is true for  $j = 0, 1$ .

The ASNs for  $M$  level quantization under the two hypotheses  $H_0$  and  $H_1$  are given as follows:

$$E(K|H_0) = \frac{(1 - p_{fa}) \log\left(\frac{p_m}{1-p_{fa}}\right) + p_{fa} \log\left(\frac{1-p_m}{p_{fa}}\right)}{\sum_{m=1}^M p_m^{H_0} \log(p_m^{H_1}/p_m^{H_0})}, \quad (4.29)$$

$$E(K|H_1) = \frac{p_m \log\left(\frac{p_m}{1-p_{fa}}\right) + (1 - p_m) \log\left(\frac{1-p_m}{p_{fa}}\right)}{\sum_{m=1}^M p_m^{H_1} \log(p_m^{H_1}/p_m^{H_0})}, \quad (4.30)$$

where

$$p_m^{H_j} = \int_{Area_m} p(y|H_j) dy \quad (4.31)$$

for  $j = 0, 1$  and  $m = 2, 3, 4$  and  $6$ .

#### 4.4. Simulations for Sequential Probability Ratio Test

Monte Carlo simulations are used to evaluate the expected ASN of the MAE and MJD quantization methods as described in Algorithm 2. First, we performed simulations using the DDT method that is assuming that the sensor outputs are available error-free at the FC for both of the cases when the MAE and MJD methods are used. Then, we considered a Rayleigh fading channel to show the channel effect on the ASN with the MAE and MJD quantization methods.

##### 4.4.1. The Relation Between ASN and the Required $p_d$ for MAE

###### Method

In this section, the detection performance of SPRT with the MAE quantization method is considered using the DDT. Firstly, we show the relationship between the expected value of  $ASN_i$  and  $p_d$ , where  $i$  is the index of the hypothesis  $H_i$ ,  $i = 0, 1$ . Table 4.1 shows the mean of the average sample number  $ASN_m$  versus  $p_d$ , where  $ASN_m$  is given by

$$ASN_m = (ASN_0 + ASN_1)/2 \quad (4.32)$$

where the transmitted data is either non-quantized data or MAE quantized data. Table 4.1 shows that  $ASN_m$  is a monotonically increasing function of  $p_d$ , for  $SNR= 0$ dB and  $p_{fa} = 0.1$ . At a specific probability of detection, the  $ASN_m$  decreases monotonically with increasing quantization level. Minimum  $ASN_m$  is achieved using non-quantized data transmission, while the maximum  $ASN_m$  is achieved using binary data transmission. For example, at  $p_d = 0.9$ , the  $ASN_m$  values using 2-level and 6-level data transmissions are 54.8356 bits and 38.8559 bits, respectively, which means that 6-level data transmission will use 1.4113 times less energy compared to binary data transmission.

Table 4.1. The relation between the ASN and probability of detection for the different quantization level at  $p_{fa} = 0.1$  and  $SNR= 0$  dB.

$p_d \backslash ASN_m$	non-quant.	MAE <sub>6</sub>	MAE <sub>4</sub>	MAE <sub>3</sub>	MAE <sub>2</sub>
0.6	3.3529	5.9107	8.0312	11.7367	28.6040
0.7	4.4700	8.2943	11.8913	17.2741	36.6022
0.8	5.676	10.3643	15.7157	22.6094	45.9801
0.9	7.2518	15.0315	20.9756	30.4379	54.8356
0.99	11.0415	24.3748	35.9332	48.1852	75.5284
0.999	12.8885	31.7861	46.7581	60.6366	88.4240

#### 4.4.2. Comparison Between SPRT and FSS

In this section, we compare SPRT and FSS detection according to the number of sensors required to achieve particular false alarm and detection probabilities at FC, where MAE and MJD are the used quantization methods. The performance of the FSS was evaluated using  $k = 25$  sensors at  $SNR= 0$  dB. In order to perform a fair comparison between SPRT and FSS using MAE and MJD quantization methods, the upper and lower thresholds of SPRT are set using  $p_d$  and  $p_{fa}$  from Table 4.2.

In Table 4.3, the  $ASN_i$  under hypothesis  $H_i$ ,  $i = 0, 1$ , are evaluated for different types of data transmissions using MAE quantization method. Table 4.3 shows a significant saving in the number of used sensors using SPRT. For example, for 6-level data

transmission,  $ASN_m = 5.4710$  sensors using SPRT instead of using 25 sensors by FSS to achieve  $p_d = 0.643$  and  $p_{fa} = 0.1$ , which means saving 19.5290 sensors. It can also be seen that at  $p_{fa} = 0.1$ , we can save 17.27 sensors at  $p_d = 0.592$ , 14.75 sensors at  $p_d = 0.52$ , and 10.8639 sensors at  $p_d = 0.432$  using 4-level, 3-level and 2-level quantization, respectively. On the other hand, we can save 20.905 sensors at  $p_d = 0.665$  and  $p_{fa} = 0.1$  for non-quantized data transmissions using SPRT.

The  $ASN_i$  obtained using MJD quantization method are shown in Table 4.4. We can see that at  $p_{fa} = 0.1$ , the  $ASN_m = 15.94$  sensors at  $p_d = 0.425$ ,  $ASN_m = 11.63$  sensors at  $p_d = 0.497$ ,  $ASN_m = 9.19$  sensors at  $p_d = 0.567$  and  $ASN_m = 6.18$  sensors at  $p_d = 0.629$  for 2-level, 3-level, 4-level and 6-level of quantization respectively instead of using 25 sensors by using FSS.

The obtained results from the SPRT method with MAE and MJD are listed in Table 4.3 and Table 4.4, respectively. Table 4.3 and Table 4.4 show that the SPRT method achieves a significant saving in the transmission power by decreasing the number of transmitted observations to the FC. Moreover, it can be noticed that the  $ASN_m$  obtained using MAE is less than the  $ASN_m$  resulting from using MJD.

Table 4.2. The relation between  $p_d$  and  $p_{fa}$  for different levels of quantization obtained with MAE and MJD methods.

$p_{fa} \backslash p_d$	MJD <sub>2</sub>	MAE <sub>2</sub>	MJD <sub>3</sub>	MAE <sub>3</sub>	MJD <sub>4</sub>	MAE <sub>4</sub>	MJD <sub>6</sub>	MAE <sub>6</sub>	non-quantized
0.1	0.425	0.432	0.497	0.520	0.567	0.592	0.629	0.643	0.665
0.2	0.590	0.610	0.653	0.684	0.728	0.760	0.772	0.789	0.803
0.3	0.710	0.720	0.755	0.790	0.810	0.845	0.850	0.867	0.885
0.4	0.787	0.805	0.825	0.858	0.860	0.895	0.903	0.922	0.924

#### 4.4.3. Comparison Between MAE and MJD Methods

The simulation results of  $ASN_m$  are obtained for the quantized and non-quantized data transmissions using DDT. For the quantized data transmission, different quantization levels are used using MAE and MJD methods. In Figure 4.2  $ASN_m$  versus  $\log_{10}(p_{fa}) = \log_{10}(p_m)$  curves are shown where the quantization levels and methods are parameters. This figure depicts that at  $\log_{10}(p_{fa}) = \log_{10}(p_m) = -3$ , the  $ASN_m$ , attains the values



205.2, 175.2, 131.9 and 59.28 sensors for the cases of 2-level, 3-level, 4-level and 6-level data transmissions, respectively, with MJD, whereas it attains the values 192.5, 154, 112.5 and 49.18 sensors for the cases of 2-level, 3-level, 4-level and 6-level data transmissions, respectively, with MAE. This means that we can save 12.7, 21.2, 19.4 and 10.1 sensors with corresponding saving percentages (SPs) 6.18, 12.10, 14.7 and 17.03 for 2-level, 3-level, 4-level and 6-level data transmissions, respectively, using MAE method.

Table 4.5 shows that the achieved SP in  $ASN_m$  using the MAE method over MJD method, which is given as

$$SP = (S \times 100\%) / ASN_m^{MJD_i} \quad (4.33)$$

where  $S = (ASN_m^{MJD_i} - ASN_m^{MAE_i})$ , and  $i = 2, 3, 4, 6$ .

Figure 4.3 shows theoretical  $ASN_m$  which is evaluated for non-quantized and quantized data transmissions using (4.4), (4.29), (4.30) and (4.32), respectively. The quantization levels are 2, 3, 4 and 6 levels using MAE and MJD quantization methods. In Figure 4.4, we show a comparison between simulation and theoretical evaluation of  $ASN_m$  for quantized and non-quantized data transmission. For the quantized data transmission, both MAE and MJD methods are used with 4-level and 2-level quantization.

#### 4.4.4. SPRT for Multiple Level Data Transmission over Rayleigh Fading Channel

Figure 4.5 shows the average number of data transmission using MAE method,  $M$ -FSK modulation scheme with non-coherent demodulation over Rayleigh fading channels and the optimal fusion rule in (3.7). In this figure, we can see that the obtained  $ASN_m$  decreases by increasing the level of data transmissions i.e. at  $\log_{10}(p_{fa}) = \log_{10}(p_m) = -3$ ,  $ASN_m = 264, 179.4, 136.6$  and  $70.1$  for 2-level, 3-level 4-level and 6-level quantization, respectively. Figure 4.6 shows the results obtained using MJD method of quantization and keeping every other parameter as used for Figure 4.5. This time the obtained

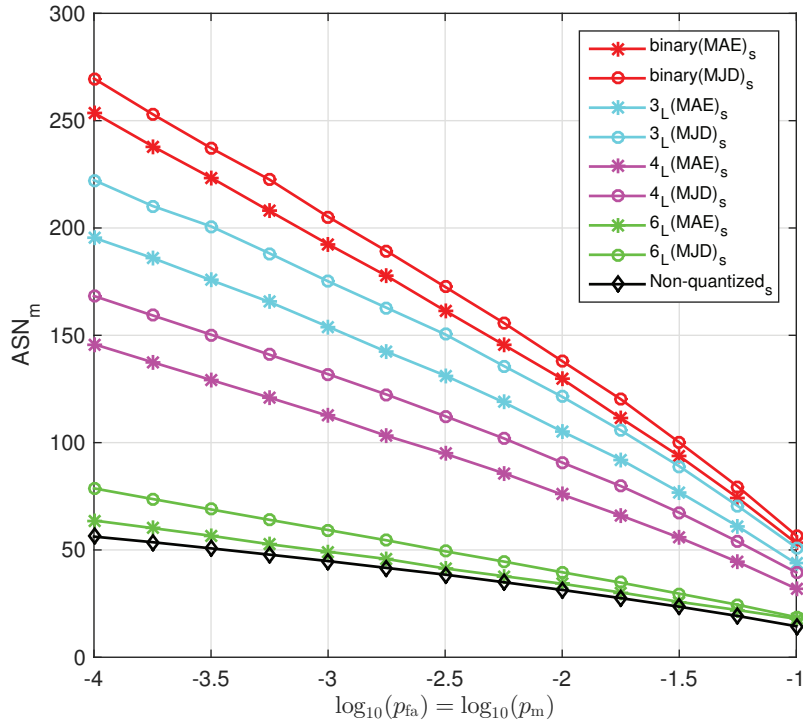


Figure 4.2. Simulated  $ASN_m$  versus logarithm of false alarm or miss probability for different levels of MAE or MJD quantization and DDT.

average  $ASN$  values are 281, 213, 163.2 and 86.91 for 2-level, 3-level 4-level and 6-level quantization, respectively. Figure 4.7 shows a comparison between the case of DDT and Rayleigh fading channels and for 2-level, 3-level 4-level and 6-level quantization. As we expected, to compensate for the resulted degradation caused by the fading channel,  $ASN_m$  increases to achieve a specific probability of error.

The increase in of  $ASN_m$  equals 71.5, 25.4, 24.1 and 20.82 when we use the fading channel instead of DDT to achieve  $\log_{10}(p_{fa}) = \log_{10}(m) = -3$  probability of error, for 2-level, 3-level 4-level and 6-level quantization, respectively.

We tabulate the achieved SP in  $ASN_m$  using MAE over MJD method for the case of fading channel in Table 4.6.

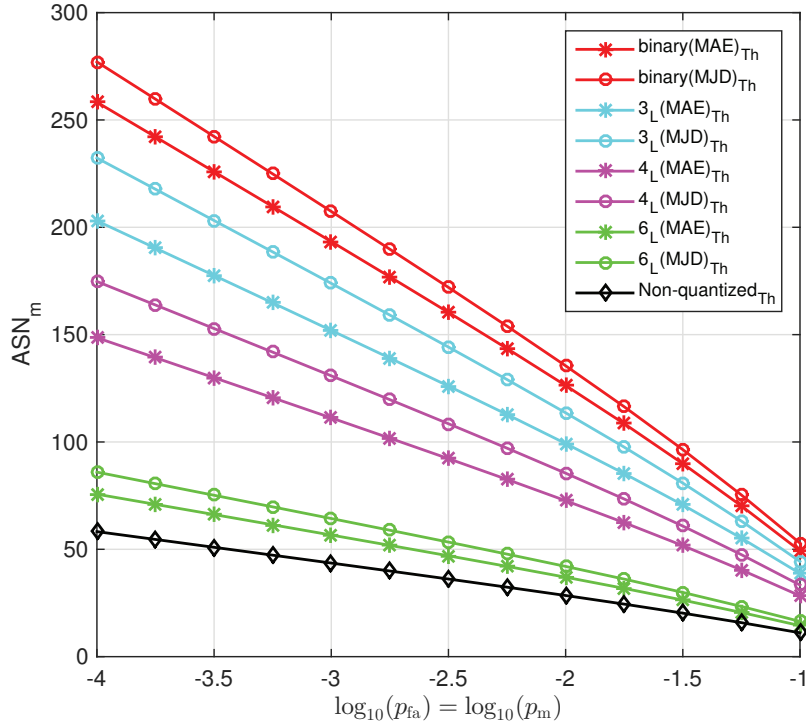


Figure 4.3. Theoretical  $ASN_m$  versus logarithm of false alarm or miss probability for different levels of MAE or MJD quantization and DDT.

#### 4.5. Conclusions on Sequential Detection

In this chapter, we have applied the MAE and MJD quantization methods to sensor outputs in the case of SPRT method which is a sequential version of detection adapted to real time analysis for streaming data flows. We have performed a comparison between FSS detection and SPRT which shows a significant saving in the number of sensors when using the SPRT method as expected. Moreover, the ASN for  $M$ -level data transmission has been evaluated theoretically and using Monte Carlo simulations for both DDT and using  $M$ -FSK modulation scheme with non-coherent demodulation over Rayleigh fading channels with optimum fusion rule. The achieved ASN using the MAE quantization method is less than the ASN resulting from using MJD for the same  $p_{fa}$  and  $p_d$ . The proposed, MAE, method performed significantly better compared to the MJD method in terms of the ASN.

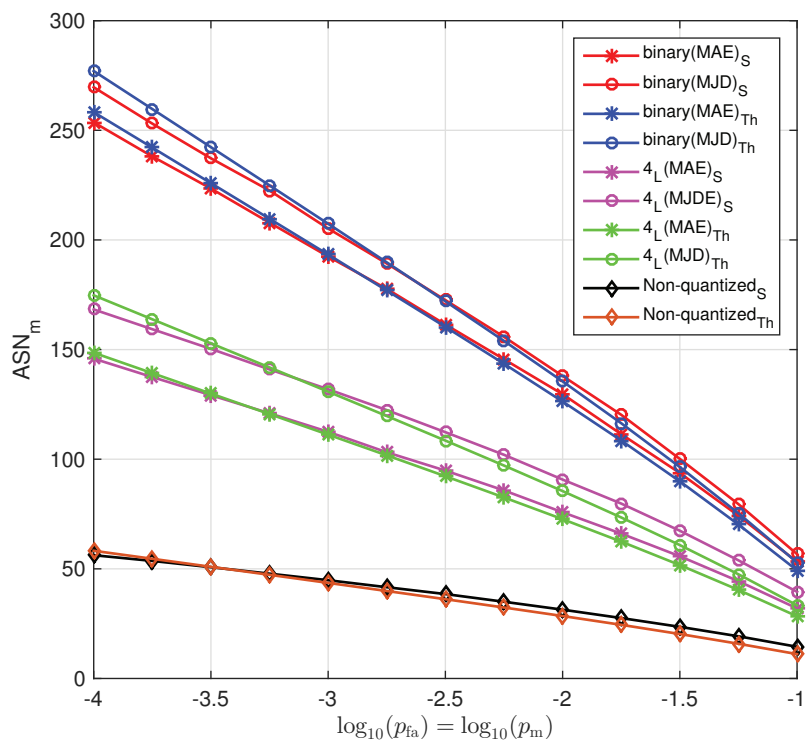


Figure 4.4. Theoretical and simulated  $ASN_m$  versus logarithm of false alarm or miss probability for different levels of MAE or MJD quantization and DDT.

Table 4.3. The ASN using MAE quantization method for different types of data transmissions.

$P_{fa}$	2-Level		3-Level		4-Level		6-Level		Non-quantized						
	$p_D$	ASN <sub>0</sub>	ASN <sub>1</sub>	$p_D$	ASN <sub>0</sub>	ASN <sub>1</sub>	$p_D$	ASN <sub>0</sub>	ASN <sub>1</sub>	$p_D$	ASN <sub>0</sub>	ASN <sub>1</sub>			
0.1	0.432	8.855	19.418	0.52	6.951	13.538	0.592	6.0543	9.398	0.643	3.503	7.439	0.665	2.58	5.61
0.2	0.610	11.943	16.048	0.684	8.067	10.926	0.76	7.226	8.3799	0.789	5.297	6.447	0.803	3.28	4.9
0.3	0.72	14.018	13.988	0.79	9.862	10.190	0.845	8.501	7.231	0.867	5.646	4.599	0.885	3.95	4.49
0.4	0.805	15.856	12.295	0.858	10.657	8.228	0.895	9.783	6.27	0.922	6.784	4.358	0.924	4.618	4.42

Table 4.4. The ASN using MJD quantization method for different types of data transmissions.

$P_{fa}$	2-Level		3-Level		4-Level		6-Level					
	$p_D$	ASN <sub>0</sub>	ASN <sub>1</sub>	$p_D$	ASN <sub>0</sub>	ASN <sub>1</sub>	$p_D$	ASN <sub>0</sub>	ASN <sub>1</sub>			
0.1	0.425	11.23	20.6619	0.497	7.9986	15.2574	0.567	6.4073	11.9740	0.629	4.17	8.19
0.2	0.59	12.941	17.6592	0.653	9.3753	13.738	0.728	7.8557	10.4080	0.772	6.10	7.0998
0.3	0.71	14.876	14.8488	0.755	11.1589	10.8331	0.81	8.8003	8.3844	0.85	6.58	5.1056
0.4	0.787	16.2556	13.2527	0.825	12.9004	10.319	0.86	10.4938	7.6306	0.903	7.7645	5.1254

Table 4.5. Achieved gain in  $ASN_m$  using the MAE method in quantization w.r.t MJD for DDT.

$p_{fa}$ \ $ASN_m$	2-Level		3-Level		4-Level		6-Level	
	S	SP	S	SP	S	SP	S	SP
$10^{-4}$	16	5.93	26.9	12.10	22.4	13.30	14.96	19.01
$10^{-3}$	12.7	6.18	21.2	12.10	19.4	14.70	10.1	17.03
$10^{-2}$	8.3	6.01	16.2	13.34	14.91	16.43	5.34	13.48
$10^{-1}$	3.57	6.29	7.5	14.67	7.55	19.10	0.77	4.12

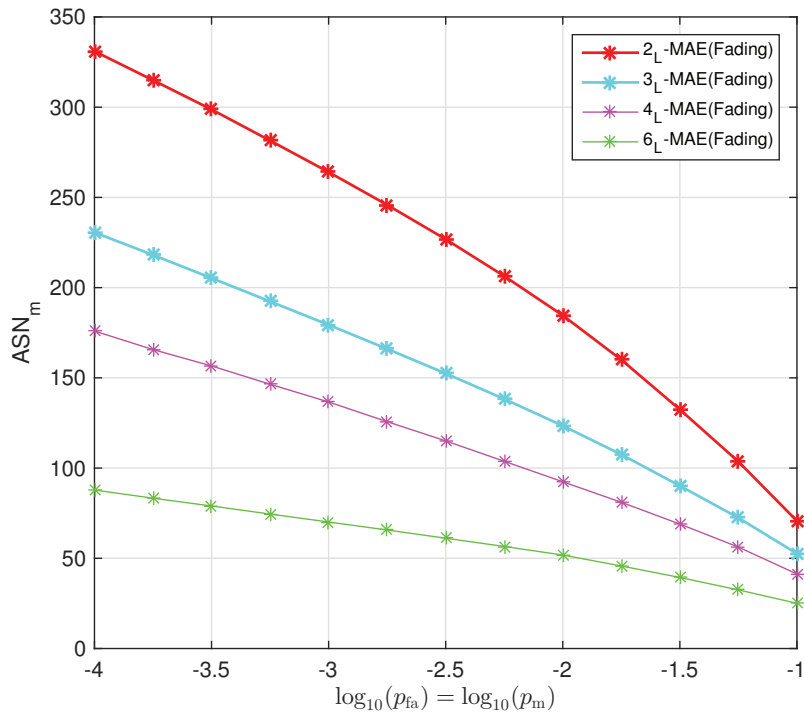


Figure 4.5. Simulated  $ASN_m$  versus logarithm of false alarm or miss probability for different levels of MAE quantization and non-coherent M-FSK communication over fading channels.

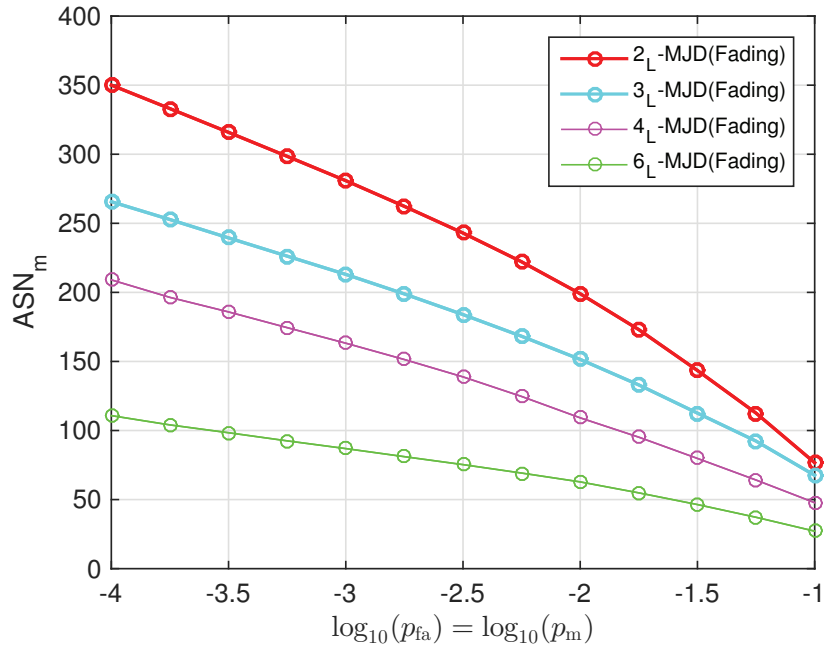


Figure 4.6. Simulated  $ASN_m$  versus logarithm of false alarm or miss probability for different levels of MJD quantization and non-coherent M-FSK communication over fading channels.

Table 4.6. Achieved gain in  $ASN_m$  by using the MAE method in quantization w.r.t MJD in the case of fading channel.

$p_{fa}$ \ $ASN_m$	2-Level		3-Level		4-Level		6-Level	
	S	SP	S	SP	S	SP	S	SP
$10^{-4}$	19.06	5.4454	35.04	13.19	33.11	15.83	22.93	20.70
$10^{-3}$	16.83	5.99	33.66	15.8	26.58	16.28	16.81	19.34
$10^{-2}$	14.39	7.23	28.05	18.52	16.76	15.35	10.99	17.51
$10^{-1}$	5.98	7.79	14.96	22.14	6.45	13.51	2.07	7.6

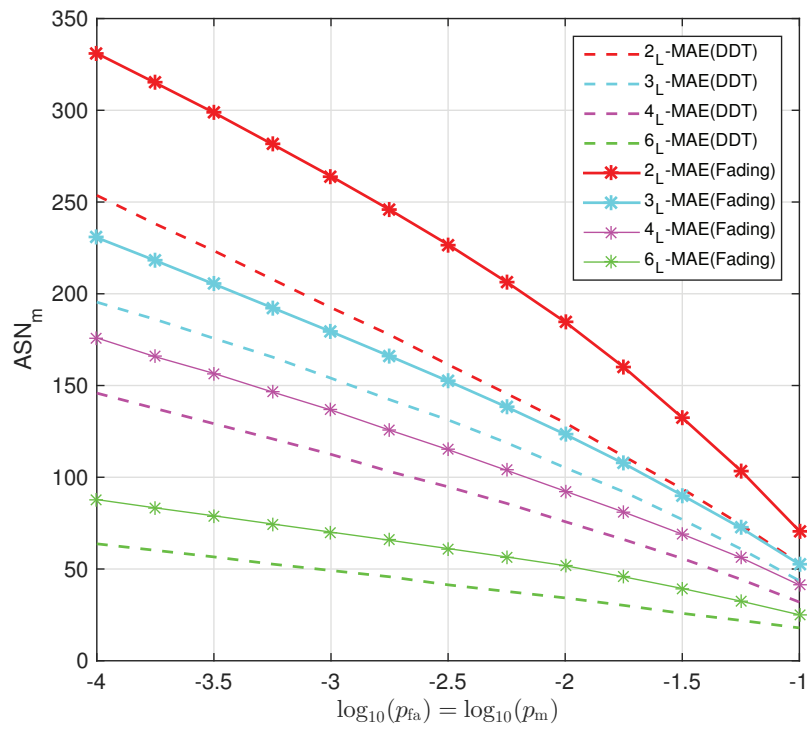


Figure 4.7. Comparison of DDT and non-coherent M-FSK communication over fading channel cases using simulated  $ASN_m$  versus logarithm of false alarm or miss probability for different levels of MAE quantization.



---

**Algorithm 2** SPRT procedure at the fusion center
 

---

- 1: Initialization  $\Lambda_k^c = 0, k=0$
  - 2: Input  $A, \sigma_N^2, p_m, p_{fa}$   $\triangleright$  Where  $A$ - distribution of signal amplitude,  $\sigma_N^2$ - noise variance,  $p_m, p_{fa}$  represent the desired probability of error needed to be achieved
  - 3:  $\eta_0 = \log_{10} \frac{p_m}{p_{1-fa}}, \eta_1 = \log_{10} \frac{1-p_m}{p_{fa}}$   $\triangleright$  Find the values of the upper and lower thresholds hypothesis testing at each sensor node
  - 4:  $H_0 : y_k = \epsilon_k, \quad \triangleright$  hypothesis testing at each sensor node  
 $H_1 : y_k = A_k + \epsilon_k$
  - 5: local likelihood ratio:  $\lambda_k$
  - 6: use Algorithm 1 to find  $\beta_M^*$  thresholds
  - 7:  $\beta_M^* = (\beta_1^*, \dots, \beta_{M-1}^*)$  thresholds
  - 8:  $p_m^{H_0} = \int_{Area_m} p(y|H_0) dy$   $\triangleright$  calculate the probability masses function under hypothesis  $H_0$
  - 9:  $\tilde{p}_m^{H_1} = \int_{Area_m} p(y|H_0) dy$   $\triangleright$  estimate the probability masses function under hypothesis  $H_1$
  - 10:  $q_{\beta_M^*} = \zeta(\log(\lambda_k), \beta_M^*)$
  - 11:  $q_{\beta_M^*} = [q_{\beta_1^*}, \dots, q_{\beta_{M-1}^*}]$
  - 12:  $\Lambda_k^c = \frac{\tilde{p}_M^{H_1}}{p_M^{H_0}}$
  - 13: **if**  $Q_s == q_1$  **then**
  - 14:  $\Lambda_k^c = \frac{\tilde{p}_1^{H_1}}{p_1^{H_0}}$
  - 15: **else if**  $Q_s == q_2$  **then**
  - 16:  $\Lambda_k^c = \frac{\tilde{p}_2^{H_1}}{p_2^{H_0}}$
  - 17: **else if**  $Q_s == q_M$  **then**
  - 18:  $\Lambda_k^c = \frac{\tilde{p}_M^{H_1}}{p_M^{H_0}}$
  - 19: **end if**
  - 20: **while**  $\Lambda_k^c \in (-\eta_0, \eta_1)$  **do**
  - 21:  $k \leftarrow k + 1$
  - 22: wait to receive the next local decision at time  $t_k$  from sensor  $k + 1$
  - 23:  $\Lambda_k^c \leftarrow \Lambda_k^c \times \lambda_k^c$
  - 24: **end while**
  - 25: Stop at time  $\mathcal{S} = t_k$
  - 26: **if**  $\Lambda_k^c \geq \eta_1$  **then**
  - 27:  $\delta = 1$
  - 28:  $ASN_1 = k$   $\triangleright$  decide in favor of hypothesis  $H_1$  with an  $ASN_1 = k$
  - 29: **else**
  - 30:  $\delta = 0$
  - 31:  $ASN_0 = k$   $\triangleright$  decide in favor of hypothesis  $H_0$  with an  $ASN_0 = k$
  - 32: **end if**
-

## CHAPTER 5

# SPATIAL CORRELATION MODELLING FOR SENSOR OBSERVATIONS

### 5.1. Introduction

WSNs are categorized by the densely deployed sensor nodes which continuously observe the physical phenomenon. Many times, sensor observations are highly correlated in the spatial domain due to their high density in the network topology. Additionally, the physical phenomenon establishes a temporal correlation between each consecutive observation of a sensor node. A correlation region is defined as an area where the values sensed by the sensor nodes are considered similar (for the application). According to both application and event type, the size of the correlation region varies. For such events whose characteristics change significantly at a short range, the sink node should decrease the size of the correlated region. On the other hand, for events whose characteristics do not change significantly at short range, the sink node can increase the size of the correlated region. The characteristics of the correlation in the WSN can be summarized as follows:

- I. Spatial Correlation: Typical WSN applications require spatially dense sensor deployment in order to achieve satisfactory coverage (Meguerdichian et al., 2001; Zytoune et al., 2010.) As a result, multiple sensors record information about a single event in the sensor field. Due to the high density in the network topology, spatially proximal sensor observations are highly correlated with the degree of correlation increasing with decreasing inter-node separation.
- II. Temporal Correlation: Some of the WSN applications such as event tracking may require sensor nodes to periodically perform observation and transmission of the sensed event features. The nature of the energy-radiating physical phenomenon constitutes the temporal correlation between each consecutive observation of a sensor node. The degree of correlation between consecutive sensor measurements may

vary according to the temporal variation characteristics of the phenomenon.

For the correlated sensor observations, accounting for the spatial correlation between the sensors, the deterministic signal propagation model which was developed in Chapter 3 is inappropriate because a particular sensor  $k$  will not give us any information about the observation at the  $j$ th sensor and their correlation coefficient defined as

$$\rho(s_k, s_j) = \frac{\text{Cov}[S_k, S_j]}{\sigma_{S_k} \sigma_{S_j}} \quad (5.1)$$

is zero where  $\text{Cov}[S_k, S_j]$  represents the covariance function given as:

$$\text{Cov}[S_k, S_j] = E[S_k, S_j] - E[S_k]E[S_j]. \quad (5.2)$$

For this reason, we have to consider a stochastic signal propagation model in this chapter. For the spatially correlated model, we decide to use the Gaussian signal model which is given in (Vuran and Akyildiz, 2006; Zheng and Tang, 2011).

In Chapter 3, we used a uniform distribution model for deploying the sensors in the surveillance zone. In this model, at each Monte Carlo run the sensor positions were changing, and finding the information content under a certain hypothesis, either  $H_0$  or  $H_1$ , requires obtaining the probability mass functions of the LRs or the observations for the selected thresholds. In the deterministic signal model the sensor observations are independent and the areas under Gaussian pdfs can be calculated analytically using the complementary cumulative distribution function. However, these probability mass functions for the correlated signals are obtained by computationally intensive two-variate bivariate screening (TVBS) and one variate bivariate screening (OVBS) methods (Bhat, 2018), which is not feasible to use in the case of averaging many Monte Carlo realizations of randomly deployed sensors. As a solution, we suggest a grid model for deploying the sensors as shown in Figure 5.1. In this case, the required covariance matrix of the signal component will be always the same and computed once only.

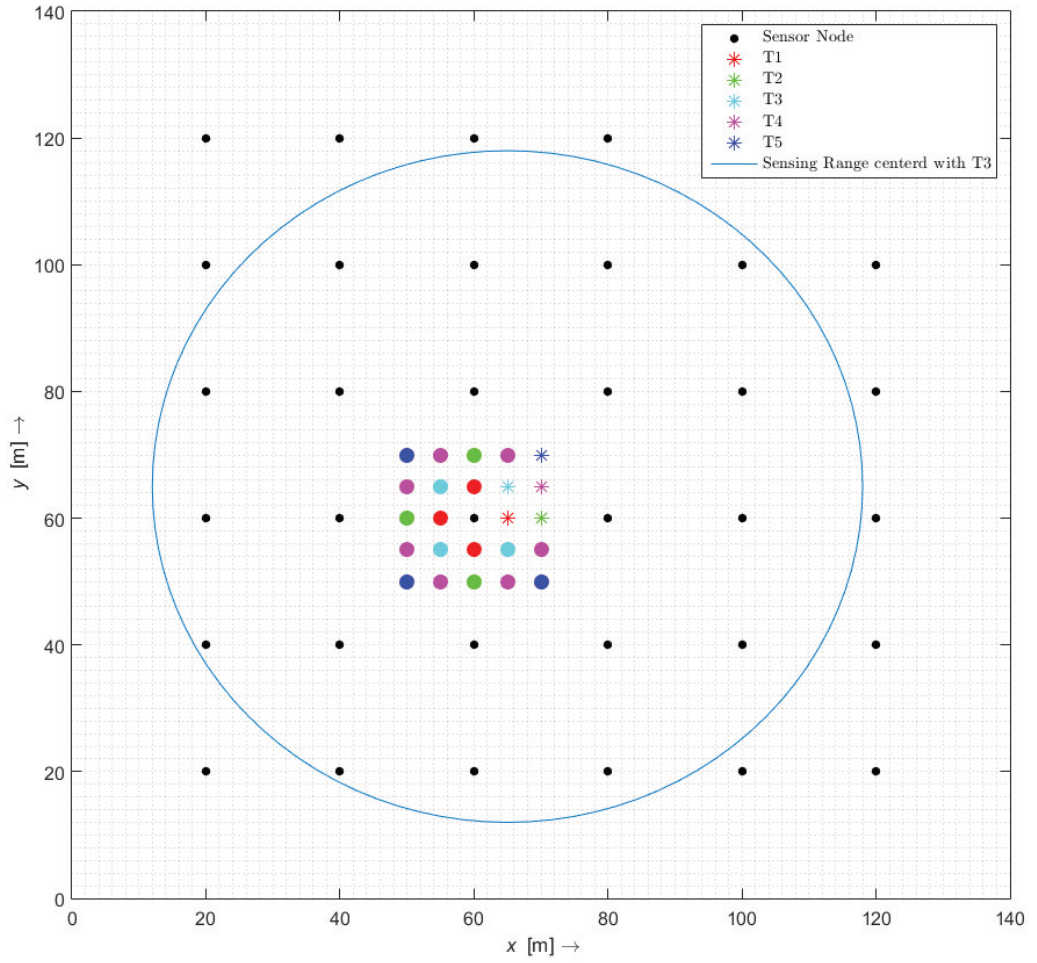


Figure 5.1. The model architecture.

## 5.2. System Model

In this chapter, we will consider a binary hypothesis testing problem, where a group of  $K$  sensors is deployed uniformly in a grid and cooperate with a FC to detect the existence of a point source as shown in Chapter 3 Figure 3.1. The hypothesis testing at each sensor node can be described as

$$\begin{aligned}
 H_0 : y_k &= \epsilon_k, \\
 \text{versus} & \\
 H_1 : y_k &= s_k + \epsilon_k,
 \end{aligned} \tag{5.3}$$

for  $k = 1, 2, \dots, K$  where  $y_k$  denotes the observation at the  $k$ th sensor and  $\epsilon_k$  denotes AWGN with variance  $\sigma_N^2$  and zero mean. The observed signal at sensor  $k$  is denoted as  $s_k$  and for that the two following statistical models are considered:

- I. Independent Gaussian signals: In this case, received signals at the sensors are independent from each other and follow a Gaussian distribution with zero mean and variance  $\sigma_{s_k}^2$ . So,  $y_k \sim \mathcal{N}(0, \sigma_N^2)$  under  $H_0$  and  $y_k \sim \mathcal{N}(0, \sigma_{s_k}^2 + \sigma_N^2)$  under  $H_1$ , where  $y_k$  are independently distributed. By using the NP detector the decision will be in favor of hypothesis  $H_1$  if

$$\Lambda(\mathbf{y}) = \frac{p(\mathbf{y}|H_1)}{p(\mathbf{y}|H_0)} = \frac{\frac{1}{(2\pi)^{\frac{K}{2}} \prod_{k=1}^K (\sigma_{s_k}^2 + \sigma_N^2)^{\frac{1}{2}}} \exp\left(-\frac{1}{2} \sum_{k=1}^K \frac{y_k^2}{(\sigma_{s_k}^2 + \sigma_N^2)}\right)}{\frac{1}{(2\pi\sigma_N^2)^{\frac{K}{2}}} \exp\left(-\frac{1}{2} \sum_{k=1}^K \frac{y_k^2}{\sigma_N^2}\right)} \underset{H_0}{\overset{H_1}{\geq}} \eta. \quad (5.4)$$

We can re-write (5.4) as

$$\Lambda(\mathbf{y}) = \left( \frac{\sigma_N^K}{\prod_{k=1}^K (\sigma_{s_k}^2 + \sigma_N^2)^{\frac{1}{2}}} \right) \exp\left( \frac{1}{2} \sum_{k=1}^K y_k^2 \left[ \frac{1}{\sigma_N^2} - \frac{1}{\sigma_{s_k}^2 + \sigma_N^2} \right] \right) \underset{H_0}{\overset{H_1}{\geq}} \eta. \quad (5.5)$$

By taking the natural logarithm of both sides we obtain:

$$\sum_{k=1}^K \left[ \frac{1}{\sigma_N^2} - \frac{1}{\sigma_{s_k}^2 + \sigma_N^2} \right] y_k^2 \underset{H_0}{\overset{H_1}{\geq}} 2 \log \left( \frac{\prod_{k=1}^K (\sigma_{s_k}^2 + \sigma_N^2)^{\frac{1}{2}}}{\sigma_N^K} \eta \right). \quad (5.6)$$

Note that, in our previous discussion, in Chapter 3, the decision statistics is a linear function of the observations, and any linear combination of a Gaussian random variable is also Gaussian. We perform the full characterization of the decision statistic for this case using the mean and the variance for the Gaussian random variable  $y_i$ . However, in this chapter, we have the summation of the squares of the random variable  $y_i$ , which is not Gaussian.

- II. Correlated Gaussian signals: In this case, we will consider the general covariance structure for the signal vector  $\mathbf{s} = [s_1, s_2, \dots, s_K]^T$  which is  $E(\mathbf{s}\mathbf{s}^T) = \mathbf{C}_s$ .

By reference to (5.3) we can see that the observed signal  $\mathbf{y}$  under  $H_0$  is a pure noise and it has the covariance structure  $E(\mathbf{y}\mathbf{y}^\top|H_0) = \sigma_N^2 \mathbf{I}_K$  where  $\mathbf{y} = [y_1, y_2, \dots, y_K]^\top$ . Under  $H_1$ , we have a summation of signal and noise components which are statistically independent. Consequently, the covariance statistic will be given as  $E(\mathbf{y}\mathbf{y}^\top|H_1) = \mathbf{C}_s + \sigma_N^2 \mathbf{I}_K$ . By using the NP detector the decision will be in favor of hypothesis  $H_1$  if

$$\frac{p(\mathbf{y}|H_1)}{p(\mathbf{y}|H_0)} = \frac{\sigma_N^K}{\sqrt{\det(\mathbf{C}_s + \sigma_N^2 \mathbf{I}_K)}} \exp\left(\frac{-1}{2} \mathbf{y}^\top (\sigma_N^{-2} \mathbf{I}_K - [\mathbf{C}_s - \sigma_N^2 \mathbf{I}_K]^{-1}) \mathbf{y}\right) \geq \eta. \quad (5.7)$$

By taking the natural logarithm of (5.7)

$$\mathbf{y}^\top (\sigma^{-2} \mathbf{I}_K - [\mathbf{C}_s - \sigma_N^2 \mathbf{I}_K]^{-1}) \mathbf{y} \geq \tilde{\eta}, \quad (5.8)$$

where  $\tilde{\eta} = 2 \log \frac{\eta}{\sigma_N} \sqrt{\mathbf{C}_s + \sigma_N^2 \mathbf{I}_K}$

### 5.3. Sensor Deployment Model

The WSN model considered in this chapter consists of  $K$  sensors deployed in a grid over the surveillance zone as shown in Figure 5.1. In order to make the results independent from the target positions, an averaging over the possible target positions is performed. As an example, the zone inside the circle indicates the event zone for the occurrence of the target  $T_3$ ; black nodes represent sensor nodes; colored nodes represent all the possible locations for the event occurrence. Notice that each set from the same color is shared by a specific number of sensors. E.g. the target with blue color is shared by 4 sensors, the target with magenta and green color are shared by 2 sensors whereas the target colors red and cyan are not shared by more than one sensor. General assumptions in the deployment of our WSN are summarized as follows:

- i. All the sensors are following the Boolean coverage model, which means each sensor has a fixed sensing range and the sensing zone is represented with a disc centered by the spatial location of that sensor.
- ii. Each sensor has a communication range in order to send its decision to the FC.
- iii. The sensors are considered to be static (no movement) which means that the location information for each sensor is known.

## 5.4. Quantizer Design

To make a global decision at the FC we use a similar assumption to what we have proposed in Section 3.3, that is each sensor transmits a single binary observation to the FC using an identical quantization threshold for detection. However this time following a variance-based partitioning of pdfs under each hypothesis as shown in Figure 5.2. The quantization thresholds are evaluated using MAE and MJD methods discussed previously in Section 3.3. Figure 5.3 shows the entropy functions  $\hat{F}_{H_0}$ ,  $\hat{F}_{H_1}$  and  $\hat{F}_{av}$  for binary quantization considering spatially correlated signals where the optimum threshold is found to be at  $\beta^* = 1$ . On the other hand, Figure 5.4 shows the corresponding maximum of  $\hat{J}$  which is found to be at  $\beta^\diamond = 2.2$  for binary quantization considering the same signals.

Similar to what we have obtained in the 1-D case in Figure 3.4, now, in this chapter we obtain multi-dimensional pmfs for  $K$  correlated signals under both hypotheses  $H_1$  and  $H_0$  using analytic approximation techniques for multivariate normal cumulative density (MVNCD). These techniques depend on a single-sided truncation of a multivariate normal distribution in which some variables are truncated while others are not. We are mainly using TVBS and OVBS which generally come out well on top relative to other methods according to the ability of evaluating the individual choice probabilities and computational time (Bhat, 2018). We perform a further approximation to find the pmfs using 4 sub-matrices instead of using only one large matrix, e.g. instead of using a matrix of dimensions  $(22 \times 22)$ , we use a diagonal matrix consisting of 4 sub-matrices  $(6 \times 6)$ ,

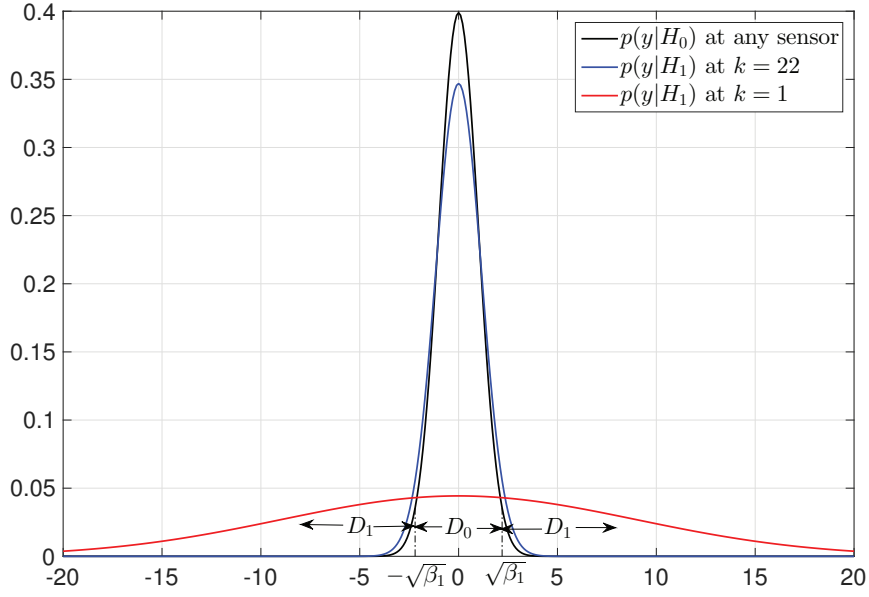


Figure 5.2. Pdfs of the observations when there is a target at  $T_3$  position and binary decision regions for hypothesis test based on change in variance ( $D_0$ : Decision Region for  $H_0$ ,  $D_1$ : Decision Region for  $H_1$ )

$(6 \times 6)$ ,  $(6 \times 6)$  and  $(4 \times 4)$  as shown in Figure 5.5 and neglect the other elements of the  $(22 \times 22)$  covariance-matrix. Also a visualization for the covariance matrix and its approximation by 4 smaller dimensional covariance matrices is given in Figure 5.6.

This simplification has been performed because the evaluation of the multi-dimensional pmfs for such a large dimensional matrix is not possible for a vector of thresholds. For the information theoretic criteria used in the thesis, namely MAE and MJD, the most costly part computationally is finding the pmf values. So, their computational costs are very similar.

The sub-diagonal matrices (covariance matrices) were obtained by considering the spatial correlation between the sensors themselves (depending on the inter sensor distances,  $d(s_i, s_j)$ ) and between the sensors and the event source (depending on the distance from the target,  $d(s_i, T_j)$ ) which is given by Algorithm 3 as composed of two steps:

1.  $\text{Max}(\text{diagonal}(C_s)); \text{Max}([\sigma_{11}^2, \sigma_{22}^2, \dots, \sigma_{kk}^2])$
2. Score of each sensor =  $\sigma_{ii}^2 \rho_{s,T} + \sum_j^K (\sigma_i \sigma_j) \rho_{i,j}$



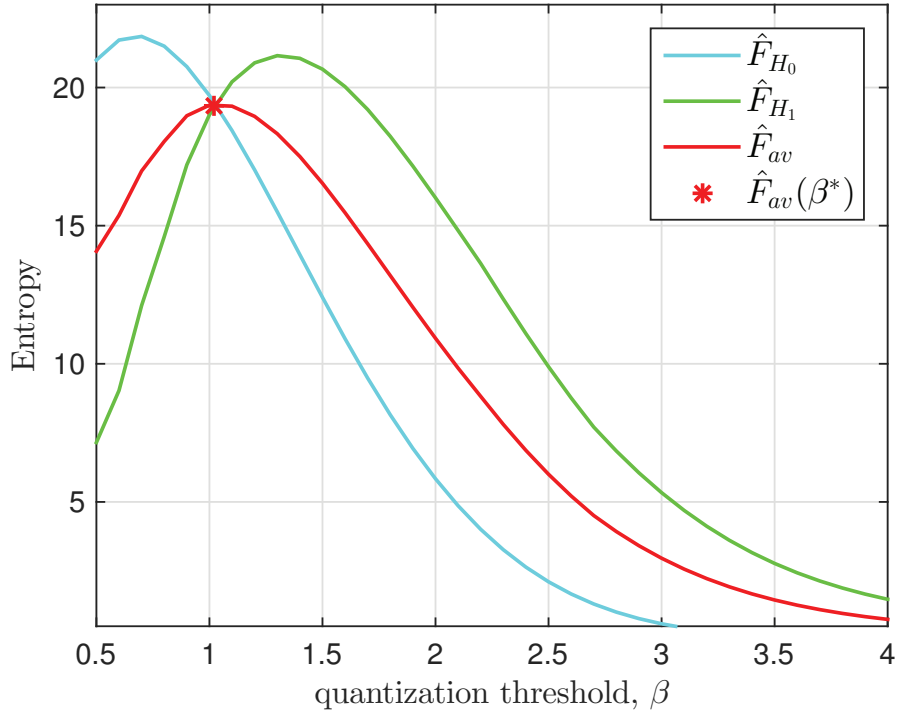


Figure 5.3. The entropy functions  $\hat{F}_{H_0}$ ,  $\hat{F}_{H_1}$  and  $\hat{F}_{av}$  for binary quantization of spatially correlated signals.

where  $\rho_{k,l}$  represents the correlation coefficient in the exponential correlation model which is given as:

$$\rho_{k,l} = \exp(-d/\theta_1)^{\theta_2} \quad (5.9)$$

with  $\theta_1 > 1$  used to control the correlation between the observations of the sensor and  $\theta_2 = 1$  for the exponential correlation model. Moreover, Figure 5.7 shows the sensors which correspond to each sub-matrix; the set of sensors, 1, 2, ..., 6 determines the first sub-matrix. Similarly, the set of sensors 7, 8, ..., 12, the set of sensors 13, 14, ..., 18 and the set of sensors 19, 20, 21, 22 determine the second, the third and the last sub-matrix, respectively, in Figure 5.5.

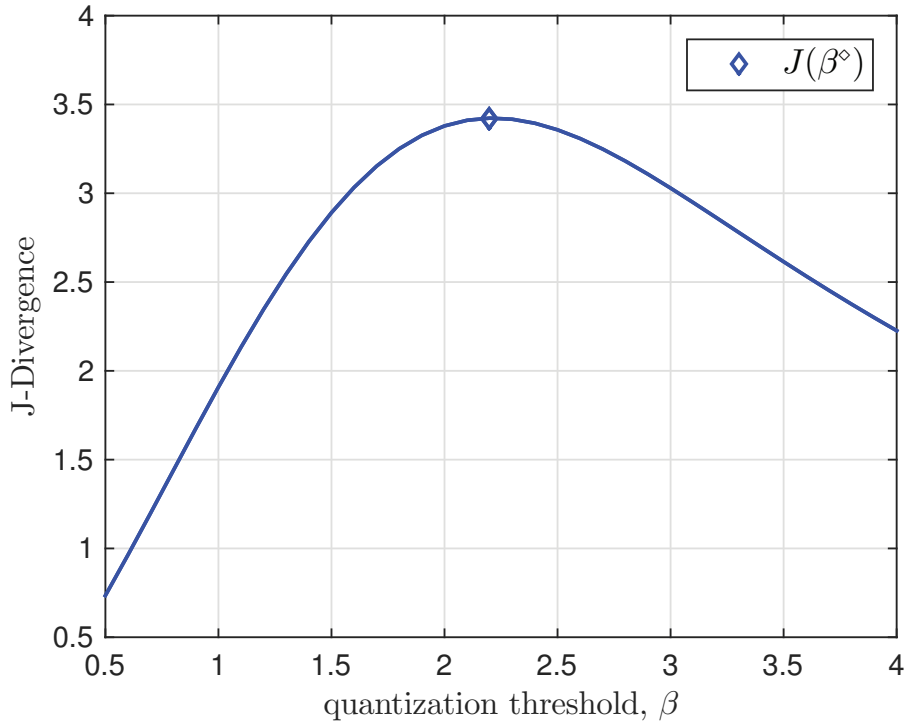


Figure 5.4. The J-divergence for binary quantization of spatially correlated signals.

## 5.5. Simulations for the Spatial Correlation Model

Monte Carlo simulations have been performed to evaluate the detection performance at SNR= 12 dB in the cases of both independent and correlated Gaussian signals which are described by the system model given in Subsection 5.2. In Figure 5.8, the ROCs are plotted for the case of independent but not identical Gaussian observations. The ROCs are evaluated using MAE and MJD methods for two levels of quantization. Two other ad hoc thresholds are provided for the comparison with MAE and MJD. In this figure, we observe that MJD quantization method outperforms MAE method and each of them performs better than the ad hoc thresholds for  $p_{FA} > 0.4$ . The ad hoc threshold value  $\beta_{ad hoc1} = 3$  performs the best for  $p_{FA} < 0.3$ .

Figure 5.9 shows the ROC curves for the correlated Gaussian signals using MAE and MJD based methods and the other two ad hoc thresholds. This figure depicts that MJD optimized threshold  $\beta^\diamond = 2.2$  and the first ad hoc threshold  $\beta_{ad hoc1} = 3$  exhibit the best performances and MAE optimized threshold  $\beta^* = 1$  performs better compared to the

---

**Algorithm 3** The procedure of choosing the sub-diagonal matrices.

---

```

1: Input  $Dim = 36$ ; ▷ number of the sensors in 5.1.
2:  $ID =$  vector with dimension  $(1 \times M)$  which includes the chosen sensors where,  $M = 1, 2, \dots, 5$ .
3:  $Covar = \mathbf{C}_s + \sigma_N^2 \mathbf{I}_K$ 
4: Initialize score =  $Zeros(1, Dim)$ 
5: for  $i = 1 : Dim$  do
6:   for  $i = 1 : Dim$  do
7:     continue
8:   end for
9:    $Score(i) = Covar(i, i)$ 
10:   $Score(i) = Score(i) + Covar(i, index(1))$ 
11:    $\vdots$ 
12:   $Score(i) = Score(i) + Covar(i, index(M))$ 
13: end for
14:  $index(M + 1) = \text{Max}(\text{score})$ ;

```

---

second ad hoc threshold  $\beta_{adhoc2} = 0.5$ . While  $\beta_{adhoc1} = 3$  is slightly more successful for  $p_{FA} < 0.35$ ,  $\beta^\diamond$  outperforms it for  $p_{FA} > 0.6$ .

Figure 5.10 shows a comparison between the resulted ROC curves using the uncorrelated and the correlated signals. This figure shows that the performance degrades for the correlated observations.

## 5.6. Conclusions on Spatially Correlated Observations

The spatial correlation of the sensors was considered in this chapter for the detection of point sources. For this case, a Gaussian isotropic event source was applied. The computational requirements in evaluating multidimensional cumulative densities necessitated proposing a rectangular grid model of sensor deployment and block-diagonal approximations of covariance matrix related to the event signal at the sensors without losing generality. The detector for the correlated signal case is an energy detector and the difference of the nature of the likelihood ratio type estimators seemed to be counting on the disadvantage of the MAE quantizer for this case. As a result, for spatially correlated signals MJD was more successful compared to MAE.

89246	14176	14176	7229	8791	8791	5259	4337	2535	1896	1896	1679	1679	1980	1980	1522	1522	847	1402	1402
14176	25849	4337	4725	2690	1572	3438	1107	913	1801	1053	1594	436	713	354	845	291	652	364	268
14176	4337	25849	4725	1572	2690	1107	3438	913	1801	1053	1594	436	713	354	291	845	652	268	364
7229	4725	4725	18805	944	944	1025	1025	395	1445	1445	1962	559	227	227	252	252	1032	133	133
8791	2690	1572	944	16096	1668	2132	687	1758	566	281	305	225	578	199	1373	442	896	245	115
8791	1572	2690	944	16096	1668	687	2132	1758	281	566	225	305	199	578	442	1373	245	896	115
5259	4337	1107	1025	2132	687	14662	349	615	1052	212	459	228	1263	116	821	165	1146	105	151
5259	1107	3438	1025	687	2132	349	14662	615	212	1052	228	459	116	1263	165	821	105	1146	151
4337	913	913	395	1758	1758	615	615	13170	174	174	108	108	166	677	677	304	304	46	870
2535	2835	913	1445	566	281	1052	212	174	13170	237	1177	379	1042	95	185	63	304	54	362
2535	913	2835	1445	566	281	212	1052	174	237	13170	379	1177	95	1042	63	185	54	304	362
1896	1801	1053	1962	305	225	459	228	108	1177	379	12733	747	387	133	84	52	121	56	840
1896	1053	1801	1962	305	225	228	459	108	379	1177	747	12733	133	387	52	84	56	121	840
1679	1594	436	559	578	578	1263	116	166	1042	95	387	133	12142	43	262	46	531	32	119
1679	436	1594	559	199	578	116	1263	166	95	1042	133	387	43	12142	46	262	32	531	119
1980	713	354	227	1373	442	821	165	677	185	63	84	52	262	46	11933	144	737	67	29
1980	354	713	227	442	1373	165	821	677	63	185	52	84	46	262	144	11933	67	737	29
1522	845	291	252	896	245	1146	105	304	304	54	121	56	531	32	737	67	1176	36	38
1522	291	845	252	245	896	105	1146	304	54	304	56	121	327	531	67	737	36	11761	38
847	652	652	1032	115	115	151	151	46	362	362	840	840	119	119	29	29	38	11617	15
1402	364	268	133	826	483	340	168	870	79	49	41	33	98	45	679	219	259	89	15
1402	268	364	133	483	826	168	340	870	49	79	33	41	45	98	219	679	89	259	15

Figure 5.5. Covariance matrix  $\times 10^4$  of sensor observations for the grid sensor network with event location  $T_3$  as shown in Figure 5.7.

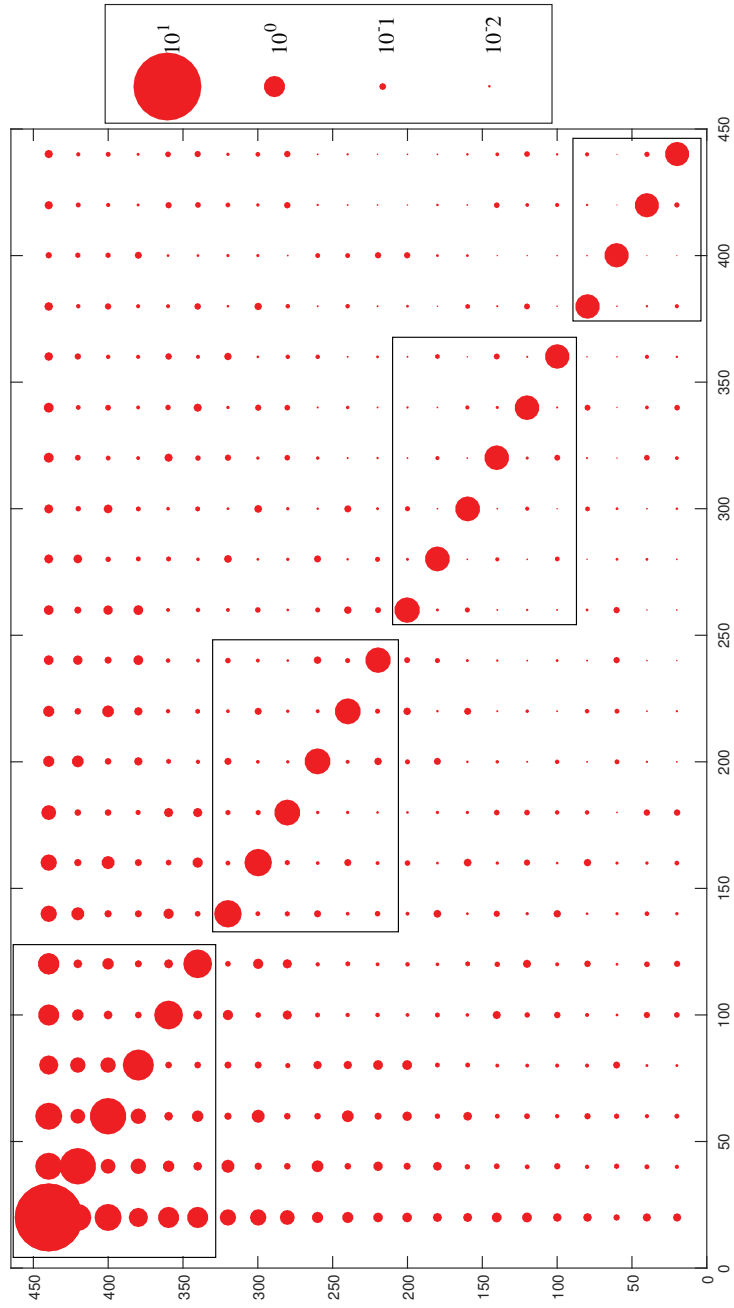


Figure 5.6. Visualization for the covariance matrix shown in Figure 5.5

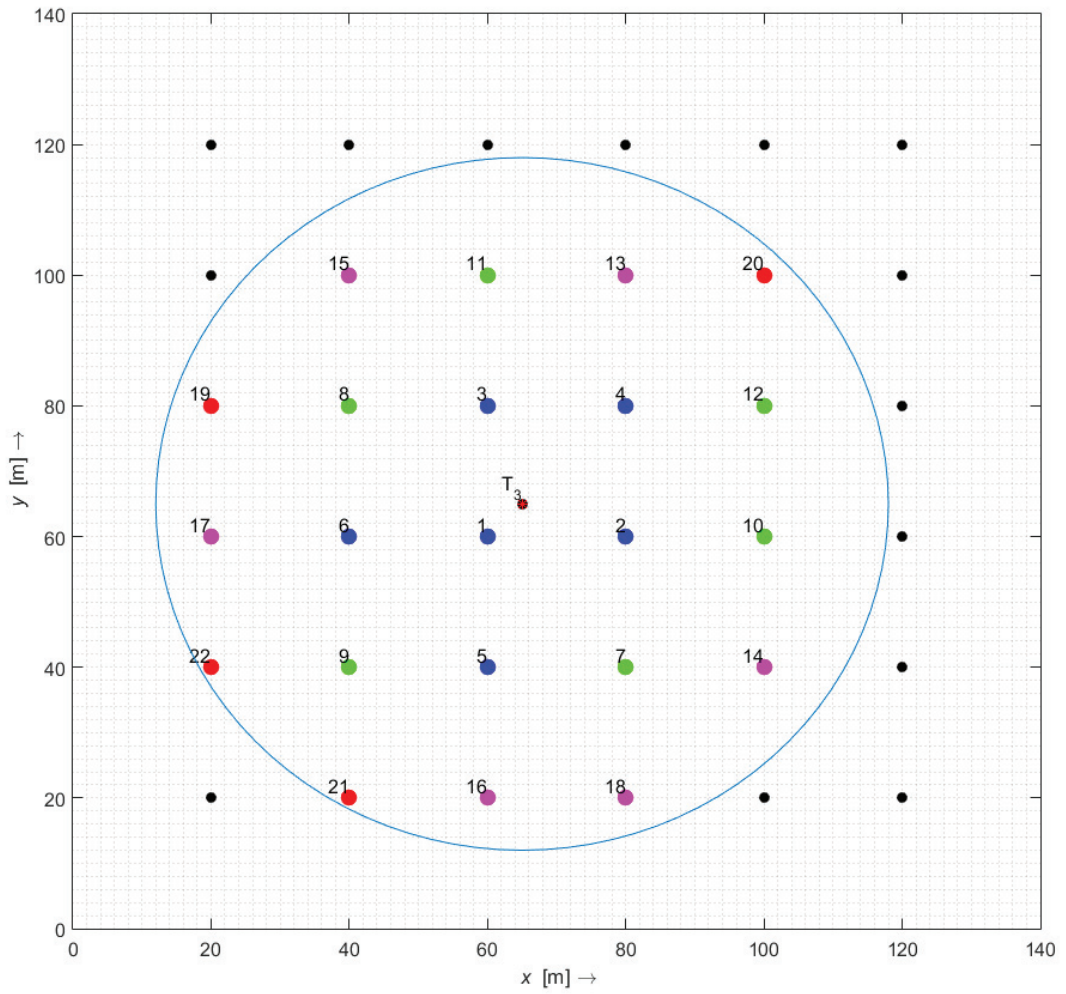


Figure 5.7. Spatially correlated sensors in the sensitivity circle of target  $T_3$  (colors of sensors: Algorithm 3 based groups, numbers: decreasing order of covariances of sensors)

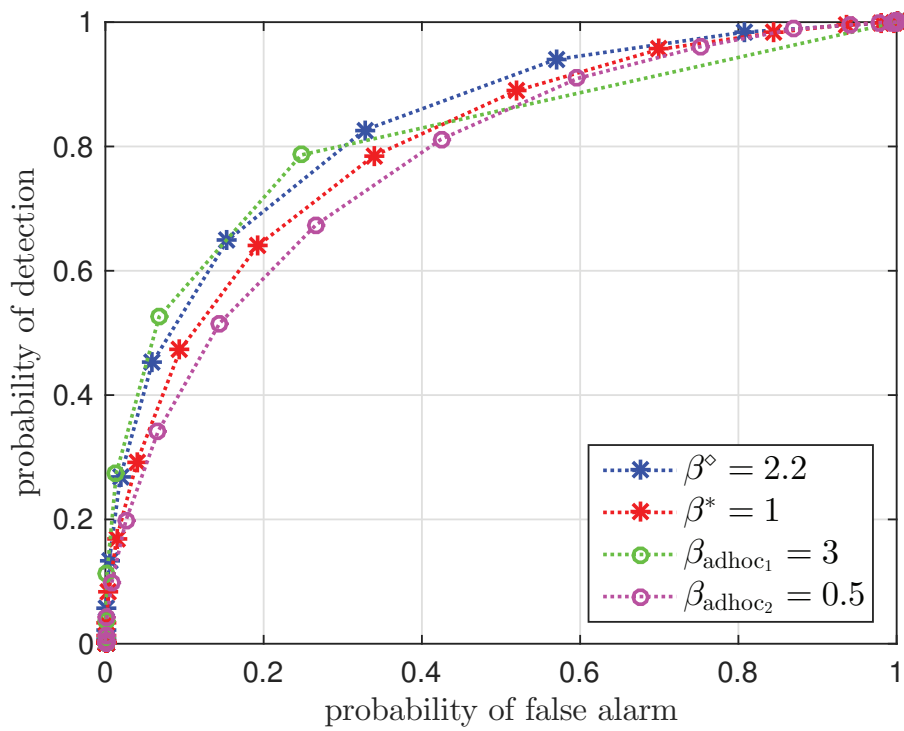


Figure 5.8. ROC curve for the independent Gaussian signals ( $\theta_1 = 21.83$ , SNR = 12 dB (w.r.t. a sensor at 5 unit distance)).

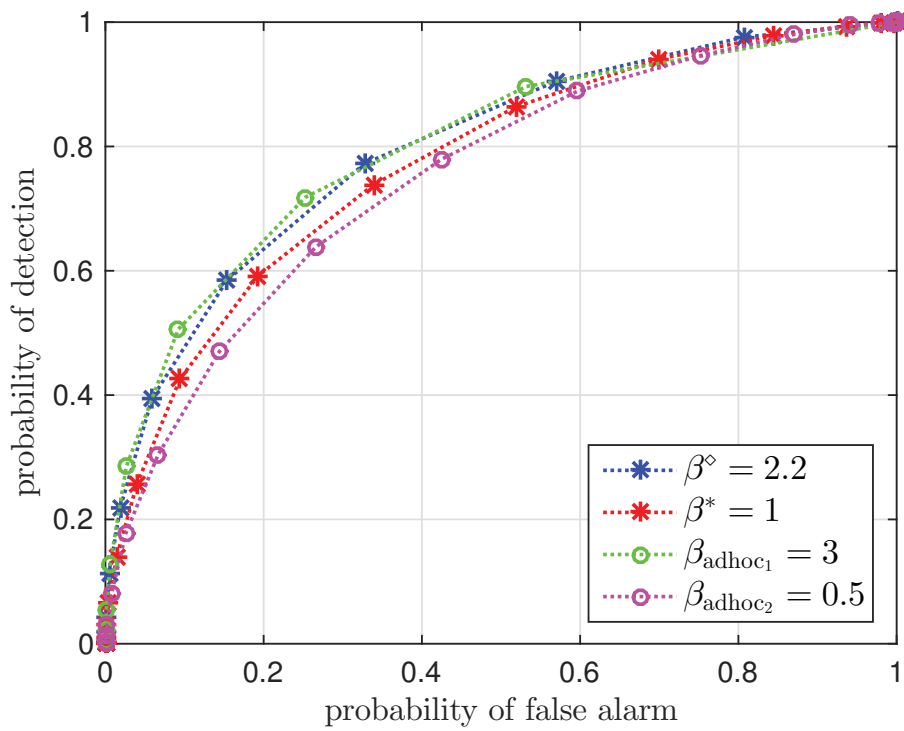


Figure 5.9. ROC curve for the correlated Gaussian signals ( $\theta_1 = 21.83$ , SNR = 12 dB (w.r.t. a sensor at 5 unit distance)).



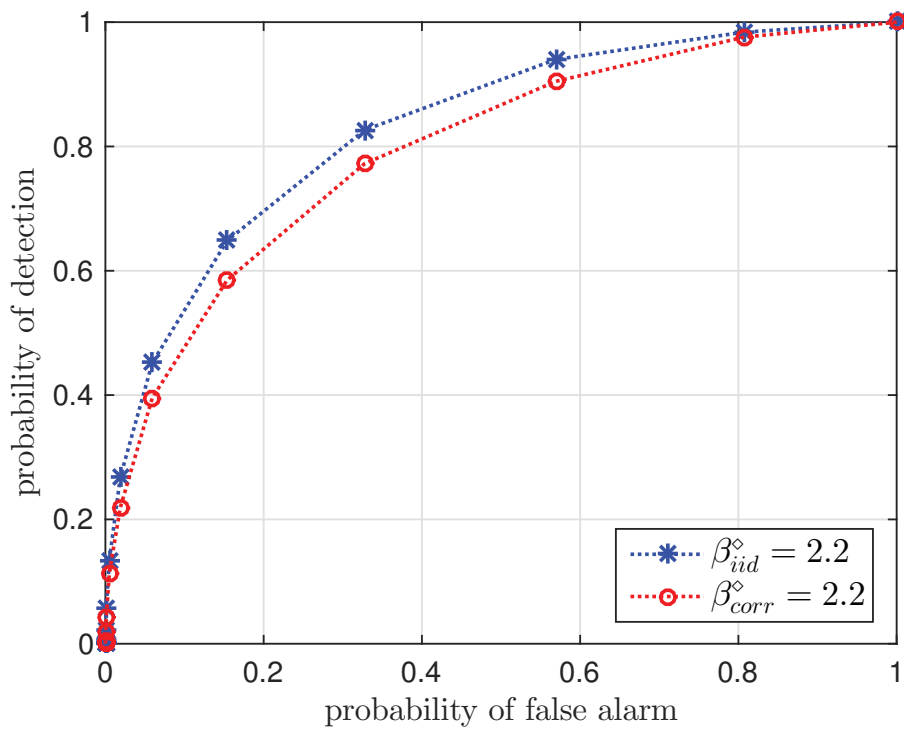


Figure 5.10. Comparison between i.i.d. and correlated observations for MJD method ( $\theta_1 = 21.83$  SNR = 12, dB (w.r.t. a sensor at 5 unit distance)).

## CHAPTER 6

# CONCLUSIONS AND SUGGESTIONS FOR FUTURE RESEARCH

### 6.1. Conclusions

In this thesis, we have proposed quantizing the sensor outputs by maximizing their average information in the cases of presence and non-presence of an event in DD. The general approach in quantization for decision processes is based on distance measures such as JD and Bhattacharyya distance. This fact may have prevented a popular information based quantization criterion for decision processes maximizing the information under both (all) hypotheses rather than the information in the difference of the distributions. Since among the distance measure based quantization approaches, JD is an information-theoretic quality, we adopted JD for comparisons of the proposed method.

One reason for suggesting another method like MAE instead of MJD is the non-symmetric nature of the considered problem and the fact that the advantage of Ali-Silvey type criteria (Poor and Thomas, 1977) which MJD is a member of, is only valid for the symmetric performance measure probability of error. Although maximizing the transferred information under each hypothesis as proposed by the MAE method is a conceptually different approach, we showed that average entropy and JD are positively proportional quantities. This means that one might expect comparable performances using either of them for determining the quantization levels which was indeed the observation in the simulation results.

In order to concentrate on the effects of how the sensor outputs are quantized on the system performance, we performed extensive simulation studies for the case that the sensor outputs are available error-free at the FC which we called DDT. The performances of considered information-based methods, namely MAE and MJD, gradually improved as the quantization level was increased from binary to six-levels and it approached the per-

formance of non-quantized data transmission. Additionally, the proposed method, MAE, performed significantly better compared to MJD for any level of quantization. Also, the effects of the Rayleigh fading channel from the sensors to the FC have been investigated using the optimal and a suboptimal fusion rule for MAE. Due to the power efficiency and small degradation in non-coherent communication MFSK was adopted as the modulation scheme for the sensor to FC communication. Using the wireless channel model similar results were obtained as in DDT. Results with 6-level quantization were comparable to non-quantized data transmission.

In this thesis, we have shown that the sequential detection outperforms the non-sequential detection in the ASN needed to achieve a specific probability of detection in the deterministic isotropic signal case with unknown SNR which is an expected result. Also, we noticed that interestingly, when sequential detection is applied, for specific values of detection, the number of transmitted bits for  $M$ -level quantization are less than the number of bits transmitted for binary quantization. For different types of quantization levels, we found that ASN decreases monotonically by increasing the quantization levels and ASN increases by increasing the target probability of detection.

The spatial correlation of the sensors is taken into the account. For this case, a Gaussian isotropic event source was applied. The computational requirements in evaluating multidimensional cumulative densities necessitated proposing a rectangular grid model of sensor deployment and block-diagonal approximations of covariance matrix related to the event signal at the sensors without losing generality.

This work showed that MAE is a valid and promising method in quantization for detection problems.

## 6.2. Suggestions for Future Research

Several research issues can be suggested based on the results of this thesis. Those are outlined below:

1. We use MAE and MJD quantization methods for binary hypothesis test. Additionally, those quantization methods can be applied to discriminate between  $M$  hypotheses, where  $M > 2$ . This type of problem is important, in which one of  $M$

signals need to be detected. Also, it arises frequently in pattern recognition systems to distinguish between different patterns.

2. This thesis considers a parallel network topology, whereas in WSNs, there are other network topologies to investigate such as tree and serial topologies.
3. Applying the MAE method to quantize the overshoot of the LR, that is, the amount by which the LR exceeds the lower or the upper thresholds of the SPRT in (Yilmaz et al., 2012b) and comparing it with non-uniform samplers which are named as event-triggered samplers. These type of quantization methods are very important to improve the performance of SPRT, e.g. in cognitive radio spectrum sensing, we have to decide if the band is used by the primary user or not, especially in time-slotted system when the decision gets late the available time for actual data transmissions will be very small and so, it would be better to make the decision as fast as possible. Moreover, in radar system applications, it is very important to decide if there is a target and start dealing with this target, i.e. start tracking this target.
4. In this thesis, we consider the cases, where the observations are i.i.d. and spatially correlated at local sensors. However, the observations can be also temporally correlated in target tracking applications. Therefore, one rich area for research is to develop proper quantization methods for temporally correlated signals.

## REFERENCES

- Al-Jarrah, M. A., A. Al-Dweik, M. Kalil, and S. S. Ikki (2018). Decision fusion in distributed cooperative wireless sensor networks. *IEEE Transactions on Vehicular Technology* 68(1), 797–811.
- Altay, C. and H. Delic (2016). Optimal quantization intervals in distributed detection. *IEEE Transactions on Aerospace and Electronic Systems* 52(1), 38–48.
- Baum, C. W. and V. V. Veeravalli (1994). A sequential procedure for multihypothesis testing. *IEEE Transactions on Information Theory* 40(6).
- Berger, J. O. (2013). *Statistical Decision Theory and Bayesian Analysis*. Springer Science & Business Media.
- Bhat, C. R. (2018). New matrix-based methods for the analytic evaluation of the multivariate cumulative normal distribution function. *Transportation Research Part B: Methodological* 109, 238–256.
- Blum, R. S. and M. C. Deans (1998). Distributed random signal detection with multibit sensor decisions. *IEEE Transactions on Information Theory* 44(2), 516–524.
- Chair, Z. and P. Varshney (1986a). Optimal data fusion in multiple sensor detection systems. *IEEE Transactions on Aerospace and Electronic Systems* (1), 98–101.
- Chair, Z. and P. Varshney (1986b). Optimal data fusion in multiple sensor detection systems. *IEEE Transactions on Aerospace and Electronic Systems* (1), 98–101.
- Chaudhari, S., V. Koivunen, and H. V. Poor (2009). Autocorrelation-based decentralized sequential detection of ofdm signals in cognitive radios. *IEEE Transactions on Signal Processing* 57(7), 2690–2700.
- Chen, B., R. Jiang, T. Kasetkasem, and P. K. Varshney (2002). Fusion of decisions transmitted over fading channels in wireless sensor networks. In *Conference Record of the*

*Thirty-Sixth Asilomar Conference on Signals, Systems and Computers, 2002.*, Volume 2, pp. 1184–1188. IEEE.

Chen, B., R. Jiang, T. Kasetkasem, and P. K. Varshney (2004). Channel aware decision fusion in wireless sensor networks. *IEEE Transactions on Signal Processing* 52(12), 3454–3458.

Chen, B. and P. K. Willett (2005). On the optimality of the likelihood-ratio test for local sensor decision rules in the presence of nonideal channels. *IEEE Transactions on Information Theory* 51(2), 693–699.

Cheng, Q., P. K. Varshney, K. G. Mehrotra, and C. K. Mohan (2005). Bandwidth management in distributed sequential detection. *IEEE Transactions on Information Theory* 51(8), 2954–2961.

Ciuonzo, D., G. Romano, and P. S. Rossi (2012). Optimality of received energy in decision fusion over rayleigh fading diversity mac with non-identical sensors. *IEEE Transactions on Signal Processing* 61(1), 22–27.

Ciuonzo, D. and P. S. Rossi (2017). Distributed detection of a non-cooperative target via generalized locally-optimum approaches. *Information Fusion* 36, 261–274.

DeGroot, M. H. (2005). *Optimal Statistical Decisions*, Volume 82. John Wiley & Sons.

Ekchian, L. K. and R. R. Tenney (1982). Detection networks. In *1982 21st IEEE Conference on Decision and Control*, pp. 686–691. IEEE.

Fellouris, G. and G. V. Moustakides (2010). Decentralized sequential hypothesis testing using asynchronous communication. *IEEE Transactions on Information Theory* 57(1), 534–548.

Ferguson, T. S. (2014). *Mathematical Statistics: A Decision Theoretic Approach*, Volume 1. Academic press.

- Hajibabaei, Z. and A. Vosoughi (2014). Impact of wireless channel uncertainty upon m-ary distributed detection systems. In *2014 IEEE 25th Annual International Symposium on Personal, Indoor, and Mobile Radio Communication (PIMRC)*, pp. 692–696. IEEE.
- Hill, J., M. Horton, R. Kling, and L. Krishnamurthy (2004). The platforms enabling wireless sensor networks. *Communications of the ACM* 47(6), 41–46.
- Hoballah, I. Y. and P. K. Varshney (1989). Distributed bayesian signal detection. *IEEE Transactions on Information Theory* 35(5), 995–1000.
- Kay, S. M. (1993). *Fundamentals of Statistical Signal Processing*. Prentice Hall PTR.
- Kotecha, J. H., V. Ramachandran, and A. M. Sayeed (2005). Distributed multitarget classification in wireless sensor networks. *IEEE Journal on Selected Areas in Communications* 23(4), 703–713.
- Kowalski, C. J. (1971). The oc and asn functions of some sprt's for the correlation coefficient. *Technometrics* 13(4), 833–841.
- Lee, C.-C. and J.-J. Chao (1989). Optimum local decision space partitioning for distributed detection. *IEEE Transactions on Aerospace and electronic systems* 25(4), 536–544.
- Levy, B. C. (2008). *Principles of Signal Detection and Parameter Estimation*. Springer Science & Business Media.
- Liu, B., A. Jeremic, and K. M. Wong (2011). Optimal distributed detection of multiple hypotheses using blind algorithm. *IEEE Transactions on Aerospace and Electronic Systems* 47(1), 317–331.
- Maleki, N. and A. Vosoughi (2012). Channel-aware m-ary distributed detection: Optimal and suboptimal fusion rules. In *2012 IEEE Statistical Signal Processing Workshop (SSP)*, pp. 644–647. IEEE.

- Meguerdichian, S., F. Koushanfar, M. Potkonjak, and M. B. Srivastava (2001). Coverage problems in wireless ad-hoc sensor networks. In *Proceedings IEEE INFOCOM 2001. Conference on Computer Communications. Twentieth Annual Joint Conference of the IEEE Computer and Communications Society (Cat. No. 01CH37213)*, Volume 3, pp. 1380–1387. IEEE.
- Messerschmitt, D. (1971). Quantizing for maximum output entropy (corresp.). *IEEE Transactions on Information Theory* 17(5), 612–612.
- Muduli, L., D. P. Mishra, and P. K. Jana (2018). Application of wireless sensor network for environmental monitoring in underground coal mines: A systematic review. *Journal of Network and Computer Applications* 106, 48–67.
- Niu, R., B. Chen, and P. K. Varshney (2006). Fusion of decisions transmitted over rayleigh fading channels in wireless sensor networks. *IEEE Transactions on Signal Processing* 54(3), 1018–1027.
- Niu, R. and P. K. Varshney (2006). Joint detection and localization in sensor networks based on local decisions. In *2006 Fortieth Asilomar Conference on Signals, Systems and Computers*, pp. 525–529. IEEE.
- Niu, R. and P. K. Varshney (2007). Performance analysis of distributed detection in a random sensor field. *IEEE Transactions on Signal Processing* 56(1), 339–349.
- Poor, H. and J. Thomas (1977). Applications of ali-silvey distance measures in the design generalized quantizers for binary decision systems. *IEEE Transactions on Communications* 25(9), 893–900.
- Poor, H. V. (1983). A companding approximation for the statistical divergence of quantized data. In *The 22nd IEEE Conference on Decision and Control*, pp. 697–702. IEEE.
- Poor, H. V. (1988). Fine quantization in signal detection and estimation. *IEEE Transactions on Information Theory* 34(5), 960–972.



- Poor, H. V. (2013). *An Introduction to Signal detection and Estimation*. Springer Science & Business Media.
- Ragam, P. and N. D. Sahebraoji (2019). Application of mems-based accelerometer wireless sensor systems for monitoring of blast-induced ground vibration and structural health: A review. *IET Wireless Sensor Systems* 9(3), 103–109.
- Rajput, A. and V. B. Kumaravelu (2019). Scalable and sustainable wireless sensor networks for agricultural application of internet of things using fuzzy c-means algorithm. *Sustainable Computing: Informatics and Systems* 22, 62–74.
- Tanataratana, S. and J. B. Thomas (1977a). On sequential sign detection of a constant signal. *IEEE Transactions on Information Theory* 23(3), 304–315.
- Tanataratana, S. and J. B. Thomas (1977b). Quantization for sequential signal detection. *IEEE Transactions on Communications* 25(7), 696–703.
- Tenney, R. R. and N. R. Sandell (1981). Detection with distributed sensors. *IEEE Transactions on Aerospace and Electronic Systems* (4), 501–510.
- Tsitsiklis, J. (1993). Decentralized detection. *Advances in Statistical Signal Processing* 2(2), 297–344.
- Tsitsiklis, J. and M. Athans (1985). On the complexity of decentralized decision making and detection problems. *IEEE Transactions on Automatic Control* 30(5), 440–446.
- Tsitsiklis, J. N. (1988). Decentralized detection by a large number of sensors. *Mathematics of Control, Signals and Systems* 1(2), 167–182.
- Van, T., K. Bell, Z. Tiany, et al. (2013). *Detection Estimation and Modulation Theory*. Wiley & Sons, Inc.
- Varshney, P. K. (2012). *Distributed Detection and Data Fusion*. Springer Science & Business Media.

- Veeravalli, V. V., T. Basar, and H. V. Poor (1993). Decentralized sequential detection with a fusion center performing the sequential test. *IEEE Transactions on Information Theory* 39(2), 433–442.
- Veeravalli, V. V. and P. K. Varshney (2012). Distributed inference in wireless sensor networks. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 370(1958), 100–117.
- Viswanathan, R. and P. K. Varshney (1997). Distributed detection with multiple sensors part i. fundamentals. *Proceedings of the IEEE* 85(1), 54–63.
- Vuran, M. C. and I. F. Akyildiz (2006). Spatial correlation-based collaborative medium access control in wireless sensor networks. *IEEE/ACM Transactions On Networking* 14(2), 316–329.
- Wahdan, M. A., M. F. Al-Mistarihi, and M. Shurman (2015). Static cluster and dynamic cluster head (scdch) adaptive prediction-based algorithm for target tracking in wireless sensor networks. In *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pp. 596–600. IEEE.
- Wald, A. and J. Wolfowitz (1948). Optimum character of the sequential probability ratio test. *The Annals of Mathematical Statistics*, 326–339.
- Warren, D. and P. Willett (1999). Optimum quantization for detector fusion: some proofs, examples, and pathology. *Journal of the Franklin Institute* 336(2), 323–359.
- Yilmaz, Y., G. V. Moustakides, and X. Wang (2012a). Cooperative sequential spectrum sensing based on level-triggered sampling. *IEEE Transactions on Signal Processing* 60(9), 4509–4524.
- Yilmaz, Y., G. V. Moustakides, and X. Wang (2012b). Cooperative sequential spectrum sensing based on level-triggered sampling. *IEEE Transactions on Signal Processing* 60(9), 4509–4524.

- Yu, L. and A. Ephremides (2006). Detection performance and energy efficiency of sequential detection in a sensor network. In *Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06)*, Volume 9, pp. 236a–236a. IEEE.
- Zanella, A., N. Bui, A. Castellani, L. Vangelista, and M. Zorzi (2014). Internet of things for smart cities. *IEEE Internet of Things Journal* 1(1), 22–32.
- Zhang, Q., P. K. Varshney, and R. D. Wesel (2002). Optimal bi-level quantization of iid sensor observations for binary hypothesis testing. *IEEE Transactions on Information Theory* 48(7), 2105–2111.
- Zheng, G. and S. Tang (2011). Spatial correlation-based mac protocol for event-driven wireless sensor networks. *Journal of Networks* 6(1), 121.
- Zytoune, O., Y. Fakhri, and D. Aboutajdine (2010). Lifetime maximisation algorithm in wireless sensor network. *International Journal of Ad Hoc and Ubiquitous Computing* 6(3), 140–149.

## APPENDIX A

### DERIVATION OF THE LOCAL DECISION RULE

By splitting the summation in (2.15) to two summation  $\sum_{\mathbf{u}^m} \sum_{u_m=1,0}$  we get

$$\mathfrak{R} = c + c_{\text{FA}} \sum_{\mathbf{u}^m} \sum_{u_m=1,0} p(u_0 = 1|\mathbf{u})p(\mathbf{u}|H_0) - c_{\text{D}} \sum_{\mathbf{u}^m} \sum_{u_m=1,0} p(u_0 = 1|\mathbf{u})p(\mathbf{u}|H_1) \quad (\text{A.1})$$

where

$\mathbf{u}^m = [u_1, u_2, \dots, u_{m-1}, u_{m+1}, \dots, u_K]^T$  represents the local decision vector excluding  $u_m$  with  $K - 1$  elements and

$$\mathbf{u}^{mj} = [u_1, u_2, \dots, u_{m-1}, u_m = j, u_{m+1}, \dots, u_K]^T, j = 0, 1.$$

Expanding  $\sum_{u_m=1,0}$  for  $u_m = 0, 1$  gives

$$\begin{aligned} \mathfrak{R} = c + c_{\text{FA}} \sum_{\mathbf{u}^m} [p(u_0 = 1|\mathbf{u}^{m0}) p(\mathbf{u}^{m0}|H_0) + p(u_0 = 1|\mathbf{u}^{m1}) p(\mathbf{u}^{m1}|H_0)] - \\ c_{\text{D}} \sum_{\mathbf{u}^m} [p(u_0 = 1|\mathbf{u}^{m0}) p(\mathbf{u}^{m0}|H_1) + p(u_0 = 1|\mathbf{u}^{m1}) p(\mathbf{u}^{m1}|H_1)]. \end{aligned} \quad (\text{A.2})$$

Using the law of total probability we can re-write  $P(\mathbf{u}^{m0}|H_j)$  in terms of  $p(\mathbf{u}^{m1}|H_j)$  as

$$p(\mathbf{u}^{m0}|H_j) = p(\mathbf{u}^m|H_j) - p(\mathbf{u}^{m1}|H_j), \quad (\text{A.3})$$

for  $j = 0, 1$ . Then by substituting (A.3) in (A.2) we have

$$\begin{aligned} \mathfrak{R} = c + c_{\text{FA}} \sum_{\mathbf{u}^m} p(u_0 = 1|\mathbf{u}^{m0}) p(\mathbf{u}^m|H_0) - p(u_0 = 1|\mathbf{u}^{m0}) p(\mathbf{u}^{m1}|H_0) \\ + p(u_0 = 1|\mathbf{u}^{m1}) p(\mathbf{u}^{m1}|H_0) - c_{\text{D}} \sum_{\mathbf{u}^m} p(u_0 = 1|\mathbf{u}^{m0}) p(\mathbf{u}^m|H_1) - p(u_0 = 1|\mathbf{u}^{m0}) \\ + p(u_0 = 1|\mathbf{u}^{m1}) p(\mathbf{u}^{m1}|H_1). \end{aligned} \quad (\text{A.4})$$

After re-arranging the terms in (A.4) we can re-write it as follows:

$$\begin{aligned}
\mathfrak{R} &= c + \underbrace{\sum_{\mathbf{u}^m} p(u_0 = 1 | \mathbf{u}^{m_0}) [c_{\text{FA}} p(\mathbf{u}^m | H_0) - c_{\text{D}} p(\mathbf{u}^m | H_1)]}_{c_m} \\
&+ \sum_{\mathbf{u}^m} \left[ \underbrace{(p(u_0 = 1 | \mathbf{u}^{m_1}) - p(u_0 = 1 | \mathbf{u}^{m_0}))}_{A(\mathbf{u}^m)} p(\mathbf{u}^{m_1} | H_0) c_{\text{FA}} \right. \\
&\quad \left. - \underbrace{(p(u_0 = 1 | \mathbf{u}^{m_1}) - p(u_0 = 1 | \mathbf{u}^{m_0}))}_{A(\mathbf{u}^m)} p(\mathbf{u}^{m_1} | H_1) p_{\text{D}} \right], \tag{A.5}
\end{aligned}$$

which is equivalent to

$$\mathfrak{R} = c_m + \sum_{\mathbf{u}^m} A(\mathbf{u}^m) [c_{\text{FA}} p(\mathbf{u}^{m_1} | H_0) - c_{\text{D}} p(\mathbf{u}^{m_1} | H_1)], \tag{A.6}$$

where  $A(\mathbf{u}^m) = p(u_0 = 1 | \mathbf{u}^{m_1}) - p(u_0 = 1 | \mathbf{u}^{m_0})$ .

Note that the summation  $\sum_{\mathbf{u}^m}$  includes  $2^{K-1}$  terms.

We will perform further manipulations for  $p(\mathbf{u}^{m_1} | H_j)$ ,  $j = 0, 1$  in (A.6) by writing it using a multidimensional integral as

$$p(\mathbf{u} | H_j) = \int_{\mathbf{y}} \underbrace{p(u_m | y_m)}_{\text{decision rule}} \left[ \prod_{k=1, k \neq m}^K p(u_k | y_k) \right] p(y_m | H_j) p(\mathbf{y}^m | y_m, H_j) d\mathbf{y}, \tag{A.7}$$

where  $\mathbf{y}^m = [y_1, \dots, y_{m-1}, y_{m+1}, \dots, y_K]^T$ . Note that the underbraced term in the previous equation represents the decision rule for the local detector which is needed to be designed; i.e.  $p(u_m = 1 | y_m)$  is the probability to decide  $H_1$  at a local detector given the observation  $y_m$ . If  $p(u_m = 1 | y_m) = 0$  or  $1$  it means that we have a deterministic detector (for a certain range of  $y_m$  we will decide 1 and for another range we will decide 0) but also it could be a randomized detector.

Using the previous equations we can write the average risk as follows:

$$\begin{aligned} \mathfrak{R} = c_m + \sum_{\mathbf{u}^m} A(\mathbf{u}^m) \{ c_{\text{FA}} \int_{\mathbf{y}} p(u_m = 1 | y_m) \left[ \prod_{k=1, k \neq m}^K p(u_k | y_k) \right] \\ p(y_m | H_0) p(\mathbf{y}^m | y_m, H_0) d\mathbf{y} - c_{\text{D}} \int_{\mathbf{y}} p(u_m = 1 | y_m) \left[ \prod_{k=1, k \neq m}^K p(u_k | y_k) \right] \\ p(y_m | H_1) p(\mathbf{y}^m | y_m, H_1) \} d\mathbf{y}. \end{aligned} \quad (\text{A.8})$$

Taking the term which represents the local detector;  $p(u_m = 1 | y_m)$  in (A.8), as a common factor produces the following equation

$$\begin{aligned} \mathfrak{R} = \int_{y_m} p(u_m = 1 | y_m) \left[ \sum_{\mathbf{u}^m} \int_{\mathbf{y}^m} A(\mathbf{u}^m) \left\{ c_{\text{FA}} \left[ \prod_{k=1, k \neq m}^K p(u_k | y_k) \right] p(y_m | H_0) \right. \right. \\ \left. \left. p(\mathbf{y}^m | y_m, H_0) - c_{\text{D}} \left[ \prod_{k=1, k \neq m}^K p(u_k | y_k) \right] p(y_m | H_1) p(\mathbf{y}^m | y_m, H_1) \right\} \right] d\mathbf{y}. \end{aligned} \quad (\text{A.9})$$

Utilizing basic concepts of Bayesian detector design (minimum risk Bayesian detector); if the value between outer brackets in (A.9) is negative, we will put  $p(u_m = 1 | y_m) = 1$  which means decide  $H_1$  otherwise we will put it as zero which means decide  $H_0$ .

$$\begin{aligned} c_{\text{FA}} p(y_m | H_0) \sum_{\mathbf{u}^m} \int_{\mathbf{y}^m} A(\mathbf{u}^m) \left[ \prod_{k=1, k \neq m}^K p(u_k | y_k) \right] p(\mathbf{y}^m | y_m, H_0) d\mathbf{y} \stackrel{H_1}{\gtrsim} \stackrel{H_0}{\lesssim} \\ c_{\text{D}} p(y_m | H_1) \sum_{\mathbf{u}^m} \int_{\mathbf{y}^m} A(\mathbf{u}^m) \left[ \prod_{k=1, k \neq m}^K p(u_k | y_k) \right] p(\mathbf{y}^m | y_m, H_1) d\mathbf{y}. \end{aligned} \quad (\text{A.10})$$

We can re-write the local decision rule in (A.10) similar to (Viswanathan and Varshney, 1997) as

$$\frac{p(y_m | H_1)}{p(y_m | H_0)} \stackrel{H_1}{\gtrsim} \frac{c_{\text{FA}} \sum_{\mathbf{u}^m} \int_{\mathbf{y}^m} A(\mathbf{u}^m) \left[ \prod_{k=1, k \neq m}^K p(u_k | y_k) \right] p(\mathbf{y}^m | y_m, H_0) d\mathbf{y}}{c_{\text{D}} \sum_{\mathbf{u}^m} \int_{\mathbf{y}^m} A(\mathbf{u}^m) \left[ \prod_{k=1, k \neq m}^K p(u_k | y_k) \right] p(\mathbf{y}^m | y_m, H_1) d\mathbf{y}}. \quad (\text{A.11})$$

From the previous discussion we can see that optimizing any decision rule can be performed keeping the remaining decision rules fixed which is called as person-by-person optimization (PBPO) and it necessitates solving  $(K + 2^K)$  coupled nonlinear equations.

## APPENDIX B

### DESIGN EXAMPLE OF A BAYESIAN PARALLEL FUSION NETWORK WITH TWO SENSORS

Consider a parallel topology with FC consisting of two local sensors with conditionally independent and identically distributed observations, (Varshney, 2012). From (2.17) we can write the LRT at sensor  $k = 1$  as follow:

$$\frac{c_D p(y_1|H_1)}{c_{FA} p(y_1|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\sum_{u_2} A(u_2) \int_{y_2} p(u_2|y_2) p(y_2|, H_0)}{\sum_{u_2} A(u_2) \int_{y_2} p(u_2|y_2) p(y_2|H_1)} \quad (\text{B.1})$$

where

$$A(u_2) = p(u_0 = 1|u_1 = 1, u_2) - p(u_0 = 1|u_1 = 0, u_2).$$

Assuming  $\eta_{ijk} = p(u_0 = i|u_1 = j, u_2 = k)$ ,

$$A(u_2) = \eta_{11u_2} - \eta_{10u_2}, \quad \text{for } i, j, k = 0, 1.$$

For a binary local detector  $u_1$  we can write (B.1) as follows:

$$\frac{c_D p(y_1|H_1)}{c_{FA} p(y_1|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\eta_{110} - \eta_{100} + (\eta_{111} - \eta_{101} - \eta_{110} + \eta_{100}) \int_{y_2} p(u_2 = 1|y_2)p(y_2|H_0)}{\eta_{111} - \eta_{101} + (\eta_{101} - \eta_{111} + \eta_{110} - \eta_{100}) \int_{y_2} p(u_2 = 1|y_2)p(y_2|H_1)}, \quad (\text{B.2})$$

then

$$\frac{p(y_1|H_1)}{p(y_1|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} t_1 \quad (\text{B.3})$$

where

$$t_1 = \frac{c_{FA} \left[ \eta_{110} - \eta_{100} + (\eta_{111} - \eta_{101} - \eta_{110} + \eta_{100}) \int_{y_2} p(u_2 = 1|y_2)p(y_2|H_0) \right]}{c_D \left[ \eta_{111} - \eta_{101} + (\eta_{101} - \eta_{111} + \eta_{110} - \eta_{100}) \int_{y_2} p(u_2 = 1|y_2)p(y_2|H_1) \right]}.$$

Following the equations from (B.1) to (B.3) we can write the decision rule for sensor  $k = 2$  as:

$$\frac{p(y_2|H_1)}{p(y_2|H_0)} \underset{H_0}{\overset{H_1}{\geq}} t_2 \quad (\text{B.4})$$

where

$$t_2 = \frac{c_{\text{FA}} \left[ \eta_{101} - \eta_{100} + (\eta_{111} - \eta_{101} - \eta_{110} + \eta_{100}) \int_{y_1} p(u_1 = 1|y_1)p(y_1|H_0) \right]}{c_{\text{D}} \left[ \eta_{111} - \eta_{110} + (\eta_{101} - \eta_{111} + \eta_{110} - \eta_{100}) \int_{y_1} p(u_1 = 1|y_1)p(y_1|H_1) \right]}.$$

We notice from (B.3) that the threshold  $t_1$  at the first sensor is a function of the decision at the second sensor, which means the threshold on the first sensor is coupled with the threshold on the second sensor.  $t_1$  is a function of  $t_2$  because  $t_1$  involves  $p(u_2 = 1|y_2)$  which is a function of  $t_2$  also.

$$\text{LR}(y_1) \underset{H_0}{\overset{H_1}{\geq}} t_1 = \text{function of } t_2,$$

$$\text{LR}(y_2) \underset{H_0}{\overset{H_1}{\geq}} t_2 = \text{function of } t_1.$$

To be more specific, we assume that we have Gaussian likelihoods,  $y_i \sim \mathcal{N}(0, 1)$ , under  $H_0$ , and  $y_i \sim \mathcal{N}(\mu_i, 1)$ ,  $i = 1, 2$  under  $H_1$  where,  $\mu_1, \mu_2 > 0$  as given in (Varshney, 2012). The LR for the first and second detector are given respectively as:

$$\frac{p(y_1|H_1)}{p(y_1|H_0)} = \exp\left(\frac{y_1^2 - (y_1 - \mu_1)^2}{2}\right) = \exp\left(\mu_1 y_1 - \frac{1}{2}\mu_1^2\right) \quad (\text{B.5})$$

and

$$\frac{p(y_2|H_1)}{p(y_2|H_0)} = \exp\left(\mu_2 y_2 - \frac{1}{2}\mu_2^2\right). \quad (\text{B.6})$$

Simplifying (B.5) and (B.6) will give

$$y_1 \underset{p(u_0|y_0)=0}{\overset{p(u_1|y_1)=1}{\geq}} \frac{1}{\mu_1} \log(t_1) + \frac{\mu_1}{2} \quad (\text{B.7})$$

and

$$y_2 \underset{P(u_2=0|y_2)=0}{\overset{P(u_2=1|y_0)=1}{\geq}} \frac{1}{\mu_2} \log(t_2) + \frac{\mu_2}{2}, \quad (\text{B.8})$$



respectively. Now, we can find the values of the integrals inside the expression of thresholds  $t_1$  and  $t_2$ . For threshold  $t_1$  the probability of  $u_2 = 1$  given a local observation  $y_2$  in the numerator is given as:

$$\begin{aligned} \int_{y_2} p(u_2 = 1|y_2) p(y_2|H_0) &= \int_{\frac{1}{\mu_2} \log(t_1) + \frac{\mu_2}{2}}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-y_2^2}{2}\right) dy_2 \\ &= Q_{\text{func}}\left(\frac{1}{\mu_2} \log(t_2) + \frac{\mu_2}{2}\right), \end{aligned} \quad (\text{B.9})$$

and in the denominator as:

$$\begin{aligned} \int_{y_2} p(u_2 = 1|y_2) p(y_2|H_1) &= Q_{\text{func}}\left(\frac{\mu_2}{2} + \frac{1}{\mu_2} \log(t_2) - \mu_2\right) \\ &= Q_{\text{func}}\left(\frac{1}{\mu_2} \log(t_2) - \frac{\mu_2}{2}\right). \end{aligned} \quad (\text{B.10})$$

In the same way, we can evaluate the integrals inside  $t_2$  in the numerator and denominator respectively, as follows:

$$\int_{y_1} p(u_1 = 1|y_1) p(y_1|H_0) = Q_{\text{func}}\left(\frac{1}{\mu_1} \log(t_1) + \frac{\mu_1}{2}\right) \quad (\text{B.11})$$

and

$$\int_{y_1} p(u_1 = 1|y_1) p(y_1|H_1) = Q_{\text{func}}\left(\frac{1}{\mu_1} \log(t_1) - \frac{\mu_1}{2}\right). \quad (\text{B.12})$$

We can re-write the thresholds  $t_1$  and  $t_2$  using the equations from (B.9) to (B.12) as

$$\frac{c_D}{c_{FA}} t_1 = \frac{\eta_{110} - \eta_{100} + (\eta_{111} - \eta_{101} - \eta_{110} + \eta_{100}) Q_{\text{func}}\left(\frac{1}{m_2} \log(t_2) + \frac{m_2}{2}\right)}{\eta_{110} - \eta_{100} + (\eta_{111} - \eta_{101} - \eta_{110} + \eta_{100}) Q_{\text{func}}\left(\frac{1}{m_2} \log(t_2) - \frac{m_2}{2}\right)} \quad (\text{B.13})$$

$$\frac{c_D}{c_{FA}} t_2 = \frac{\eta_{101} - \eta_{100} + (\eta_{111} - \eta_{101} - \eta_{110} + \eta_{100}) Q_{\text{func}}\left(\frac{1}{m_1} \log(t_1) + \frac{m_1}{2}\right)}{\eta_{101} - \eta_{100} + (\eta_{111} - \eta_{101} + \eta_{110} + \eta_{100}) Q_{\text{func}}\left(\frac{1}{m_1} \log(t_1) - \frac{m_1}{2}\right)}. \quad (\text{B.14})$$

Re-arranging the previous equations gives:

$$t_1 = \frac{c_{FA} \left[ \eta_{110} - \eta_{100} + (\eta_{111} - \eta_{101} - \eta_{110} + \eta_{100}) Q_{\text{func}}\left(\frac{1}{\mu_2} \log(t_2) + \frac{\mu_2}{2}\right) \right]}{c_D \left[ \eta_{111} - \eta_{101} + (\eta_{101} - \eta_{111} + \eta_{110} - \eta_{100}) Q_{\text{func}}\left(\frac{1}{\mu_2} \log(t_2) - \frac{\mu_2}{2}\right) \right]} \quad (\text{B.15})$$

and

$$t_2 = \frac{c_{FA} \left[ \eta_{101} - \eta_{100} + (\eta_{111} - \eta_{101} - \eta_{110} + \eta_{100}) Q_{\text{func}}\left(\frac{1}{\mu_1} \log(t_1) + \frac{\mu_1}{2}\right) \right]}{c_D \left[ \eta_{111} - \eta_{110} + (\eta_{101} - \eta_{111} + \eta_{110} - \eta_{100}) Q_{\text{func}}\left(\frac{1}{\mu_1} \log(t_1) - \frac{\mu_1}{2}\right) \right]}. \quad (\text{B.16})$$

The coupled equations, (B.15) and (B.16) can be solved to obtain the values of  $t_1$  and  $t_2$ . Solving the coupled equations, (B.15) and (B.16) requires to assign values for  $\eta_{ijk}$  which represents the probability of global decision,  $u_0 = i$  for given local decisions,  $u_1 = j$  and  $u_2 = k$ , where  $i, j, k = 0, 1$ . In this case, we have to try all the possibilities of  $u_1$  and  $u_2$  for a certain fusion rule and then assign the value for  $\eta_{ijk}$ . For each possible fusion rule and its corresponding  $\eta_{ijk}$ s we will check if the coupled equations give a solution for  $t_1$  and  $t_2$  which is consistent with the  $2^K$  in (2.20). In case  $t_1$  and  $t_2$  have more than one solution, we will choose the one which minimizes the average risk, e.g. using OR rule as global decision at FC gives  $\eta_{100} = 0$  and  $\eta_{101}, \eta_{110}, \eta_{111} = 1$ . The possible number of fusion rules for  $K$  sensors equals to  $2^{2^K}$ ; for our case  $K = 2$  which gives 16 logical function rule. From these 16 possibility we will chose AND, OR, decision in favor of sensor 1, and decision in favor of sensor 2. To minimize average probability of error, let  $c_{10} = c_{01} = 1$  and  $c_{00} = c_{11} = 0$  which gives  $c_{FA} = p(H_0)$  and  $c_D = p(H_1)$ . We will consider the thresholds and the fusion rule for three different sets of values  $\mu_1 = 1, \mu_2 = 1$ ;  $\mu_1 = 1, \mu_2 = 1.5$  and  $\mu_1 = 1, \mu_2 = 2$ . The thresholds at 2 sensors and its corresponding average probability of error  $p_e$  for the three cases are shown in the Figures from B.1 to B.6. Figure B.1 shows the average probability of error,  $p_e$ , as a function of the prior probability,  $p(H_0)$ , for the case  $\mu_1 = \mu_2 = 1$ . The worst  $p_e$  is achieved when the global decision is taken according to  $u_1$  and disregards  $u_2$ , the red curve, which is identical to take the global decision according  $u_2$  and disregards  $u_1$ , blue squared curve. Using OR rule gives the minimum  $p_e$  for  $p(H_0) < 0.5$ , black curve whereas AND rule gives the minimum  $p_e$  for  $p(H_0) > 0.5$ , green curve. The minimum  $p_e$  obtained from the lower envelope of the  $p_e$  curves corresponding to different fusion rules, black crossed curve; which obtained from both OR and AND rules; represents the minimum  $p_e$  in this case. For the same case the identical thresholds for the two sensors are shown in Figure B.2.

Figure B.3 shows the average probability of error,  $p_e$ , when  $\mu_1 = \mu_2 = 1.5$ . In this figure, we can see the maximum  $p_e$  is obtained when the global decision is taken according to the first sensor only, the red curve. The  $p_e$  decreases and we can achieve better performance in terms of  $p_e$  when we take the global decision according to the second sensor, the blue curve. This is logical because both sensors have zero mean observations

under hypothesis  $H_0$  but the second sensor has a mean,  $\mu = 1.5$  under  $H_1$  which is greater than the mean of the first sensor,  $\mu = 1$ . We can achieve a better performance in term of  $p_e$  if we increase the mean of the second sensor to be  $\mu = 2$  as shown in Figure B.5. (Increasing the distance between the two hypotheses decreases the  $p_e$ ). The corresponding thresholds for the cases ,  $\mu_1 = \mu_2 = 1.5$  and  $\mu_1 = \mu_2 = 2$  are shown respectively in Figures B.4 and B.6.

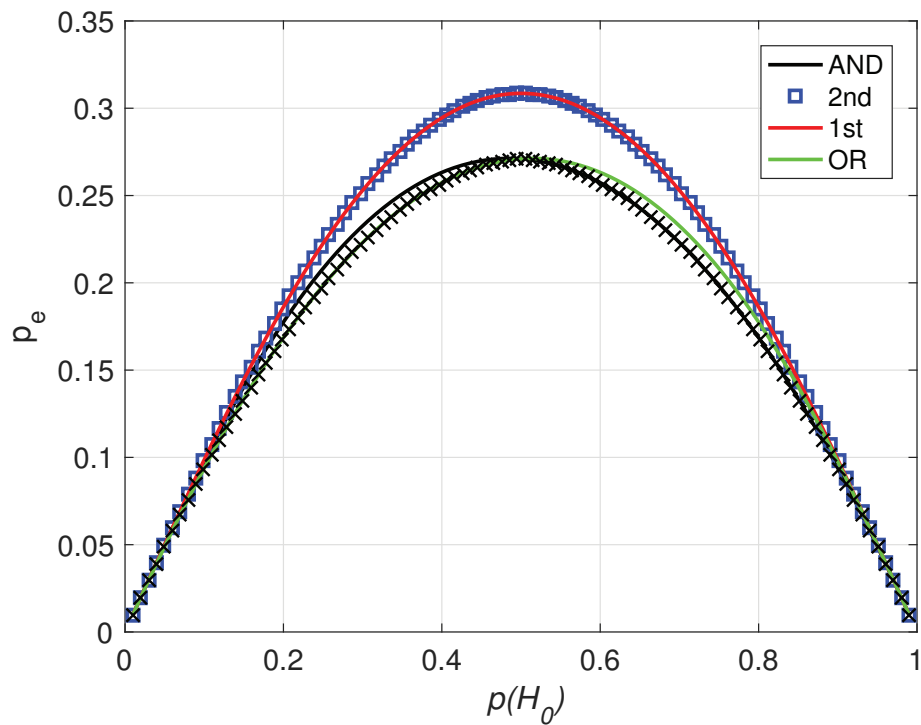


Figure B.1. Average probability of error, case 1:  $\mu_1 = 1, \mu_2 = 1$ .

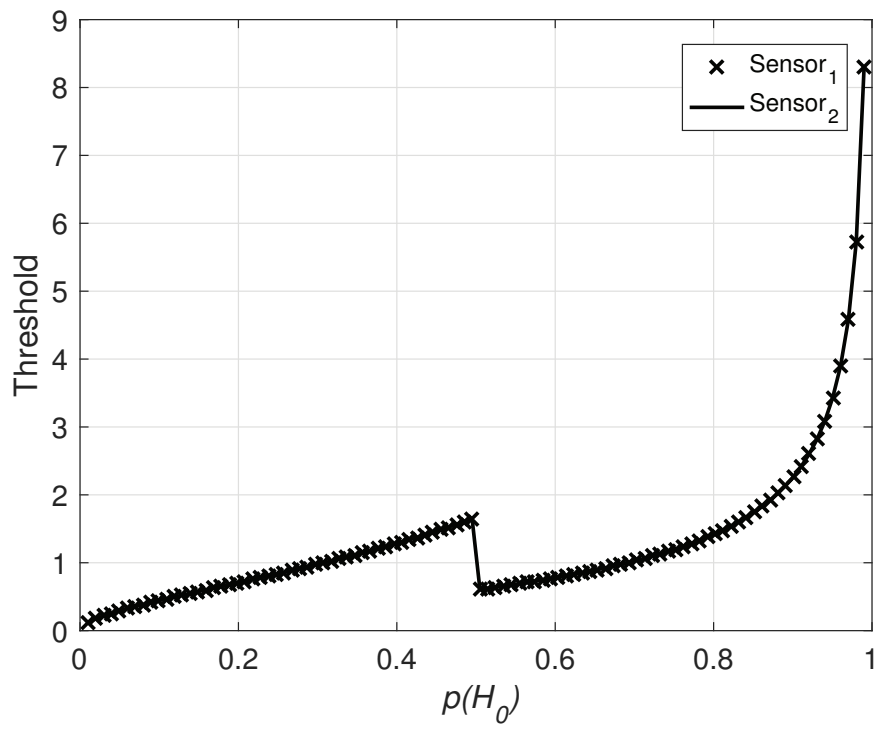


Figure B.2. The threshold values, case 1:  $\mu_1 = 1, \mu_2 = 1$ .

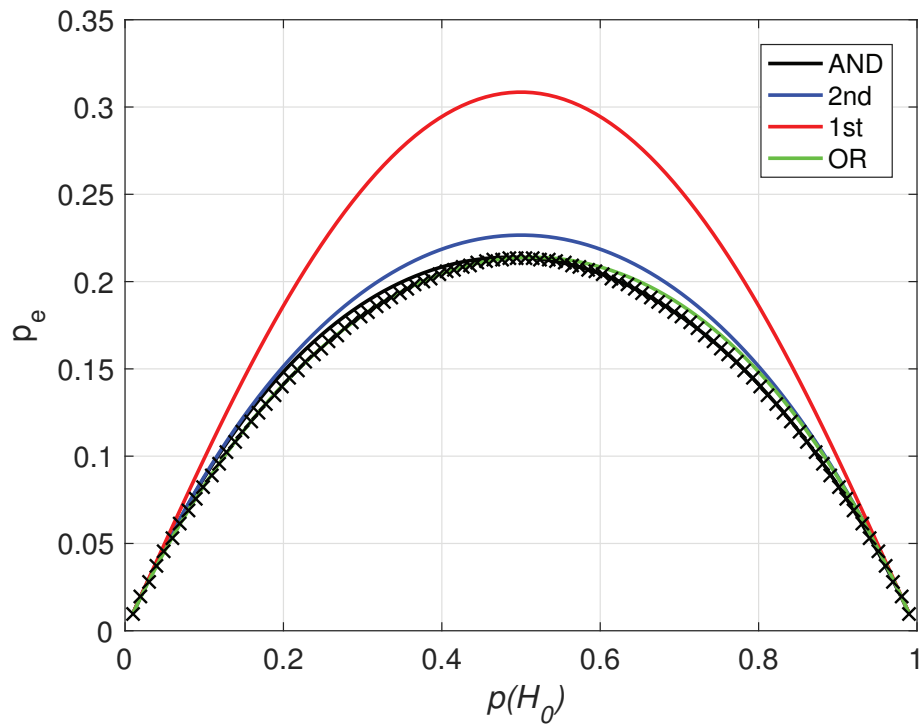


Figure B.3. Average probability of error, case 2:  $\mu_1 = 1, \mu_2 = 1.5$ .

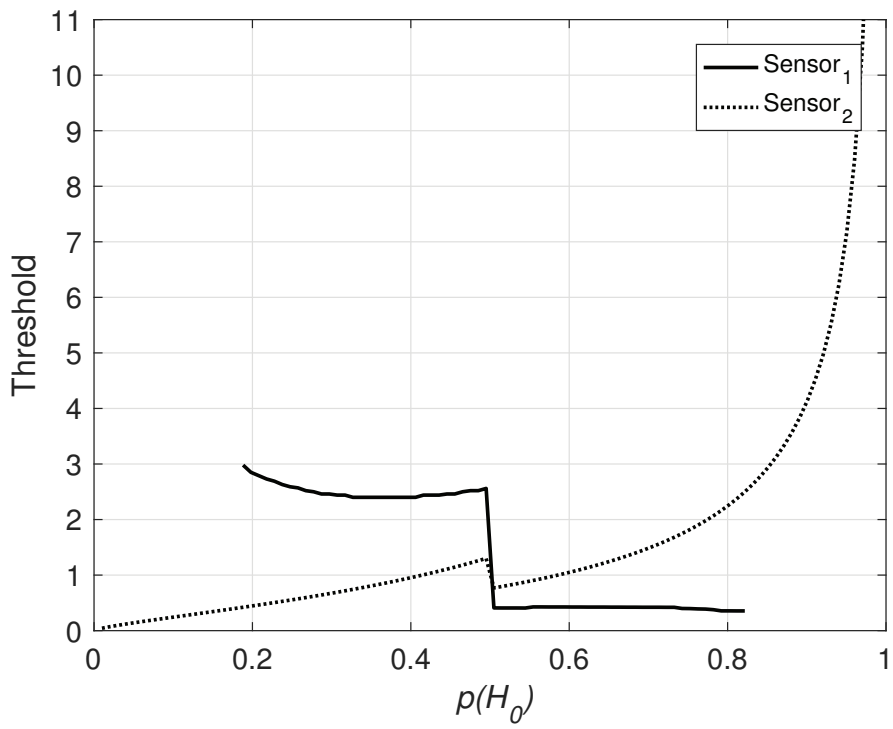


Figure B.4. The threshold values, case 2:  $\mu_1 = 1, \mu_2 = 1.5$ .

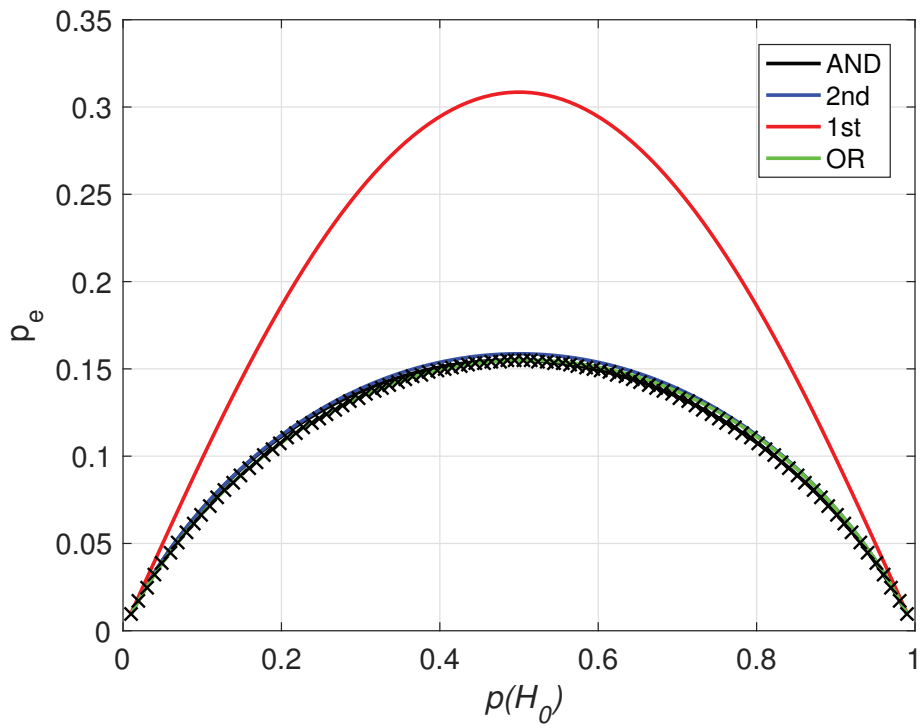


Figure B.5. Average probability of error, case 3:  $\mu_1 = 1, \mu_2 = 2$ .

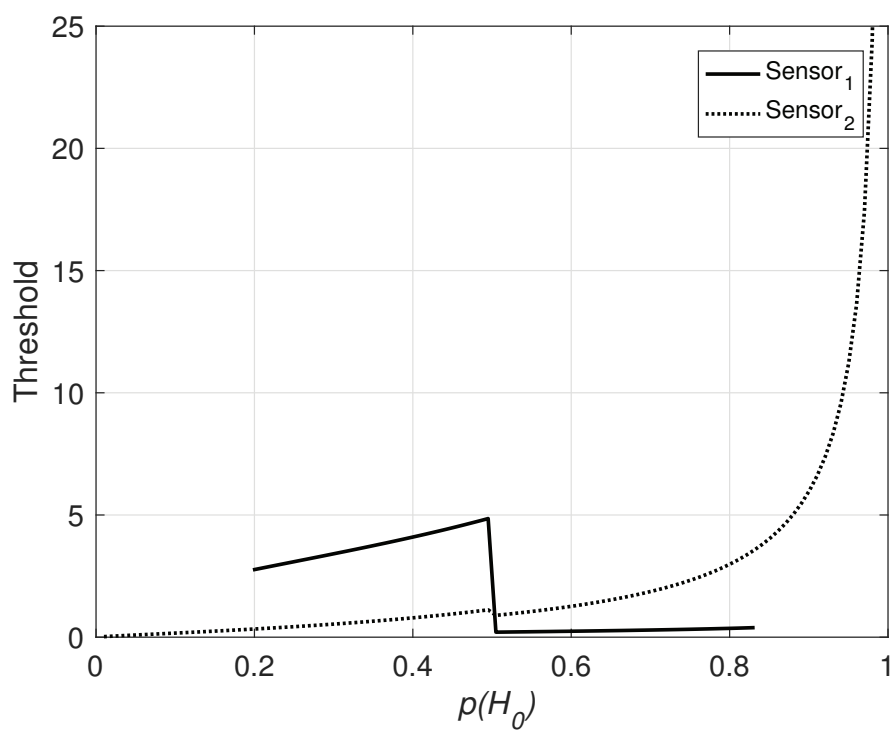


Figure B.6. The threshold values, case 3:  $\mu_1 = 1, \mu_2 = 2$ .

## APPENDIX C

### TOY EXAMPLES FOR APPLYING MAE AND MJD

It is expectable that for binary quantization with symmetrical pdfs under both hypotheses as in Case 1, one will obtain the same threshold and consequently the same ROC performance using either of the methods. However with asymmetrical pdfs as in Case 2, the two methods result in different thresholds and consequently different ROC performance.

- Case 1: In this case, we consider two pdfs symmetric around the point  $\lambda=0.5$ ,  $q(\lambda) = -2(\lambda - 1)$  and  $p(\lambda) = 2\lambda$  as shown in Figure C.1, where the areas  $A_1$ ,  $A_3$  represent the probabilities of the partitions under  $p(\lambda)$  and the areas  $A_2$ ,  $A_4$  represent the probabilities of the partitions under  $q(\lambda)$  (note:  $p(\lambda)$  and  $q(\lambda)$  are nonzero only in  $\lambda$  in  $(0,1)$  (these should be defined)).

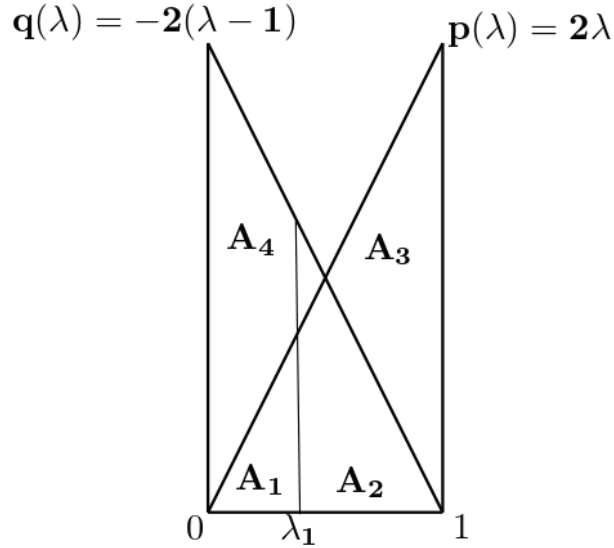


Figure C.1. Symmetric pdf

The average entropy and the J-divergence equations for this example are given respectively as follows:

$$F_{av} = 0.5(-A_1 \log_2(A_1) - A_2 \log_2(A_2) - A_3 \log_2(A_3) - A_4 \log_2(A_4)), \quad (C.1)$$

$$J = A_2 \log_2 \frac{A_2}{A_3} + A_1 \log_2 \frac{A_1}{A_4} - A_3 \log_2 \frac{A_2}{A_3} - A_4 \log_2 \frac{A_1}{A_4}. \quad (\text{C.2})$$

The threshold value  $\tilde{\lambda}$  that maximizes both the average entropy and J-divergence can be found by solving (C.3) and (C.4) which give  $\tilde{\lambda}_{MAE} = \tilde{\lambda}_{MJD} = 0.5$ , where

$$\tilde{\lambda}_{MAE} = \arg \max_{\lambda} \hat{F}_{av}, \quad (\text{C.3})$$

$$\tilde{\lambda}_{MJD} = \arg \max_{\lambda} J. \quad (\text{C.4})$$

- Case 2: An asymmetrical pdf set where  $q(\lambda) = 1$  and  $p(\lambda) = 2\lambda$  as shown in Figure C.2. The average entropy and the J-divergence equations are given by (C.1) and (C.2), respectively.

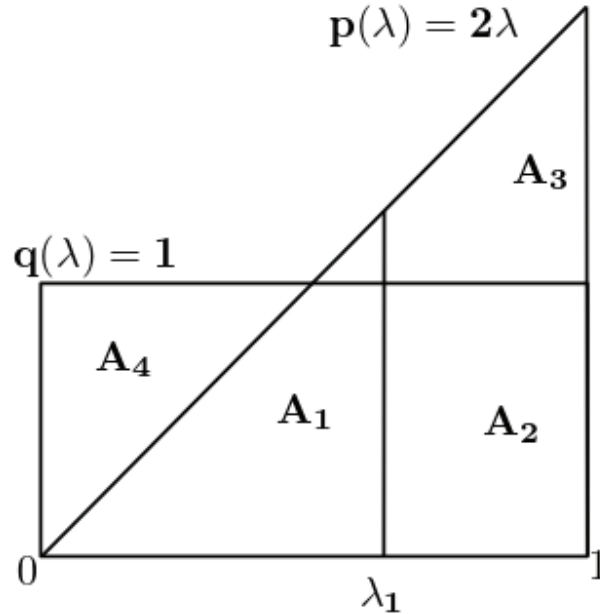


Figure C.2. Asymmetric pdf

The threshold values  $\tilde{\lambda}$  that maximize both the average entropy and J-divergence can be found by solving (C.3) and (C.4) which give  $\tilde{\lambda}_{MAE} = 0.6295$  and  $\tilde{\lambda}_J = 0.318$ .

The ROC curves have been considered for both symmetrical and asymmetrical pdfs as shown in Figure C.3 and Figure C.4, respectively. In the case of symmetric pdfs, MAE



outperforms the MJD for the probability of false alarm between 0 and 0.5 and is inferior to it for the probability of false alarm between 0.5 and 1.

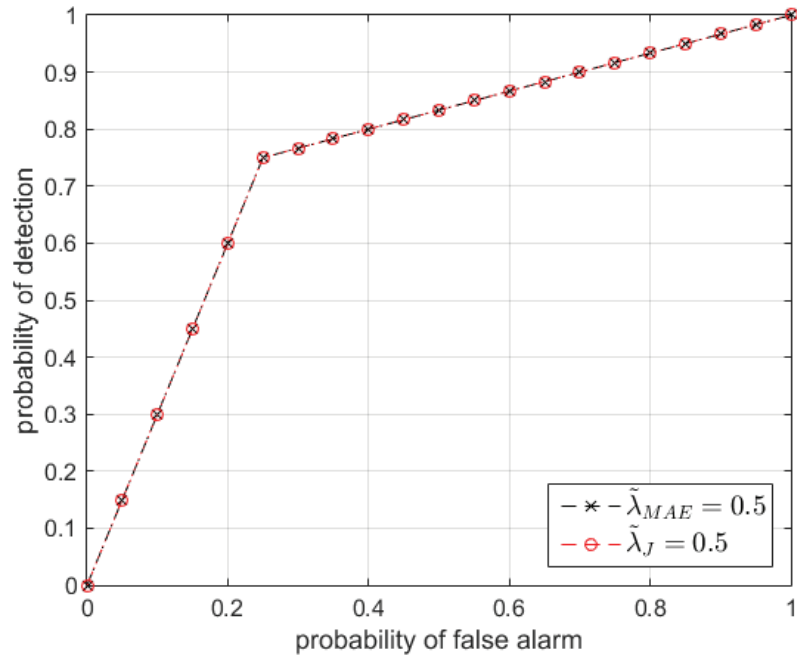


Figure C.3. ROC symmetrical pdfs

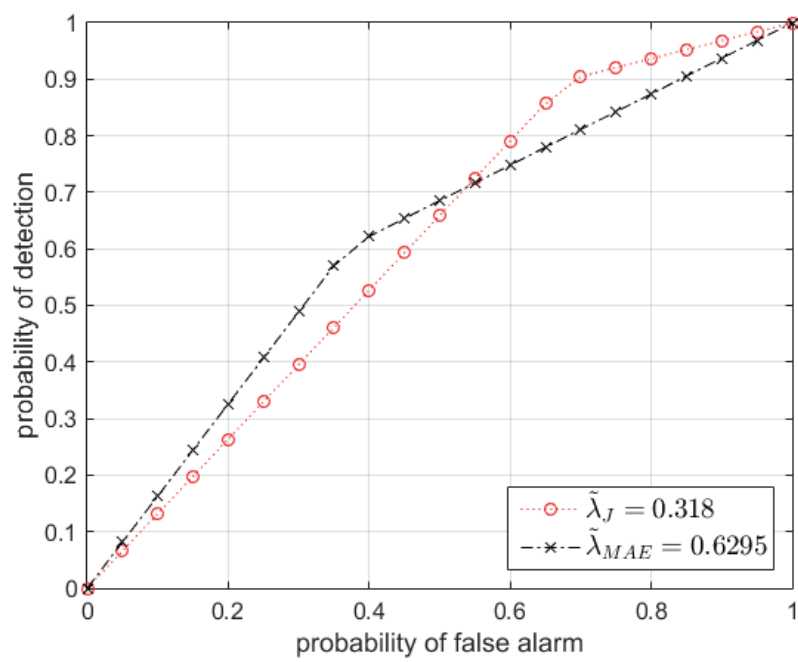


Figure C.4. ROC unsymmetrical pdfs

# VITA

**Date and Place of Birth:** 15.01.1987, Palestine

## EDUCATION

### **2014 - 2020 Doctor of Philosophy in Electrical and Electronics Engineering**

Graduate School of İzmir Institute of Technology, İzmir - Turkey

Thesis Title: MEDIUM-AWARE INFERENCE FOR WIRELESS SENSOR NETWORKS

Supervisor: Assoc. Prof. Dr. Mustafa A. Altinkaya

### **2009 - 2012 Master of Science in Electrical Engineering**

Graduate School of Engineering and Sciences, Jordan University of Science and Technology, Irbid-Jordan

Thesis Title: ENERGY EFFICIENT ADAPTIVE PREDICTION-BASED ALGORITHM for TARGET TRACKING in WIRELESS SENSOR NETWORK

Supervisor: Assoc. Prof. Dr. Mamoun F. Al-Mistarihi

### **2005 - 2009 Bachelor of Electrical Engineering**

Department of Electrical Engineering, Palestine Technical University, Palestine

## PROFESSIONAL EXPERIENCE

### **Jan 2013- September 2013 Lecturer**

Instructor for many communication engineering courses, Tulkarm- Palestine

### **September 2009- June 2012 Teaching and Research Assistant**

Teaching and Research Assistant, Irbid- Jordan

## PUBLICATIONS

WAHDAN, M., M. A. Altinkaya (in Prep.). Maximum Average Entropy-Based Quantization of Local Observations for Distributed Detection.

WAHDAN, M., M. A. Altinkaya (2019). Optimal Quantization in Decentralized Detection by Maximizing the Average Entropy of the Sensors. In 27th Signal Processing and Communications Applications Conference (SIU), pp.1-4. IEEE.