摘要

在無線感測器網路系統下，現存有關於整合通道資訊之偵測法設計皆假設融合中心(Fusion Center)具有本地感測器偵測機率的資訊，但在實際狀況下，本地感測器對事件發生的偵測機率多隨時間和環境不同而變化。本計劃中，假設本地感測器偵測率未知的情況下，吾人透過二元對稱通道(BSC)傳送本地之一位元判斷報告給融合中心。為解決此類存有未知的問題，傳統上多利用廣義似然比檢驗法(GLRT)，然而此方法並無法達到最佳效能，且難以分析。本計劃乃針對此兩缺點，先是提出了最大似然估計法(ML estimate)的化簡，接著根據此化簡，為融合中心設計出較 GLRT 簡單之判斷法，並加以分析其效能。吾人由效能公式中觀察出通道效應對整體效能的影響，更進而設計出分配本地感測器傳送功率之方法。吾人所提出的判斷法輔以傳送功率之分配，和 GLRT 相比，不但大幅降低複雜度，效能更有顯著的改善，甚至接近偵測系統所能達到的理論最佳值。

Abstract

In the field of wireless sensor networks, existing works of channel-aware fusion rule design assume that the fusion center (FC) knows the local sensor detection probabilities. However, this paradigm ignores the possibility of unknown sensor alarm responses to the event occurrences. This work focuses on the case where the local detection probability is unknown and assumes sensors transmit their one-bit reports through binary symmetric channels to FC. Traditionally, Generalized Likelihood Ratio Test (GLRT) can tackle this scenario, but it does not guarantee optimal performance and is too complicated to analyze. To solve these problems, a simpler fusion rule is proposed based on the simplified ML estimate, and its performance is analyzed. By investigating the channel effects, a power allocation scheme is then proposed to further improve the performance. Being far less complicated than GLRT, the proposed fusion rule with power allocation outperforms GLRT significantly and can even achieve the performance of LRT, which is the optimal rule for any possible detectors.

Table of Contents

List of Figures

Acronym Glossary

AWGN	additive white Gaussian noise
BER	bit error rate
BSC	binary symmetric channel
CDMA	code division multiple access
CSI	channel state information
EGC	equal gain combining
FDMA	frequency division multiple access
FC	fusion center
GLRT	generalized likelihood ratio test
LRT	likelihood ratio test
MIMO	multiple-input multiple-output
ML	Maximum likelihood
MRG	Maximum ratio combining
MSE	mean square error
PDF	probability density function
PBPO	person-by-person optimization
ROC	receiver operating curve
SNR	signal-to-noise ratio
TDMA	time division multiple access
WSN	wireless sensor network

Notations

$\mathcal{H}_0 \,/\, \mathcal{H}_1$	the state of absent/present event
N	total number of sensors
s_i	1-bit report from sensor i
ε_i	cross-over probability of BSC between sensor i and FC
r_i	the hard decision of s_i at FC
s_0	global decision made by FC
$p_d \,/\, p_f$	local detection/false-alarm probability
$P_d \,/\, P_f$	global detection/false-alarm probability
γ	threshold used in fusion rule
$\Pr\{\bullet\}$	probability of an event
$Q(\bullet)$	Q function
P	total transmit power
p_i	Transmit power for sensor i
N_0	two times the power of noise

Chapter 1

Introduction

Distributed detection systems are typically composed of many sensors and an FC, working collaboratively to detect an event of interest. With significant progresses in wireless communications, networking and microprocessors, distributed detection using wireless sensor networks (WSN) has become an active research area, [1-3]. In WSN, sensors are connected by wireless channels to each other and FC so that they can be flexibly deployed, enhancing the surveillance coverage and the sensing potential, especially in applications like battlefield and monitoring for security or environment. In contrast to traditional communication systems, sensors are usually cheap device with low power usage; together with limited channel capacity, the stringent communication resources make the system design quite challenging. Conventionally, signal processing algorithms are treated as independent part of the communication block, and thus most of the earlier studies are based on the idealized assumption that the sensor reports can be received at FC without errors [3, 4], which are so-called classical distributed detection problems. As is proved that designers should integrate the signal processing algorithms design and the communication aspect to reach the optimal performance, it is especially true in WSN where channels cannot be assumed reliable anymore compared to those in classical distributed detection. Recently there have been several proposals further taking into account the communication channel impairments [5-9]; see [10] for a tutorial introduction to distributed detection in the presence of non-ideal channel links. A common assumption made in these channel-aware schemes is that the local sensor detection performance, characterized by the detection and false-alarm probabilities, is known at the FC. This, however, ignores the possibly unknown sensor alarming responses to the occurrence of events. Consider, for example, that a sensor network is deployed for monitoring the increase in the room temperature, as in the scenario of home security against fire. The local detection probability (under a fixed threshold) could be unknown due to the response to the uncertain temperature of fire events. As the local detection probability being indispensable in most fusion rule design, to reflect the variation of the sensing field conditions, a conceivable approach is thus to model the local detection probability as an unknown parameter, and to accordingly design the global decision rule for tackling such uncertainty.

In this report, we propose a channel-aware distributed detection scheme for the above-mentioned scenario. The communication links between the sensor nodes and the FC are modeled as binary symmetric channels. In the proposed approach each sensor, when triggered, just sends a single bit to inform the FC of its local decision; no further communication overhead is needed for conveying the message about the current local detection performance. The FC treats the local detection probability as an unknown parameter. Based solely on the received sensor reports, the global decision rule is naturally formulated as GLRT [11]. The implementation of GLRT calls for the maximum likelihood (ML) estimate of the unknown parameter which, in our case, does not lead to a closed-from solution. Under the high signal-to-ratio (SNR) assumption this work derives an approximate ML estimate that is *affine* in the received data. It is seen that, even with the approximation of the ML solution, the performance of GLRT remains quite difficult to characterize. Based on the approximated ML scheme, we then propose a simple alternative fusion rule in which the test statistic is affine in the received data. The main advantages of this alternative are threefold. Firstly, it allows deriving closed-form performance results for facilitating analytic characterization of the channel effect. Secondly, it is shown that, under certain conditions, the global detection performance can be improved by enhancing the communication-link quality, e.g., reducing the average link bit-error rate (BER). Hence, this work then proposes an optimal power allocation scheme to minimize the mean BER subject to a total power budget. Thirdly, simulations show that the proposed alternative scheme outperforms GLRT. The rest of this report is organized as follows. Chapter 2 traces the main developments in the field of distributed detection and gives reviews of related works in WSN. Chapter 3 starts with the problem formulation, presents the GLRT based detection scheme and then derives the approximate ML solution. Chapter 4 introduces the alternative approach and derives the associated analytic performance results. The issue of channel impairment mitigation for improving the detection performance is then addressed. In Chapter 5 the simulation results are shown and interpreted. Finally, Chapter 6 concludes this report and suggests some future works.

Chapter 2

Wireless Sensor Network for Detection

WSN have received greater research interests in recent years. Depending on various applications and environment settings, many different system models have been proposed, dedicating to solve different problems inherent in their environmental assumptions. In this chapter, the concepts and models of the development of distributed detection systems are reviewed. In the beginning, this report introduces the classical detection problem, which assumes reliable transmission from sensors to the center controller. A popular model called canonical distributed detection systems is illustrated, which is similar to the model settings in this work.

WSN are distributed detection systems built on the wireless infrastructure, and many new design challenges emerge. We investigate some fundamental works in areas of WSN and point out the possible room for improvements which motivates our work.

2.1 Review of Distributed Detection Systems

Distributed detection systems refer to the systems where multiple sensors work in some way to distinguish between two or more hypothesis, which is also often denoted as the states of the environments. As can be traced back to the precedent human-like activities such as voting when people try to make decisions, the first formal treatments can be found in the work of Radner [12], where the problem of decision making from multiple persons is addressed. Afterwards, many applications and the corresponding problem formulations are proposed, and one kind of the prevailing applications is the system involving both distributed sensors and an FC. Specifically, the fusion center makes the final decision based on the information gathered from local sensors. In the following context, let's confine the notation "distributed detection systems" to such systems without otherwise noted. If the raw observations of the local sensors are accessible at FC, this scenario is just the classical hypothesis testing problem [13]. Unfortunately, due to the limited communication resources such as channel capacity between sensors and FC, the observations are often compressed at local sensors and then transmitted to FC.

In distributed detection systems, the classical distributed detection had been an active research field following the seminal work of Tenney and Sandell in 1981 [14], where the so-called canonical distributed detection system is established. This system assumes that the distributed sensors communicate directly through parallel channels, as illustrated below,

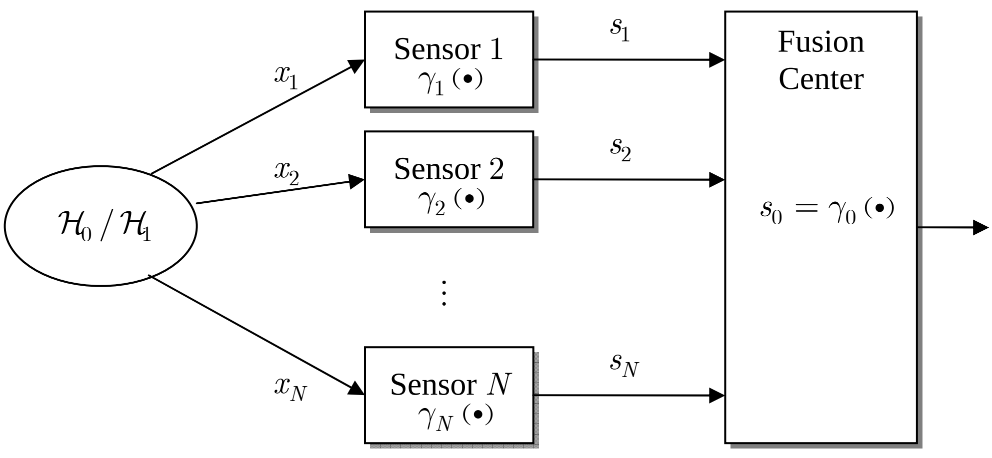

Figure 2.1: Canonical distributed detection systems

Note that canonical distributed detection systems assume local sensor outputs can be received reliably at FC, and the only uncertainty comes from the observation noise. In such systems, two problems are to be solved: one is the decision rule (or fusion rule) at FC and the other is the signal processing schemes at the local sensors. These two problems are wined with each other and have to be jointly designed.

For the first problem, if FC knows perfectly the PDF of the gathered information under every hypothesis, then the optimal fusion rule is the Likelihood Ratio Test (LRT) [11]. This rule holds no matter the received signals at FC are soft or hard decided, as long as FC knows the PDF of the soft/hard decided signals. However, when there are unknown parameters in these PDF's under different hypothesis, a commonly used heuristic approach is GLRT [11], although it is not really the optimal fusion rule in such cases. LRT and GLRT are relevant to our works and they are introduced in more detail in Chapter 2.2. The second problem is a more complicated one. Under the conditional independence assumption, the optimality of LRT at local sensors is established. However, the LRT thresholds for the sensors are connected with each other. The dominating approach finding the thresholds for local sensors is the person-by-person optimization (PBPO) [15], where every sensor's threshold is optimized assuming the decision rules of all other sensors and FC are fixed. This report does not deal with the second problem so we are not going any further here.

2.2 Neyman-Pearson Detection Rule

The performance of detectors can be measured in many ways, and one popular indicator is the Receiver Operating Characteristic (ROC) curve. The ultimate goal of distributed detection systems is to make a global decision s_0 as in Figure 2.1, and $s_0 = 1$ or 0 correspond to claiming $\mathcal{H}_1$ or $\mathcal{H}_0$, respectively. There are four relevant values describing the accuracy of the global decision, which are $\Pr\{s_0 = 1 \mid \mathcal{H}_1\}$, $\Pr\{s_0 = 1 \mid \mathcal{H}_0\}$, $\Pr\{s_0 = 0 \mid \mathcal{H}_1\}$ and $\Pr\{s_0 = 0 \mid \mathcal{H}_0\}$. The first one and the second one is called the detection probability P_d and the false-alarm probability P_f, respectively. In most cases of detector designs, the values of P_d and P_f are a trade-off, which means that designers can hardly increase P_d but decrease P_f at the same time. The ROC curve plots these two values of the global decision, with P_f being the x-axis and P_d being the y-axis. A typical ROC curve is illustrated as follows:

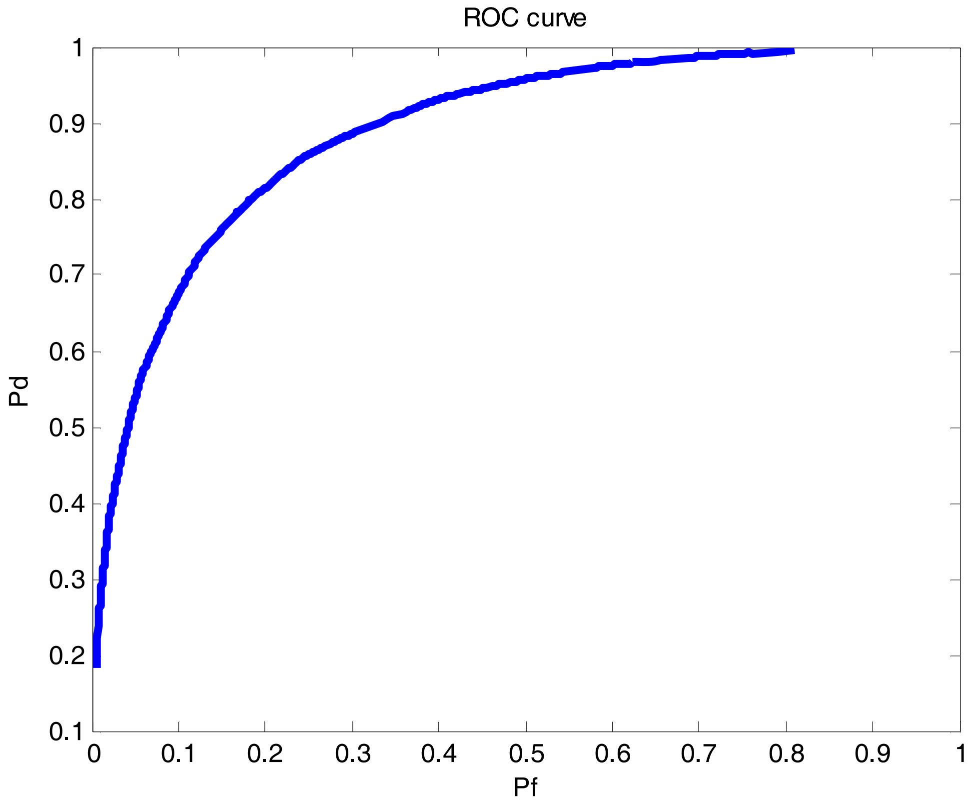

Figure 2.2: An example of ROC curve

According to classical distributed detection theory, a detector generally calculates a statistic as a function of the received reports from all sensors and then compares the statistic with a predetermined threshold. If the statistic is larger than the threshold, the detector claims the event happens, vice versa. The P_d and P_f of the detector move along the ROC curve as the threshold changes, that is, the detector operates along the ROC curve. A reasonable detector design always leads to a concave ROC curve above the straight line connecting points (0,0) and (1,1). To see this, consider a detector which tosses a biased coin in making decision whether the event happens. The performance of the detector moves along the straight line connecting points (0,0) and (1,1) as the bias of the coin changes, without using any information of observations. Therefore, any detectors utilizing the information of observations guarantees a better performance, which in turn results in an ROC curve above the straight line.

In the cases of binary hypothesis test, namely the detector decides only between $\mathcal{H}_0$ and $\mathcal{H}_1$, a popular detection rule LRT proposed by Neyman-Pearson [11]. It guides how to obtain the maximum detection probability given a value of acceptable false alarm probability. Specifically, denote $\mathbf{x}$ the received signal. The probability distributions of $\mathbf{x}$ depend on the underlying hypotheses, which are denoted as $p(\mathbf{x};\mathcal{H}_0)$ and $p(\mathbf{x};\mathcal{H}_1)$. The theory is stated as follows:

To maximize P_d for a given $P_f = \alpha$, decide $\mathcal{H}_1$ if

$$L(\mathbf{x}) = \frac{p(\mathbf{x};\mathcal{H}_1)}{p(\mathbf{x};\mathcal{H}_0)} > \gamma$$

where the threshold γ is found from

$$P_f = \int_{\{\mathbf{x}:L(\mathbf{x})>\gamma\}} p(\mathbf{x};\mathcal{H}_0)\,d\mathbf{x} = \alpha$$

The function $L(\mathbf{x})$ is indeed the ratio between the two likelihood functions, and that is why the entire test is called Likelihood Ratio Test.

LRT assumes the prior knowledge $p(\mathbf{x};\mathcal{H}_0)$ and $p(\mathbf{x};\mathcal{H}_1)$ is available. However, in many cases there are unknown parameters in the prior knowledge, for example the prior knowledge is probably $p\left(\mathbf{x};\theta_0,\mathcal{H}_0\right)$ and $p\left(\mathbf{x};\theta_1,\mathcal{H}_1\right)$, where θ_0 and θ_1 are unknowns. LRT then serves as the foundation of modification in such cases. Two major approaches to these situations are Bayesian approach and GLRT.

Bayesian approach considers the unknown parameters as realizations of random variables and assigns to each of them a prior PDF. Following the previous example of binary hypothesis, it has

$$p\left(\mathbf{x};\mathcal{H}_0\right)=\int p\left(\mathbf{x}\mid\theta_0;\mathcal{H}_0\right)p\left(\theta_0\right)d\theta_0$$

$$p\left(\mathbf{x};\mathcal{H}_1\right)=\int p\left(\mathbf{x}\mid\theta_1;\mathcal{H}_1\right)p\left(\theta_1\right)d\theta_1$$

It then applies the optimal LRT and decides $\mathcal{H}_1$ if

$$\frac{p\left(\mathbf{x};\mathcal{H}_1\right)}{p\left(\mathbf{x};\mathcal{H}_0\right)}=\frac{\int p\left(\mathbf{x}\mid\theta_1;\mathcal{H}_1\right)p\left(\theta_1\right)d\theta_1}{\int p\left(\mathbf{x}\mid\theta_0;\mathcal{H}_0\right)p\left(\theta_0\right)d\theta_0}>\gamma$$

This approach has to make further assumption about the unknowns θ_0 and θ_1, and it often require multidimensional integrations, which are often too complicated to implement in practice.

GLRT still borrows the way LRT detects because it is the optimal rule when the prior knowledge of probability distribution is known. Instead of making the probabilistic assumption of the θ_0 and θ_1, GLRT replaces the unknown parameters by their maximum likelihood estimates. This replacement is straight forward but there are no optimal criteria involved, and it has been shown that GLRT is generally not the best detector in the cases where unknown parameters are present. Although it is probably not optimal, it is still a popular approach in practice due to easy implementations.

In analyzing detectors for situations with unknown parameters, one often generates the optimal LRT by replacing the estimates of the unknowns with the real values, which are inaccessible in practice. It is the so-called clairvoyant detector, which is of theoretical interests for it serves as the performance upper bound for all possible detectors.

2.3 Review of Wireless Sensor Networks

WSN for detection are distributed detection systems using wireless technology, and sensors are not physically wired to FC. Each sensor is equipped with an antenna and transmits its report through the air. We still focus on the model of canonical distributed detection system, and note that the parallel channels from sensors to FC can be realized using orthogonal wireless transmission schemes such as TDMA, FDMA or CDMA. In developing the signal processing algorithms in WSN for detection, the designer often confronted not only the scarce resource constraints already appearing in the classical distributed detection systems, but also the unreliable channels. The effects of channel fading and interference in wireless channels invoke another uncertainty to WSN, and open up another dimension for the system design. The system block diagram is illustrated below,

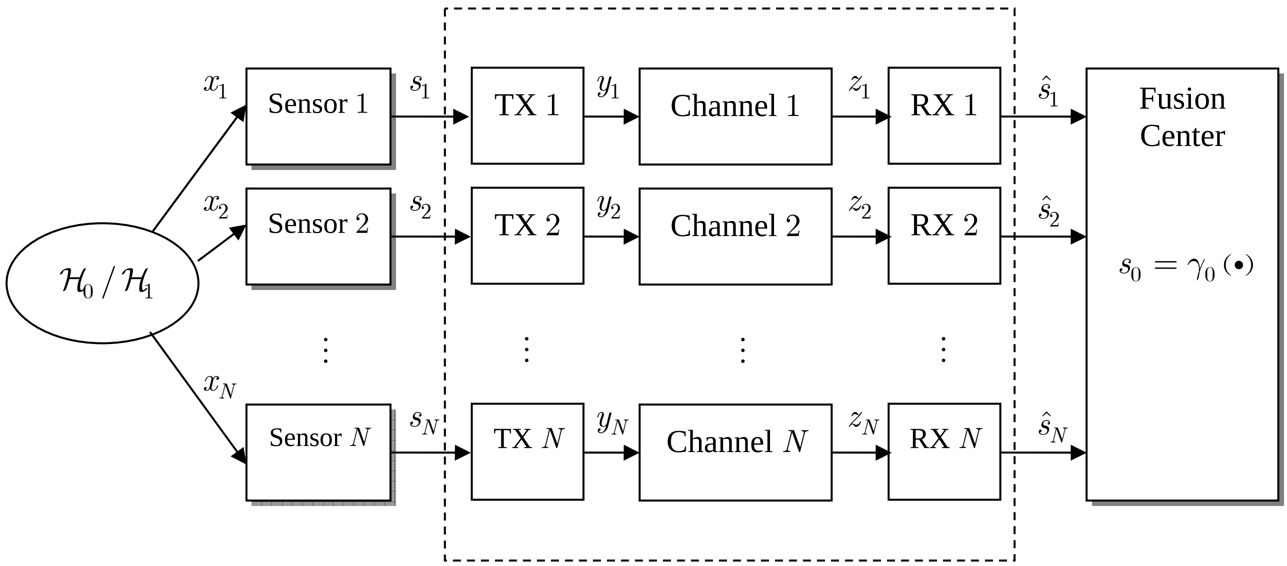

Figure 2.3: Canonical distributed detection systems with channel blocks

In the block diagram, the blocks in the dotted line can be regarded as the communication block. A practical and straightforward approach [16] is to separate the global decision task into a two stage process – first, $\hat{s}_i$ is used to infer about s_i, and then apply the optimum fusion rule based on s_i. This methodology treats the communication block and the classical distributed detection system as two independent parts so that once the communication problems are solved, the fusion rules developed in the field of classical distributed detection can be applied directly.

A more basic question arises whether one can design the wireless communication block alone irrespective of the signal processing algorithm at the local sensors and FC. The answer is unfortunately negative. When one designs the communication block, all the efforts are actually placed in recovery the s_i's from the $\hat{s}_i$'s. However, the ultimate goal is not to recover the s_i's but to infer the underlying hypothesis, and any deviating intermediate processing of the signal flow is vulnerable to losing information for inference. It means that the signal processing algorithms designed for the classical distributed detection systems cannot be implemented directly to the systems with unreliable channels, such as WSN. Designers have to regard the detection problem and the unreliable communication block as a whole integration, and then derive the optimal strategy in terms of the detection performance.

Similar problems appear in detection using WSN that signal processing algorithm at FC and at local sensors should be redesigned taking into account the channel state information (CSI). Some channel-aware algorithms have been proposed and indeed perform better than directly applying the classical detection methods. For the former problem, if the local sensor decision rules are given, the fusion rule design goes back to the centralized detection problem and classical detection rules apply, for example the LRT rule. The previously mentioned two-stage process is shown to approximate the optimal LRT rule when SNR is high, which is often a demanding requirement for WSN. Note that the degree of knowledge of channel actually affects the corresponding fusion rules, for example, whether FC can coherently detects or not results in different fusion rules.

The proposal [5] gives an important example of fusion rule design incorporating CSI. In its scenario of canonical WSN, in addition to knowing full CSI, FC has to know the performances of all local sensors in terms of local detection and false alarm probability. Remind that given the local sensor performance, the fusion rule design goes back to the centralized detection problem, and the optimal fusion rule is LRT. However, the formula of LRT is too complicated and the paper then proposes its approximations in high and low SNR, termed Λ_1 and Λ_2. The low-SNR fusion rule assembles MRC statistic for diversity combining, which in turn motivates a heuristic fusion rule assembling EGC, termed Λ_3. The paper also provides a new fusion rule called LRT-CS which requires only the channel statistics instead of the instantaneous CSI. Surprisingly, the high-SNR approximation of LRT-CS turns to Λ_1 and low-SNR approximation of low-SNR turns to the heuristic alternative Λ_3. Simulations show that the proposed LRT-CS rule outperform Λ_1 and Λ_3, and it performs better than Λ_2 for most practical SNR values.

After surveying relevant papers, we find that in general, most papers in channel-aware fusion rule assume that the local sensors' decision rules, or alternatively their local detection performance characterized by detection probabilities and false-alarm probabilities, are known to FC. However, these scenarios ignore the possibly unknown sensor responses to the occurrence of the interested event. In our work, the false-alarm probability of sensors is properly assumed not to change significantly with the environments because false-alarm probability is defined under the condition where the event is not happening. However, the detection probability is probably varying with the intensity of the event. For example, consider a sensor network deployed to monitor the rise in temperature in a room to detect the outbreak of a fire. In practice, the characteristics of a fire are uncertain, e.g. the mean temperature may vary from 100 to 1000 degrees depending on the severity of the fire or the type of the fire. Moreover, the characteristics of the fire are probably time-varying. As is introduced above, when there are unknowns in the PDF of the received signals, the optimal LRT rule is not applicable anymore and the alternative method GLRT puts the ML estimate of the unknown into the original LRT statistic and then compare it with the predetermined threshold. Our work starts from this popular GLRT and tries to simplify the complex ML estimates. However, the birth of GLRT is from heuristic thoughts so that it works fine in practice but not optimally in theory. Moreover, the formulation of GLRT statistic inherited from LRT is too complicated to analyze even if the ML estimates are simplified. These two main reasons motivate our design of new simple detection rule, which even performs far better than GLRT. In the following sections, the performance of the proposed simple detection rule is analyzed and the effects of channel impairments on the performance are also investigated. Finally, the corresponding power allocation strategy is proposed, which further improves our system in terms of ROC curves.

Chapter 3

GLRT Based Detection Method

3.1 System Model

Consider WSN of N identical sensors for detecting the occurrence of the event of interest. Specifically, the sensors monitor a certain parameter in this event. If this parameter exceeds a certain value, they should claim the event happening and transmit their decisions to the FC. The FC then combines all these decisions from the sensors and makes a global decision on whether the event is happening. Detailed system model description for each stage is introduced as follows. In the first stage, the status of the event can be regarded as binary hypothesis with $\mathcal{H}_0$ and $\mathcal{H}_1$ denoting the absence and presence of the event, respectively. Each sensor makes its binary decision on the hypothesis, transmitting $s_i = 1$ if it claims $\mathcal{H}_1$ and remaining silent if it claims $\mathcal{H}_0$, i.e., $s_i \in \{0,1\}$. That is, the reports to FC are hard decisions. Assume uniformity of the phenomenon of interests in the area where WSN is deployed, so that each sensor subjects to a known identical false-alarm probability $p_f = \Pr\{s_i = 1 \mid \mathcal{H}_0\} \triangleq \pi_0$, which can be measured before deploying the WSN. Assume also that each sensor possesses an unknown identical detection probability $p_d = \Pr\{s_i = 1 \mid \mathcal{H}_1\} \triangleq \pi_1$, which has to be estimated at FC.

The sensors then transmit their 1-bit decisions s_i through binary symmetric channels with different cross-error probabilities. Assume FC knows the CSI and is in turn aware of these cross-error probabilities. The signals received by FC are denoted as r_i and

$$\Pr\{r_i = 1\} = \begin{cases} \pi_0(1-\varepsilon_i) + (1-\pi_0)\varepsilon_i, & \text{(event is absent)}, \\ \pi_1(1-\varepsilon_i) + (1-\pi_1)\varepsilon_i, & \text{(event is present)}. \end{cases} \tag{3.1}$$

Receiving $\mathbf{r} = [r_1, r_2, \ldots, r_N]$ the FC then applies the fusion rule and makes global decision $s_0 = f(\mathbf{r})$ on the event. The system then has a global performance featured by the global detection probability $P_d = \Pr\{s_0 = 1 \mid \mathcal{H}_1\}$ and global false-alarm probability $P_f = \Pr\{s_0 = 1 \mid \mathcal{H}_0\}$. Detector performance of the system is evaluated using the ROC curve. The block diagram is illustrated as follows:

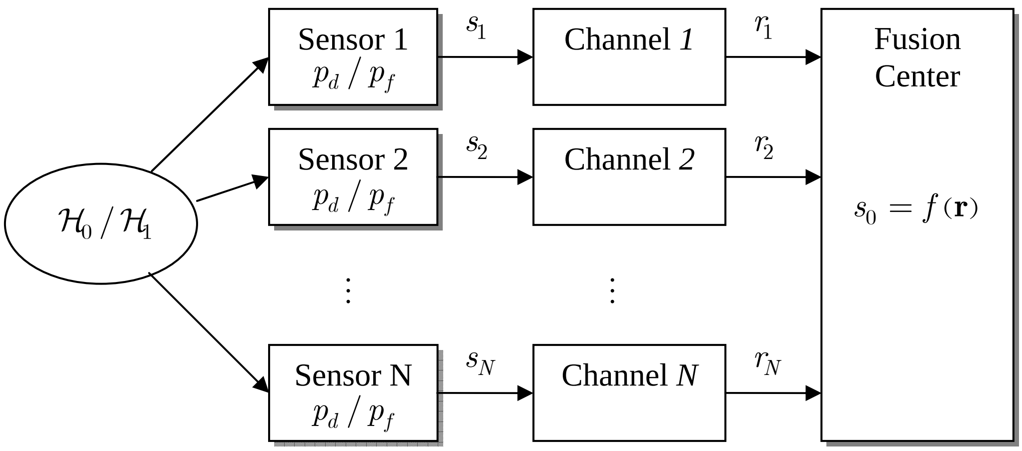

Figure 3.1: System model

3.2 GLRT based detection method

Assume that the set of Bernoulli random variables $\{r_i\}$ are conditionally independent given the event under the test, the joint probability mass functions of

$\mathbf{r} = [r_1, r_2, \ldots, r_N]$ under π_0 and π_1 are

$$p(\mathbf{r};\pi_m) = \prod_{i=1}^{N} \left[(1-2\varepsilon_i)\pi_m + \varepsilon_i\right]^{r_i} \left[-(1-2\varepsilon_i)\pi_m + (1-\varepsilon_i)\right]^{1-r_i},\ m = 0,1 \quad (3.2)$$

The binary hypothesis thus reforms as

$$\begin{cases} \mathcal{H}_0 : \mathbf{r} \sim p(\mathbf{r};\pi_0) & \text{(event is absent),} \\ \mathcal{H}_1 : \mathbf{r} \sim p(\mathbf{r};\pi_1), \pi_1 > \pi_0 & \text{(event is present).} \end{cases} \quad (3.3)$$

As described in Chapter 2, a commonly used detection method is GLRT, which claims $\mathcal{H}_1$ if

$$\ln \frac{p(\mathbf{r};\hat{\pi}_1)}{p(\mathbf{r};\pi_0)} = \begin{pmatrix} \sum_{i=1}^{N} r_i \log\left[\dfrac{(1-2\varepsilon_i)\hat{\pi}_1 + \varepsilon_i}{(1-2\varepsilon_i)\pi_0 + \varepsilon_i}\right] \\ + \sum_{i=1}^{N} (1-r_i) \log\left[\dfrac{(1-2\varepsilon_i)\hat{\pi}_1 - (1-\varepsilon_i)}{(1-2\varepsilon_i)\pi_0 - (1-\varepsilon_i)}\right] \end{pmatrix} \geq \gamma \quad (3.4)$$

where $\hat{\pi}_1$ is the ML estimation of π_1, and γ is the predetermined threshold. The ML estimation solves $\dfrac{\partial \ln p(\mathbf{r};\pi_1)}{\partial \pi_1} = 0$, which turns to

$$\sum_{i=1}^{N} \frac{r_i}{\pi_1 + [\varepsilon_i/(1-2\varepsilon_i)]} + \sum_{i=1}^{N} \frac{1-r_i}{\pi_1 - [(1-\varepsilon_i)/(1-2\varepsilon_i)]} = 0 \quad (3.5)$$

3.3 Proposed Simplified ML solution

To obtain the ML solution of π_1, one has to solve certain roots of the polynomial with order $N-1$. It can probably be achieved by numerical techniques, but no analytic solution exists. Thus for higher SNR, an approximated solution is proposed.

In the high SNR case, ε_i's are small and we has

$$\frac{\varepsilon_i}{1-2\varepsilon_i} = \left(1 + 2\varepsilon_i + 4\varepsilon_i^2 + \cdots\right)\varepsilon_i \approx \varepsilon_i \quad \text{and}$$

$$\frac{1-\varepsilon_i}{1-2\varepsilon_i} = \left(1 + 2\varepsilon_i + 4\varepsilon_i^2 + \cdots\right)(1-\varepsilon_i) \approx 1 + \varepsilon_i \quad (3.6)$$

by neglecting the high order terms. With (3.6), Equation (3.5) approximates to

$$\sum_{i=1}^{N}\frac{r_i}{\pi_1+\varepsilon_i}+\sum_{i=1}^{N}\frac{1-r_i}{\pi_1-(1+\varepsilon_i)}=\sum_{i=1}^{N}\frac{r_i(\pi_1-1-\varepsilon_i)+(1-r_i)(\pi_1+\varepsilon_i)}{\pi_1^2-\pi_1-\varepsilon_i(1-\varepsilon_i)}=0 \quad (3.7)$$

Again by keeping only the first-order term in the denominator in each summand and with rearrangement, (3.7) then becomes

$$\sum_{i=1}^{N}\frac{\pi_1+\varepsilon_i-(2\varepsilon_i+1)r_i}{\pi_1^2-\pi_1-\varepsilon_i}=0 \quad (3.8)$$

Also assume that ε_i's are small so that $\pi_1^2-\pi_1-\varepsilon_i \approx \pi_1^2-\pi_1$, and (3.8) is further simplified to

$$\frac{1}{\pi_1^2-\pi_1}\sum_{i=1}^{N}\left[\pi_1+\varepsilon_i-(2\varepsilon_i+1)r_i\right]=0 \quad (3.9)$$

Thus, as long as $\pi_1 \neq \{0,1\}$, $\hat{\pi}_1$ is obtained by solving

$$\sum_{i=1}^{N}\left[\pi_1+\varepsilon_i-(2\varepsilon_i+1)r_i\right]=0 \quad (3.10)$$

and the resulting approximated ML scheme is

$$\hat{\pi}_1=\frac{1}{N}\sum_{i=1}^{N}\left[(1+2\varepsilon_i)r_i-\varepsilon_i\right] \quad (3.11)$$

3.4 Computer Simulations

To examine the accuracy of the proposed simplified ML estimate, this work uses mean square error (MSE) as the indicator of accuracy. The parameter to be estimate $\Pr\{s_i=1 \mid \mathcal{H}_1\}=\pi_1$ is set to be 0.4, and we normalize the MSE to this value. Figure 3.2 shows the normalized MSE versus the mean cross-error probability.

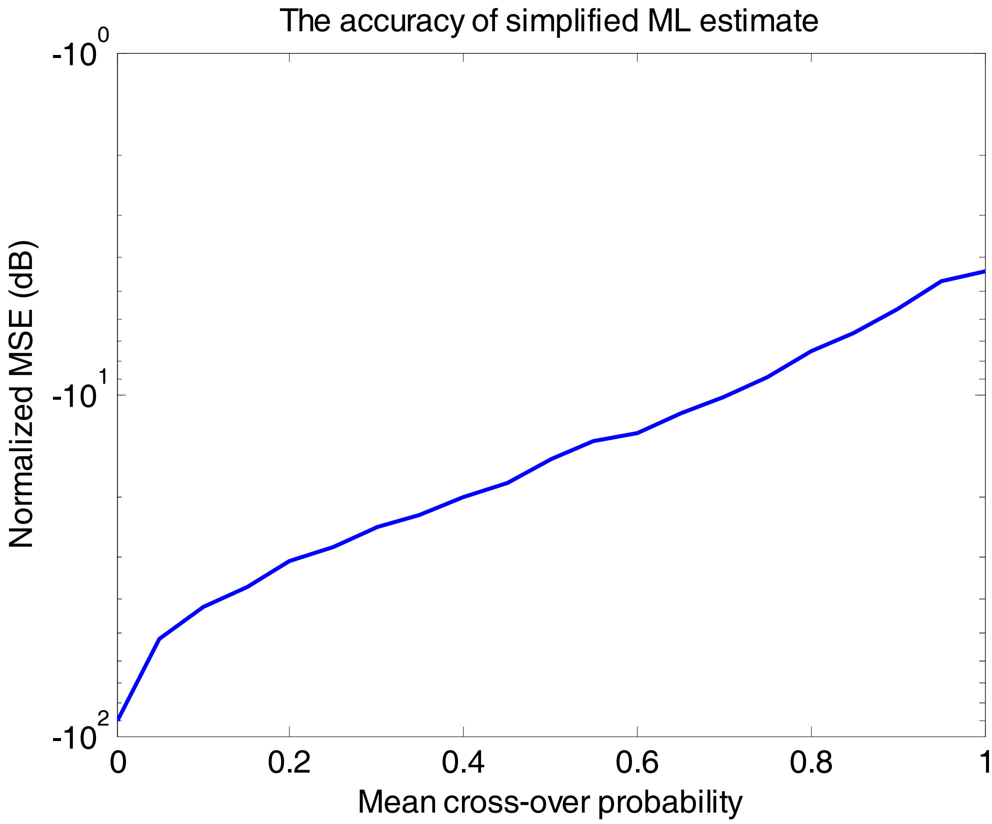

Figure 3.2: Accuracy of the proposed simplified ML estimate

Simulations tell that the proposed ML estimate approximates closely to the real one when the mean cross-error probability is small, for example below 0.2.

3.5 Discussion

Although the proposed Formula (3.11) is only an approximation of the true ML solution, the property of being affine in the data r_i's makes it more attractive and potentially amenable for analysis. Simulations show that the mean square error between Equation (3.11) and the true ML solution is quite small when mean cross-error probability is small. Accordingly, the performances of GLRT by $\hat{\pi}_1$ and the true ML solution are also quite close. The solution (3.10) also has an appealing interpretation. Let's consider the case where each communication link subjects to an identical cross-over probability, i.e., $\varepsilon_i = \varepsilon$ for $1 \le i \le N$. The received data r_i's are thus regarded as i.i.d. Bernoulli random variables with "success probability" $\Pr\{r_i = 1\} = \pi_1(1-\varepsilon) + (1-\pi_1)\varepsilon$ when change is present. According to [17] and the invariant property of the ML estimate [18], the exact ML solution is obtained as

$$\begin{aligned}\hat{\pi}_1^{(\varepsilon)} &= \frac{1}{1-2\varepsilon}\left[\frac{1}{N}\sum_{i=1}^{N} r_i - \varepsilon\right] \\ &= \frac{1}{N}\sum_{i=1}^{N}\frac{r_i}{1-2\varepsilon} - \frac{\varepsilon}{1-2\varepsilon} = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{r_i}{1-2\varepsilon} - \frac{\varepsilon}{1-2\varepsilon}\right]\end{aligned}. \tag{3.12}$$

In high SNR case where ε is small, we have $(1-2\varepsilon)^{-1} = (1 + 2\varepsilon + 4\varepsilon^2 + \cdots)$ Keeping only the zero-th and first order terms, (3.12) becomes

$$\hat{\pi}_1^{(\varepsilon)} \approx \frac{1}{N}\sum_{i=1}^{N}\left[(1+2\varepsilon) r_i - \varepsilon\right]. \tag{3.13}$$

Hence, the proposed approximate ML estimate (3.11) can be regarded as a direct modification of $\hat{\pi}_1^{(\varepsilon)}$ in (3.12) to take non-uniform communication link errors into consideration.

We can also interpret the simplified ML estimate as a modification of the voting scheme. If the sensors report their signal through perfect channels, the fusion center actually receives $r_i = s_i$ and then makes an estimate on $\Pr\{s_i = 1\}$. The corresponding best unbiased estimator uses the voting strategy, which uses the mean value as the estimate of $\Pr\{s_i = 1\}$. Assume now s_i's pass through BSC's with identical cross-over probability ε, we have $\Pr\{r_i = 1\} = \pi_1 - 2\varepsilon\pi_1 + \varepsilon$ when $\mathcal{H}_1$ is true. The voting rule or the mean of r_i's is thus $\pi_1 - 2\varepsilon\pi_1 + \varepsilon$, i.e.,

$$\frac{1}{N}\sum_{i=1}^{N} r_i = \pi_1 - 2\varepsilon\pi_1 + \varepsilon = \pi_1(1-2\varepsilon) + \varepsilon \tag{3.14}$$

To obtain π_1 from (3.14), after manipulation we arrive at

$$
\begin{aligned}
\pi_1 &= \frac{\frac{1}{N}\left(\sum_{i=1}^{N}\left(r_i-\varepsilon\right)\right)}{1-2\varepsilon} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{r_i}{1-2\varepsilon}-\frac{\varepsilon}{1-2\varepsilon}\right) \\
&\approx \frac{1}{N}\sum_{i=1}^{N}\left(\left(1+2\varepsilon\right)r_i-\varepsilon\right)
\end{aligned}
\tag{3.15}
$$

We arrive at the same conclusion as (3.13), which suggests the simplified ML estimate is actually nothing but a modification from the natural voting scheme.

3.6 Summary

An accurate approximation formula (3.11) to the true ML solution is derived and it has a natural interpretation related with the straightforward voting scheme. This formula is more tractable in that it is an affine function in the received signal r_i's. To accomplish the GLRT test, the FC then adopts this simplified ML estimate in the LRT statistic and compares it with a predetermined threshold. However even with the simplified formula, the achievable detection performance of GLRT, in particular the impact from channel impairments, remains quite difficult to characterize especially when the number of sensors is finite. It motivates us to propose an alternative detection rule which can exploit the affine nature of $\hat{\pi}_1$ and result in analytic study of the link error effects.

Chapter 4

Proposed Detection Method

4.1 Proposed Simple Detection Rule

It is shown that GLRT is merely a heuristic approach, nor does it involve any optimality criteria in deriving this rule. Even if the proposed ML approximation is simple, it helps little after being adopted into the GLRT statistic, which motivates another simpler fusion rule that can benefit from the ML approximation. Note that at receiving r_i's, the FC is actually applying the ML estimate $\hat{\pi}$ for $\Pr\{s_i = 1\}$. Because $\Pr\{s_i = 1 \mid \mathcal{H}_1\} = \pi_1$ and $\Pr\{s_i = 1 \mid \mathcal{H}_0\} = \pi_0$, ideally, $\hat{\pi} \approx \pi_1$ when $\mathcal{H}_1$ occurs and $\hat{\pi} \approx \pi_0$ when $\mathcal{H}_0$ occurs. A simpler and more natural alternative is to obtain $\hat{\pi}$ first and then compare it with the known $p_f = \pi_0$. More specifically, the FC can be designed to make the following decision

$$\begin{cases} \mathcal{H}_0 : \hat{\pi} - \pi_0 \leq \gamma \\ \mathcal{H}_1 : \hat{\pi} - \pi_0 > \gamma \end{cases} \tag{4.1}$$

where γ was the predetermined threshold.

The main advantage of the proposed decision rule (4.1) is that, unlike the GLRT in (3.4), the test statistic in (4.1) is affine in the estimate $\hat{\pi}$, and hence is affine in the received data r_i's. The proposed method also directly utilizes the parameter that really reflects the different hypotheses. Based on these attractive features, performance can be characterized analytically as shown below.

4.2 Performance Analysis

To proceed, let's write $\hat{\pi}$ as

$$\hat{\pi} = \frac{1}{N}\sum_{i=1}^{N}[(1+2\varepsilon_i)r_i - \varepsilon_i] = \underbrace{\frac{1}{N}\sum_{i=1}^{N}(1+2\varepsilon_i)r_i}_{\triangleq T} - \frac{1}{N}\sum_{i=1}^{N}\varepsilon_i \tag{4.2}$$

and T is substantially the equivalent test statistic. Since $r_i \in \{0,1\}$, T assumes a finite number of alphabets, which are to be specified. First, for each $0 \le k \le N$, define $I^{(k)} \triangleq \left\{I_1^{(k)}, I_2^{(k)} \cdots, I_{C_k^N}^{(k)}\right\}$ to be the collection of all distinct k-element subsets of $\{1,\cdots,N\}$, where $C_k^N \triangleq N!/[k!(n-k)!]$ and $I^{(0)} = \{\phi\}$. Each element in $I^{(k)}$ maps to a possible value of T, thus for each $0 \le k \le N$, let's define $S^{(k)}$ be the set consisting of all possible values of T when k sensors are active, that is,

$$S^{(k)} \triangleq \{T \mid k \text{ sensors are active}\} = \left\{S_1^{(k)}, S_2^{(k)}, \cdots, S_{C_k^N}^{(k)}\right\} \tag{4.3}$$

where $S_l^{(k)} := N^{-1}\left(k + 2\sum_{i\in I_l^{(k)}} \varepsilon_i\right)$. As a result,

$$T \in \bigcup_{k=0}^{N} S^{(k)}, \text{ where } S^{(0)} = \{0\} \tag{4.4}$$

Note from (4.4) that there are totally $C_0^N + C_1^N + \cdots + C_N^N = (1+1)^N = 2^N$ possible values of T.

To facilitate further investigation, assume without loss of generality that, for each $1 \le k \le N$, the elements in $S^{(k)}$ are arranged so that $S_1^{(k)} \le S_2^{(k)} \le \cdots \le S_{C_k^N}^{(k)}$, i.e.,

$$N^{-1}\left(k + 2\sum_{i\in I_1^{(k)}} \varepsilon_i\right) \le N^{-1}\left(k + 2\sum_{i\in I_2^{(k)}} \varepsilon_i\right) \le \cdots \le N^{-1}\left(k + 2\sum_{i\in I_{C_k^N}^{(k)}} \varepsilon_i\right) \tag{4.5}$$

Also, let $1 \le k_l \le N$ be such that

$$N^{-1}\left(k_l+2\sum_{i\in I_1^{(k_l)}}\varepsilon_i\right)\le \pi_0+\frac{1}{N}\sum_{i=1}^{N}\varepsilon_i+\gamma< N^{-1}\left(k_l+1+2\sum_{i\in I_1^{(k_l+1)}}\varepsilon_i\right) \quad (4.6)$$

By definition of the detection probability $P_d=\Pr\left\{T\ge \pi_0+\frac{1}{N}\sum_{i=1}^{N}\varepsilon_i+\gamma \,\middle|\, \mathcal{H}_1\right\}$, the lower bound for P_d is then

$$P_d \ge P_d^{(L)} \quad (4.7)$$

where based on (4.6) and (4.3),

$$\begin{aligned} P_d^{(L)} &= \Pr\left\{S^{(k_l+1)}\cup S^{(k_l+2)}\cup\ldots\cup S^{(N)} \mid \pi_1 \text{ is true}\right\} \\ &= \sum_{k=k_l+1}^{N}\sum_{l=1}^{C_k^N}\left\{\prod_{i\in I_l^{(k)}}\left[\left(1-2\varepsilon_i\right)\pi_1+\varepsilon_i\right]\prod_{i\notin I_l^{(k)}}\left[-\left(1-2\varepsilon_i\right)\pi_1+\left(1-\varepsilon_i\right)\right]\right\}. \end{aligned} \quad (4.8)$$

Similarly, for the false-alarm probability $P_f=\Pr\left\{T\ge \pi_0+\frac{1}{N}\sum_{i=1}^{N}\varepsilon_i+\gamma \,\middle|\, \mathcal{H}_0\right\}$ an associated lower bound can be obtained as

$$P_f \ge P_f^{(L)} \quad (4.9)$$

where

$$\begin{aligned} P_f^{(L)} &= \Pr\left\{S^{(k_l+1)}\cup S^{(k_l+2)}\cup\ldots\cup S^{(N)} \mid \pi_0 \text{ is true}\right\} \\ &= \sum_{k=k_l+1}^{N}\sum_{l=1}^{C_k^N}\left\{\prod_{i\in I_l^{(k)}}\left[\left(1-2\varepsilon_i\right)\pi_0+\varepsilon_i\right]\prod_{i\notin I_l^{(k)}}\left[-\left(1-2\varepsilon_i\right)\pi_0+\left(1-\varepsilon_i\right)\right]\right\}. \end{aligned} \quad (4.10)$$

Observe that the performance bounds in (4.8) and (4.10) depended on the link error probability ε_i's. This allows for further discussions on the impact of the channel effect on the detection performance, as in the next sections.

The performance a thirty-sensor WSN is simulated as the blue ROC curve in Figure 2.1. The channel model is the same as that in Chapter 4.4 and the average cross-over probability is 0.023, which is small enough to validate the bound derivation. The local detection and false-alarm probability are 0.6 and 0.4, respectively. The proposed bound is plotted as the red ROC curve. It can be seen that the proposed bound is tight enough to evaluate the performance of the proposed fusion rule.

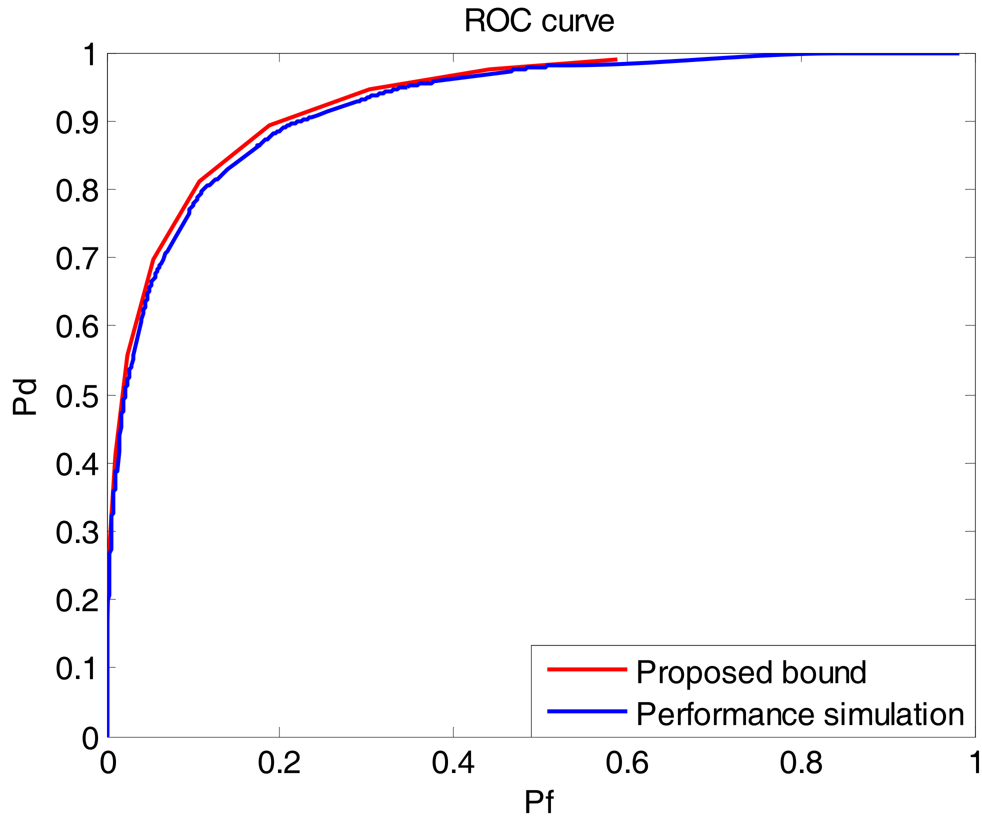

Figure 4.1: Accuracy of the proposed performance bound

4.3 Impact due to Channel Effects

The performance formulas in (4.8) and (4.10) remain non-linear functions of ε_i's. It is still difficult to assess the effects of non-ideal communication channels. Remember that the proposed ML solution is derived under the high-SNR assumption. By further exploiting the assumption that ε_i's are small, (4.8) and (4.10) can be simplified considerably, as in the next lemma.

Lemma 4.1: For small ε_i's,

$$P_d^{(L)} = \sum_{k=k_l+1}^{N} A_k + \left(\sum_{i=1}^{N} \varepsilon_i\right) \sum_{k=k_l+1}^{N} B_k \tag{4.11}$$

where

$$
\begin{aligned}
A_k &\triangleq C_k^N \left(1-\pi_1\right)^{N-k} \pi_1^k \\
B_k &\triangleq \pi_1^{k-1}\left(1-2\pi_1\right)\left(1-\pi_1\right)^{N-k-1}\left(C_{k-1}^{N-1}\left(1-\pi_1\right)-C_k^{N-1}\pi_1\right)
\end{aligned}
\tag{4.12}
$$

Similarly,

$$
P_f^{(L)} = \sum_{k=k_l+1}^{N} C_k + \left(\sum_{i=1}^{N}\varepsilon_i\right)\sum_{k=k_l+1}^{N} D_k \tag{4.13}
$$

where

$$
\begin{aligned}
C_k &\triangleq C_k^N \left(1-\pi_0\right)^{N-k} \pi_0^k \\
D_k &\triangleq \pi_0^{k-1}\left(1-2\pi_0\right)\left(1-\pi_0\right)^{N-k-1}\left(C_{k-1}^{N-1}\left(1-\pi_0\right)-C_k^{N-1}\pi_0\right)
\end{aligned}
\tag{4.14}
$$

Proof:

Based on (4.8) and (4.10), (4.11) and (4.13) are obtained by neglecting the high-order terms of ε_i's and then some manipulations. The binomial coefficient comes from the summation of all possible combinations given k sensors are active. □

While the bounds (4.8) and (4.10) are quite complicated functions of ε_i's, in the high-SNR regime, the detection performance is closely related to the summed cross-error probabilities, namely $\sum_{i=1}^{N}\varepsilon_i$. Still, the fact that $P_d^{(L)}$ and $P_f^{(L)}$ depend on $\sum_{i=1}^{N}\varepsilon_i$ does not explicitly indicate how the variation of $\sum_{i=1}^{N}\varepsilon_i$ influences the detection performance. However, under some reasonable assumptions on p_d and p_f, this work proves that minimizing $\sum_{i=1}^{N}\varepsilon_i$ guarantees a better detector performance in terms of the ROC curve, as precisely stated in the following theorem.

Theorem 4.2: Assume $\pi_0 < 0.5 < \pi_1$. Given a fixed false-alarm probability P_f, let

$P_d^{(L)}$ and $P_d'^{(L)}$ be two detection probability lower bounds associated with two different summed link errors $E = \sum_{i=1}^{N} \varepsilon_i$ and $E' = \sum_{i=1}^{N} \varepsilon_i'$, respectively. If $E' < E$, then $P_d'^{(L)} > P_d^{(L)}$. $\square$

Proof: See Appendix A

Theorem 4.2 suggests that, when the condition $\pi_0 < 0.5 < \pi_1$ is fulfilled, the global detection performance improves if the summed link error rate can be made small. The assumption $\pi_0 < 0.5 < \pi_1$ is actually not too demanding for any reasonable sensors. Inspired by Theorem 4.2, a sensor power allocation scheme for enhancing the global detection performance is developed next.

4.4 Proposed Power Allocation Strategy

Recall that the ith sensor transmits $s_i = 1$ when it claims $\mathcal{H}_1$ and transmits nothing when it claims $\mathcal{H}_0$. Namely, the sensors report their one-bit decisions using on-off keying to conserve energy. After incorporating the power allocation strategy, the ith sensor actually transmits $s_i \in \{0,1\}$ multiplied with an amplitude factor a_i, which is to be designed later on, and the corresponding power allocated to this sensor is $p_i = a_i^2$. Assume the communication channel between the ith sensor and the FC is flat and Rayleigh distributed with the current channel coefficient h_i, with the average power normalized to 1. Knowing these current channel coefficients, the FC can then apply the coherent detection and the received signal y_i from the ith sensor could be described by a commonly used discrete-time baseband model

$$y_i = h_i a_i s_i + n_i,\ 1 \le i \le N \tag{4.15}$$

where n_i is the zero-mean Gaussian noise of variance $N_0/2$. The corresponding cross-over probability of the ith link is then $\varepsilon_i = Q\left(\sqrt{\frac{h_i^2 p_i}{2N_0}}\right)$, where $Q(t) \triangleq (\sqrt{2\pi})^{-1}\int_t^{\infty} \exp\left[-u^2/2\right] du$ is the Q-function.

Under a total transmit power budget P, the optimization problem can be formally stated as

$$\begin{aligned} &\{p_i^*\}_{i=1}^N = \arg\min_{p_1,\ldots,p_N} \sum_{i=1}^N Q\left(\frac{h_i a_i}{\sqrt{2N_0}}\right) = \arg\min_{p_1,\ldots,p_N} \sum_{i=1}^N Q\left(\sqrt{\frac{h_i^2 p_i}{2N_0}}\right) \\ &\text{subject to } \sum_{i=1}^N p_i = P,\ p_i \ge 0 \end{aligned} \tag{4.16}$$

The optimization problem of the form (4.16) has been addressed in the context of MIMO wireless communications [19, 20]. Note that the cost function and the inequality constraint are convex and the equality constraint is linear. The optimization problem is thus convex and the Kuhn-Tucker conditions are necessary and sufficient conditions for finding $\{p_i^*\}_{i=1}^N$.

Define the Lagrangian function as:

$$\mathscr{L} = \sum_{i=1}^N Q(\sqrt{\gamma_i p_i}) - \sum_{i=1}^N u_i p_i + \lambda\left(\sum_{i=1}^N p_i - P\right),\ \gamma_i = \frac{h_i^2}{2N_0} \tag{4.17}$$

where $\{u_i\}_{i=1}^N$ and λ are KKT multipliers and Lagrange multiplier, respectively. According to the theory of optimization, $\{p_i^*\}_{i=1}^N$, $\{u_i\}_{i=1}^N$ and λ should satisfy the following conditions:

(1.) $\dfrac{\partial \mathscr{L}}{\partial p_i^*} = 0, \quad 1 \le i \le N$

(2.) $\sum_{i=1}^N p_i^* = P, \quad 1 \le i \le N$

(3.) $p_i^* \geq 0\,, \qquad 1 \leq i \leq N$

(4.) $u_i \geq 0\,, \qquad 1 \leq i \leq N$

(5.) $u_i p_i^* = 0\,, \qquad 1 \leq i \leq N$

Condition (1.) turns to

$$\sqrt{\frac{\gamma_i}{p_i^*}} e^{-p_i^* \gamma_i} = u_i - \lambda,\ 1 \leq i \leq N \tag{4.18}$$

For a fixed value of λ, u_i and p_i^* can be chosen for every i as follows

1. If there is a $P \geq p_i \geq 0$ that solves (4.18), then choose $u_i = 0$ and $p_i^* = p_i$

2. If there is no such p_i, choose $p_i^* = 0$ and $u_i = \infty$

Thus, the minimizer $\{p_i^*\}_{i=1}^N$ can be expressed as $p_i^* = \max(0, p_i)$.

To clarify further, demonstrated below is the flow chart of the power allocation procedure:

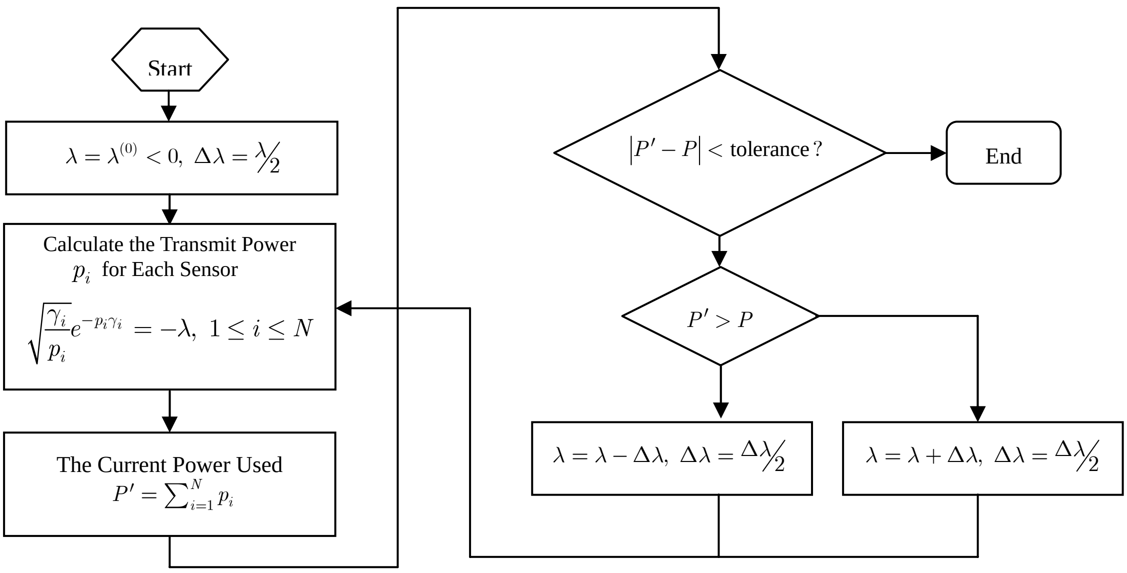

Figure 4.2: Flow chart of power allocation strategy

4.5 Summary

To solve the problems that GLRT is not optimal and the GLRT statistic is too complicated, a simpler and more straightforward fusion rule is proposed. The proposed fusion rule conserves the affine properties of the approximated ML estimate and is in turn affine in the received signals, which enables the performance analysis. A tight bounds of P_d and P_f are proposed and then simplified under the assumption that the SNR is moderately high. These bounds are derived not only to evaluate the performance, but also to facilitate investigation of the channel effects, which is accomplished by locating the simple channel-related term $\sum_{i=1}^{N} \varepsilon_i$ in the approximation formulas. However, it remains unclear how to improve the ROC curve because both of these approximations remain complicated functions of $\sum_{i=1}^{N} \varepsilon_i$. Despite of the fact, it is proved that under some more reasonable assumptions, minimizing $\sum_{i=1}^{N} \varepsilon_i$ guarantees a better performance. At the end, a power allocation strategy aiming at minimizing $\sum_{i=1}^{N} \varepsilon_i$ is proposed. Simulations of this work are demonstrated in the next chapter.

Chapter 5

Computer Simulations and Discussions

5.1 Computer Simulations

The performances are simulated for the proposed detection rule with and without transmit power allocation, and then are compared with those for GLRT. The simulations of the LRT performance are also provided, which serve as the upper bound of any possible detector designs. In all simulations, the channel coefficients are assumed flat and Rayleigh distributed with the average power normalized to 1.

Figure 5.1 shows the ROC curve for an WSN of twenty sensors with uniform local detection probability $p_d = 0.6$ and local false-alarm probability $p_f = 0.4$. The noise power is $N_0 = 0.05$. The blue line is the performance of the LRT detector with power allocation proposed in this work. The black solid and the red dash curves are the ROC curves of the proposed fusion rule with and without power allocation, respectively. The black solid and the red dash curves with circles are the ROC curves of GLRT with and without power allocation designed for the proposed fusion rule, respectively.

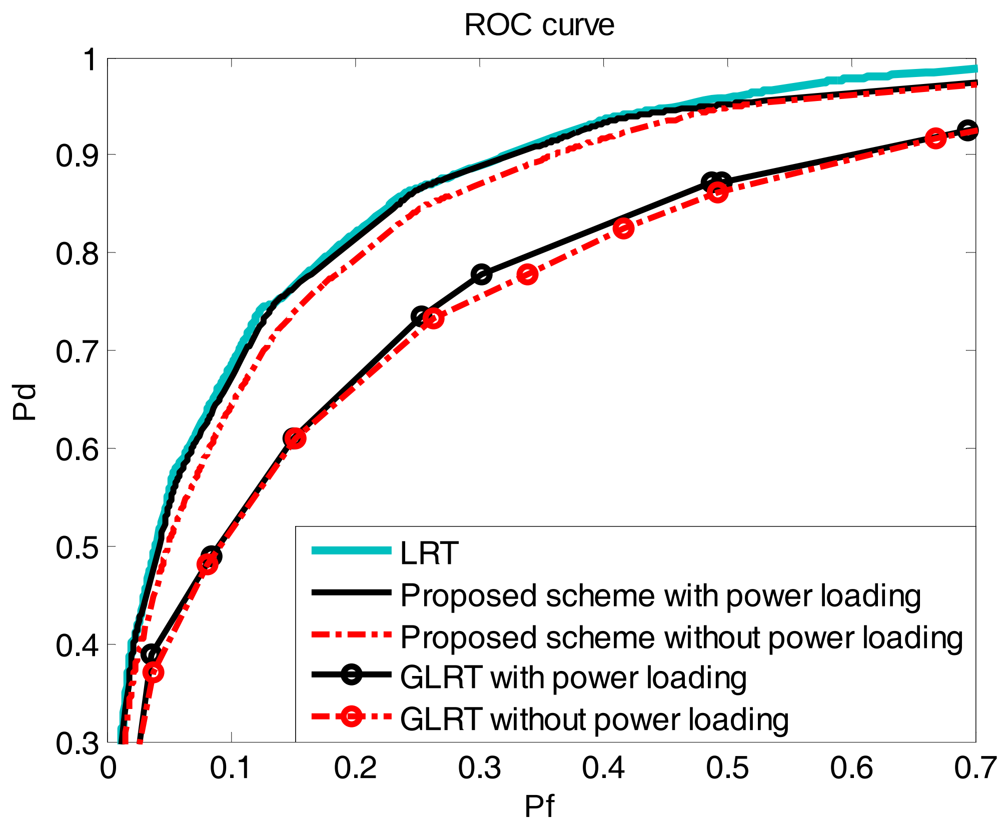

Figure 5.1: ROC curve of twenty-sensor network

From the above figure, the proposed fusion rule without power allocation (red dash line) outperforms GLRT in P_d by 10 to 15 percent given a value of P_f small enough to have practical interests. After power allocation, the performance of the proposed fusion rule even approaches the optimal LRT bound (blue solid line).

Figure 5.2 and Figure 5.3 are based on similar environment settings as Figure 5.1, but the number of sensors changes to 30 and 50, respectively. As can be seen, similar results appear but the extent of increase in P_d becomes smaller. The phenomenon will be described shortly in this section.

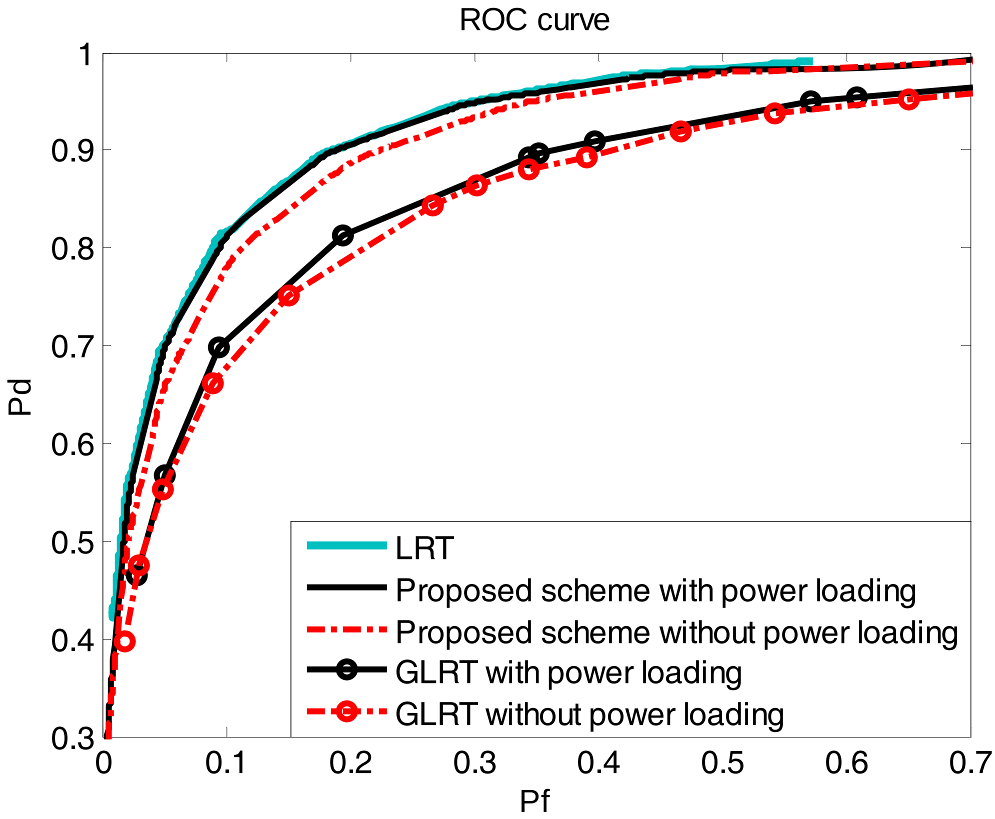

Figure 5.2: ROC curve of thirty-sensor network

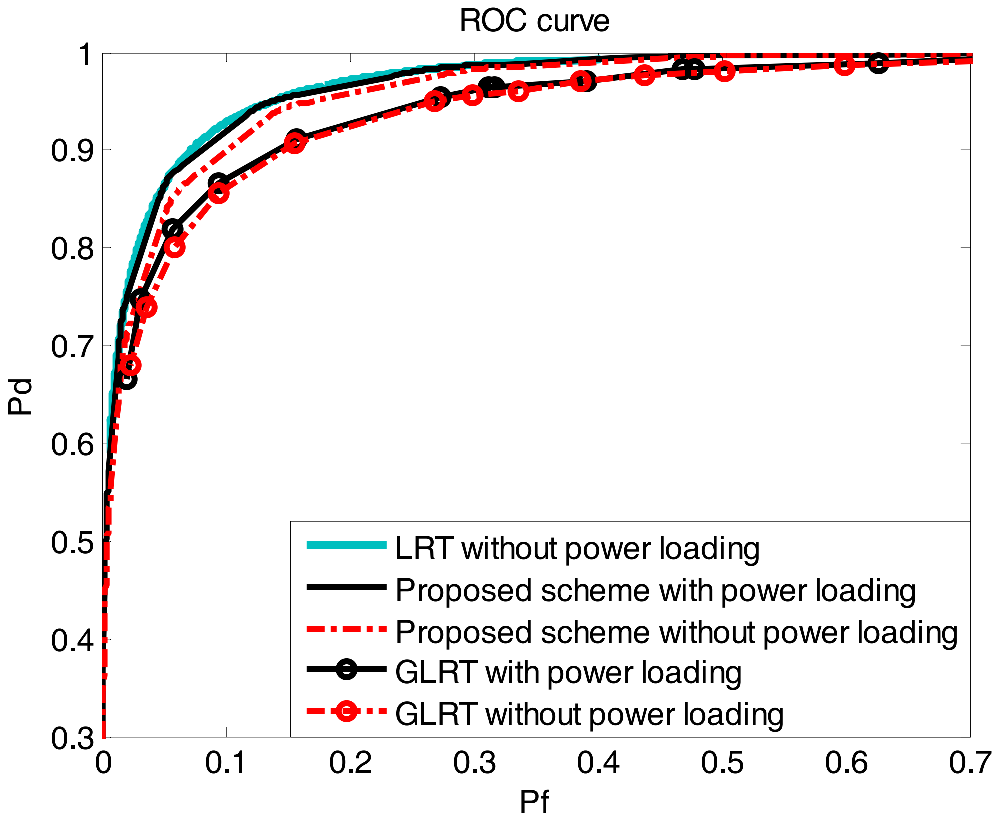

Figure 5.3: ROC curve of fifty-sensor network

Figure 5.4 illustrates the influence of the number of sensors on the relative increase in P_d of the proposed fusion rule with power allocation over GLRT, given the global $P_f = 0.1$. As the number of sensors increases, the detector is expected to perform better because it has more information-bearing reports to make a correct decision. Asymptotically, when the number of sensors grows to infinity, all detector designs that use the information in the received signals have so much information available that their ROC curves approach to the left-upper corner, same for GLRT. That is why the improvement of the proposed fusion rule over GLRT diminishes as the number of sensors increases, and the proposed fusion rule improves P_d significantly for smaller number of sensors.

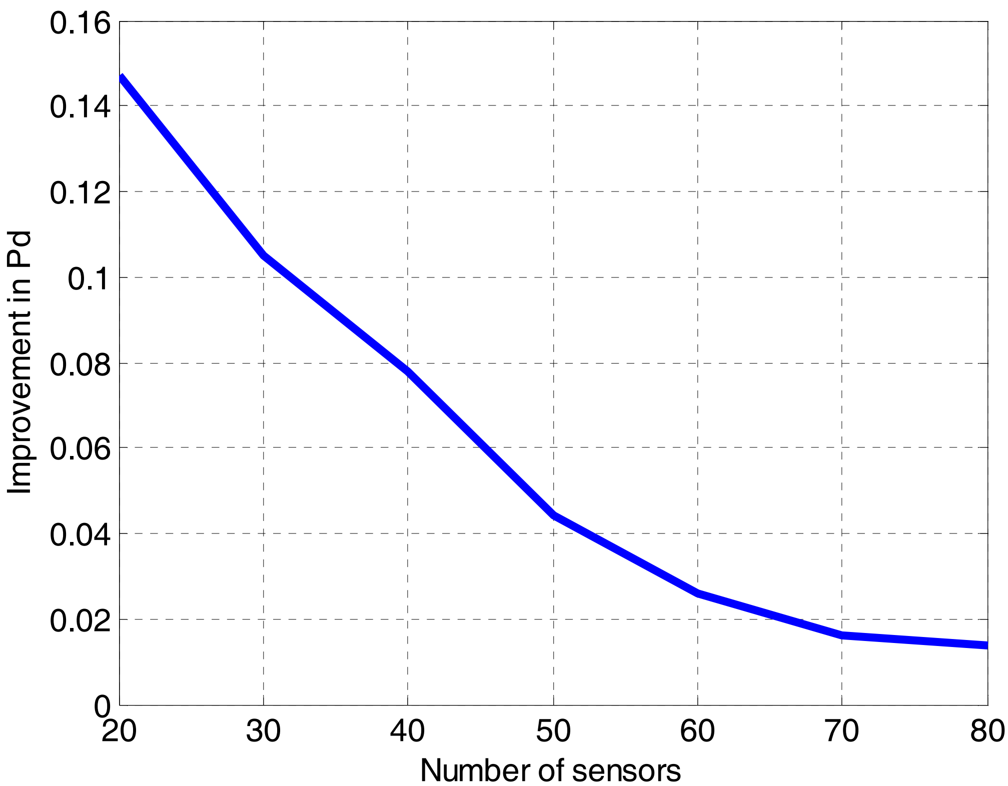

Figure 5.4: Relative increases in P_d over GLRT vs. number of sensors

Figure 5.5 demonstrates the relations of the relative increase in P_d of the proposed rule with power allocation over that without power allocation versus the average transmit power per sensor. The P_f is fixed to 0.1, and other environment settings remain the same, i.e., $p_d = 0.6$, $p_f = 0.4$ and $N_0 = 0.05$. Interestingly, the improvement in P_d is smaller when the average transmit power is too high or too low, and there is a peak improvement when the average transmit power is equal to about 0.8.

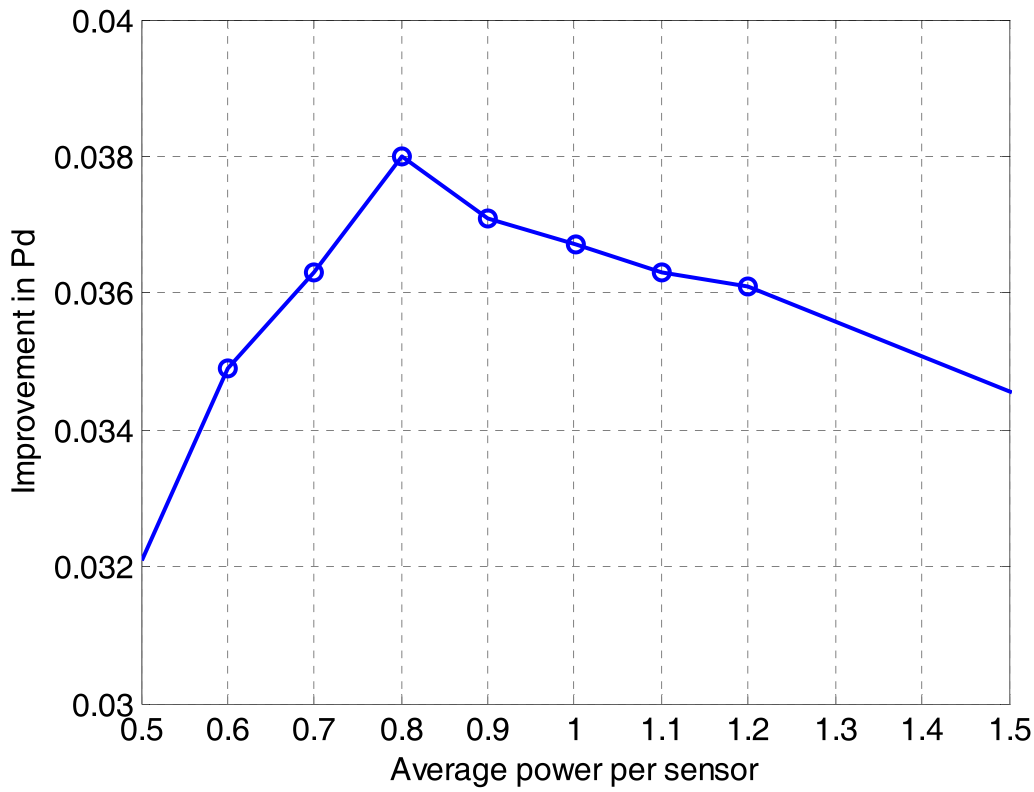

Figure 5.5: Relative increase in P_d from power allocation

vs. average power for $N_0 = 0.05$

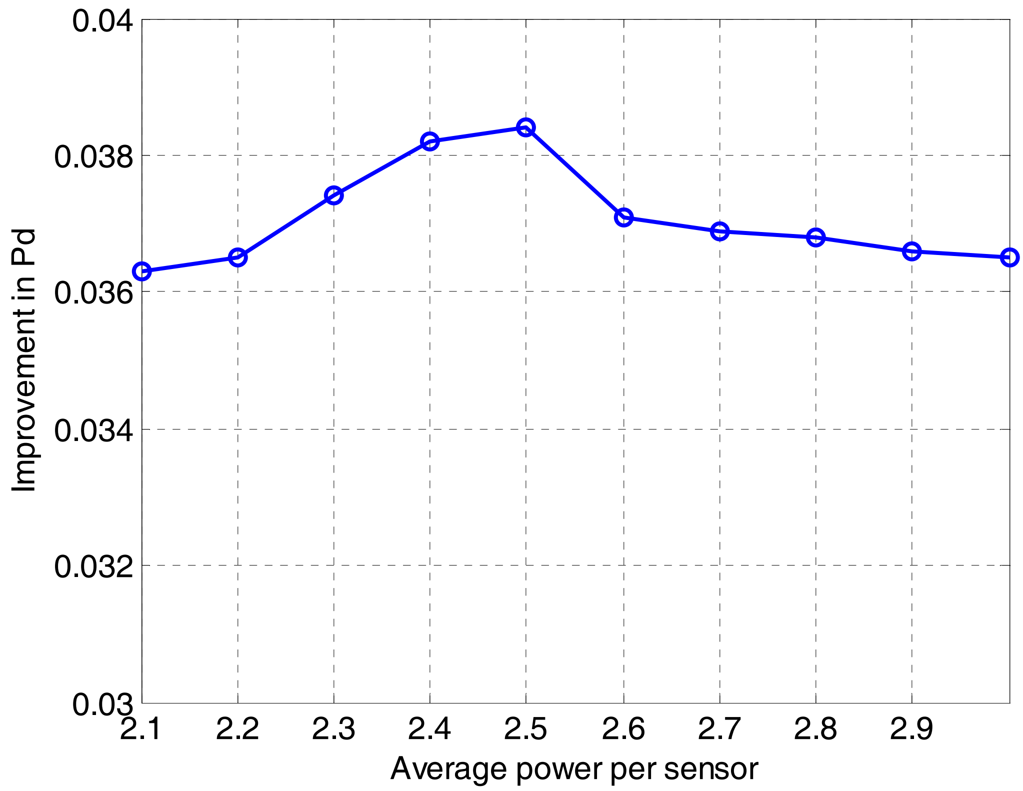

Figure 5.6: Relative increase in P_d from power allocation

vs. average power for $N_0 = 0.15$

Figure 5.6 also demonstrates the relations of the improvement in P_d from power allocation versus the average transmit power per sensor, but this time the noise variance increases by three times, i.e., $N_0 = 0.15$. Note that the peak of the improvement in P_d now moves to the average power of about 2.4, which is three times larger than the peak average power in Figure 5.5. The reason for the relation between Figure 5.5 and Figure 5.6 is explained as follows. In Equations (4.11) and (4.13), the channel effects only go into the value of ε_i's. Given a realization of h_i's, $\varepsilon_i = Q\left(\sqrt{h_i^2 p_i / 2N_0}\right)$ only depends on p_i and N_0. If the average power p_i and N_0 keeps the same ratio, they would produce the same value of ε_i's. The cross-error probabilities before and after power allocation will be the same for all the same ratio of

p_i to N_0. The effects of fixing p_i and changing N_0 match that of fixing N_0 and changing p_i. Specifically, the improvement in P_d at average power 0.8 in Figure 5.5 matches that at average power 2.4 in Figure 5.6.

To illustrate why too high or too low average power leads to smaller improvement in P_d, two ROC curves are illustrated for these two extreme scenarios. Figure 5.7 shows the case where the average transmit power is 0.1, which is extremely small, and Figure 5.8 shows the case where the average transmit power is 3, which is comparatively high.

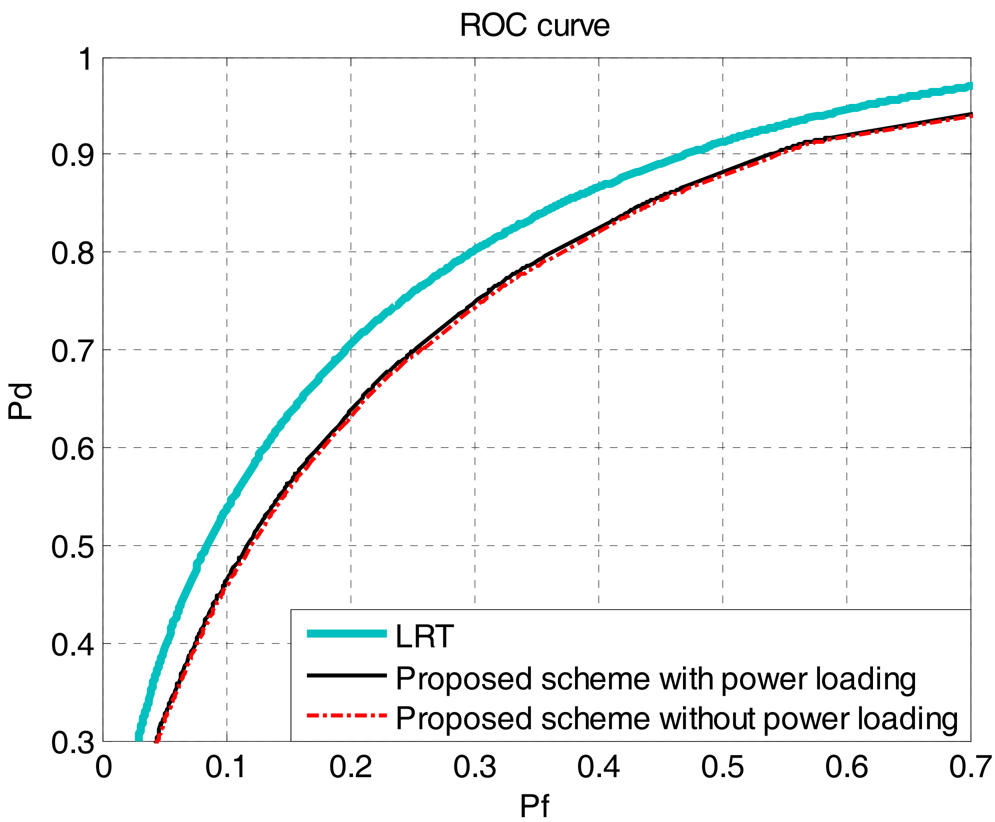

Figure 5.7: ROC curve of thirty-sensor network with average power 0.1

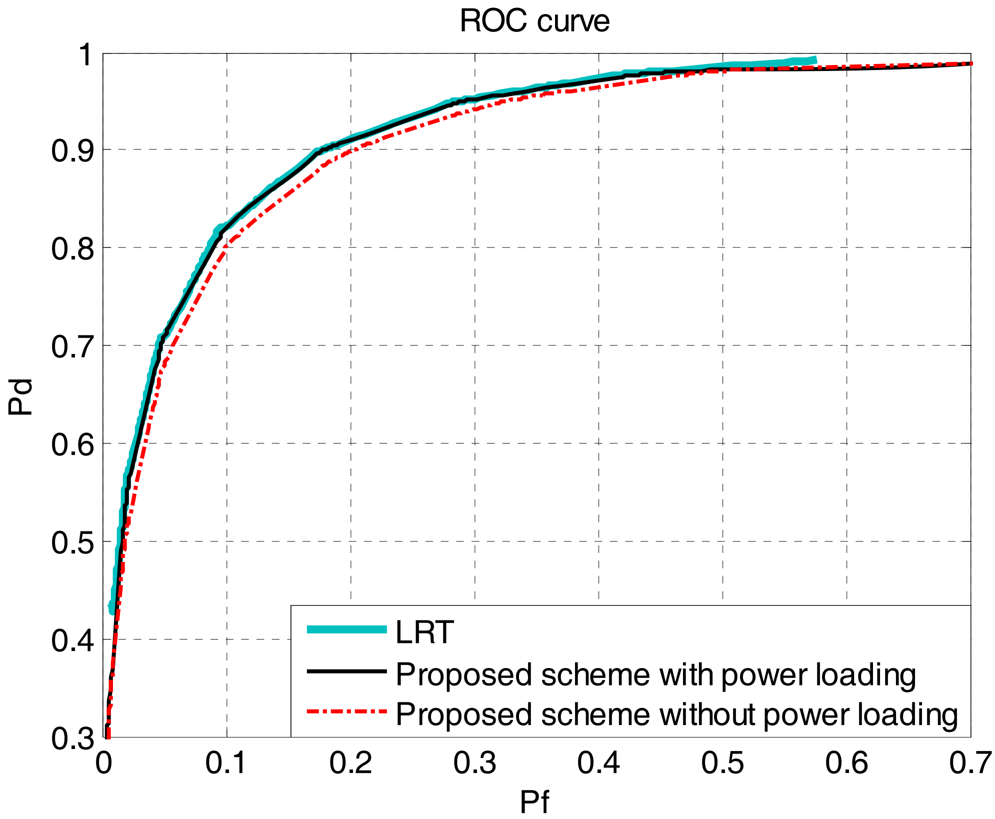

Figure 5.8: ROC curve of thirty-sensor network with average power 3

As expected, when there is no power allocation, the performance of the proposed fusion rule with higher transmit power is already better than that with lower transmit power. Also note that in the case of higher average power, the gap between the proposed fusion rule without power allocation and LRT (with power allocation) is already smaller than that in the low-average-power case. It is reasonable because when the average transmit power is high, even if there are no power allocations, the ε_i 's are already quite small. Power allocation cannot make the performance of the proposed detector better than that of LRT and thus the increase in P_d is small. It can also be explained as follows: the improvement in P_d becomes small because the power allocation changes ε_i 's from already small ones to smaller ones.

When the average transmit power is small, it is seen from Figure 5.7 that the proposed fusion rule without power allocation performs much poorly than LRT with power allocation. In contrast to high-average-power case where the improvement in P_d is somewhat "upper bounded" by LRT, the low average transmit power deprives the FC of the margin to tackle with the channel impairments. Thus, the ε_i's decrease insignificantly after power allocation, which make the improvement in P_d small.

5.2 Discussion on Proposed Method

Simulations indicate that the proposed detection rule outperforms GLRT significantly, in terms of ROC curve. After the transmit power allocation, the ROC curve of the proposed rule moves to the left-upper corner and even approaches the LRT bound. It means that the proposed method with power allocation is nearly optimal. Also, we can see that for any given value of P_f, the improvement in P_d has a peak at a certain average transmit power, termed p_{peak}. Average transmit power lower than p_{peak} is too small to have large improvement in P_d, and the ROC curve of the proposed method after power allocation still has gap from that of LRT. Average transmit power greater than p_{peak} is large enough to raise the ROC curve of the proposed fusion rule to the LRT's ROC curve. But as the performance gap between LRT and the proposed fusion rule without power allocation becomes less, the extent of improvement is thus less. The ROC curve of the proposed rule with power allocation is somewhat "upper bounded" by that of the optimal LRT. Although p_{peak} generates the peak improvement, it is not really the so-called optimal transmit power from the global perspective. Indeed, this work suggests using the average power greater than p_{peak} because the system performance can approach that of the optimal LRT, but note that while the transmit power is larger, the ROC curve of LRT still moves to left-upper corner. Since the performance of our work with power allocation approaches that of LRT, the resulting P_d is still increasing with the total power, although the extent of improvement in P_d is decreasing. After all, designers can choose a certain transmit power larger than p_{peak} to obtain the desired optimal ROC curve.

To find the reason GLRT performs poorly, we first notice that in LRT all the parameters in $p\left(\mathbf{r};\mathcal{H}_1\right)$ and $p\left(\mathbf{r};\mathcal{H}_0\right)$ are assumed known by FC, which can directly adopt these parameters in calculating the likelihood ratio. Most importantly, these values of parameters do not depend on whether the underlying situation is $\mathcal{H}_0$ or $\mathcal{H}_1$. However in GLRT, we only obtain the formula for estimating these parameters and these estimates are different in $\mathcal{H}_0$ and $\mathcal{H}_1$. Take the system in this work for example, although GLRT asks for $\Pr\left\{s_i = 1\right\}$ given that $\mathcal{H}_1$ is true, the derived ML estimate approximates π_1 only when $\mathcal{H}_1$ is actually happening. In other words, it is not possible to obtain the estimate of π_1 if it is $\mathcal{H}_0$ that is happening. In such case, the FC can only obtain $\Pr\left\{s_i = 1 \mid \mathcal{H}_0\right\}$, which is close to π_0. Consequently, although GLRT takes a similar form as LRT, it does not actually behave the same. Even in asymptotic case, i.e. through extensive computer simulations, there is still a gap between the ROC performance of LRT and GLRT.

5.3 Summary

Simulations indicate that the proposed simple fusion rule with power allocations outperforms GLRT rule significantly, especially when the number of sensors is small. With average transmit power moderately large, the performance of the proposed fusion rule with power allocations can even approach the ROC curve of the optimal clairvoyant LRT detector. When the transmit power is too low, power allocation does not improve much and there is still a gap from the optimal ROC curve of LRT. When the transmit power is larger than a threshold, power allocations can raise the ROC curve of the proposed fusion rule to that of LRT. However, the relative increase in P_d diminishes because the performance with power allocation is upper bounded by the ROC curve of LRT.

Chapter 6

Conclusions and Future Works

In the beginning, this report traces some important developments of distributed detection systems, and a popular system model called canonical distributed detection system is introduced, where sensors transmit their reports reliably and directly to FC through parallel channels. Two problems in canonical distributed detection systems are to be solved: the fusion rule design and the signal processing algorithm at local sensors, and our work focuses on the first one. In the WSN cases where channels cannot be assumed reliable anymore, following the important concept that fusion rule design and the channel effects should be considered jointly, this report surveys many works of channel-aware fusion rule design and finds out that most of these works do not address the problems where sensor performances are not known to FC. If the sensor performances are necessary in fusion rule design, estimations must be conducted at FC. Although GLRT can be applied to tackle these problems, there are rooms for improvement because firstly, the GLRT statistic is too complicated and secondly, GLRT does not guarantee the optimal performance.

In Chapter 3, the system model in our work is described in detail. The probability distributions of the received signal at FC are derived, which are indispensable in many fusion rule design, including GLRT. The formula of GLRT statistic is then derived for our system, and as is mentioned above, it is too complicated to analyze; moreover, the ML solution in the GLRT statistic is also complicated. This report then proposes an accurate approximation of the ML solution in high SNR, which is an affine function in the received signals, and can be reasonably interpreted as a modification of the voting scheme. However, even after replacing the simplified ML solution in the GLRT statistic, the GLRT statistic remains complicated.

Aiming at reducing the complexity of GLRT and at seeking for a better performance, this work then proposes a simpler fusion rule in Chapter 4. The greatest advantage of the proposed fusion rule is that it is simple and retains the affine properties of the proposed ML approximation, making it easy to analyze. The rule also uses the core information reflecting the state of the environments. A tight bound for the ROC curve is proposed and for high SNR, it further indicates that the channel effects come into the bound of global detection and false-alarm probability only through the summation of cross-over probability. A proof is also given under some reasonable assumption, claiming that minimizing the summation term guarantees better performance in ROC curve, and a power allocation strategy is then proposed to meet this goal. Simulations show that the proposed fusion rule outperforms GLRT, and after power allocation, the ROC curve of the proposed fusion rule can even approach the optimal LRT benchmark.

The key reason GLRT performs poorly is that in contrast to LRT where the environmental parameters in LRT statistic are known and identical in all hypotheses, GLRT replaces these fixed parameters with the ML estimates, whose values vary between different hypotheses. In sum, GLRT only borrows the formula but behaves differently. Another interesting feature of the proposed fusion rule is that channel effects come into the performance by the summation of cross-over probabilities. It means that the sensors with poor channels should transmit more power as compensation, which is different from the conventional communication system that to maximize capacity, water-filling is applied so that some parts of system with poor channels are allocated less power or even turned off. Note that the basic difference between these two systems is that in conventional communication systems, the transmitted signals from the transmitter are the same; given limited communication resources, a plausible way is to use these resources as efficiently as possible. In contrast, sensors in WSN make their decisions independently and each sensor should have “the same rights to speak.” The formulas of the performance reflect this reasoning because it is the summation of the cross-error probabilities that have the strongest impact. To summarize, in the scenario where there are unknowns, instead of adopting the commonly used GLRT rule, this work highlights the potential of designing a simpler fusion rule that outperforms GLRT, and then indeed proposes one fusion rule whose performance can even approach that of LRT, which is the optimal fusion rule for the cases where there are no unknowns.

Some issues are not addressed in this work. Recall that the sensor report s_i and the received signal r_i at FC are elements of $\{0,1\}$, namely they are hard decisions. Hard decisions in many cases lose the original information and thus perform worse. Accordingly, we expect that changing from hard to soft decisions results in better performance, but the analyses will be more involved. Another issue is that in WSN, it is probably demanding to obtain the instantaneous CSI. In our work, utilizing the instantaneous CSI and applying power allocation lead to an ROC curve close to that of optimal LRT. However without CSI and the resulting power allocation, there is still gap between the performance of the proposed fusion rule and that of optimal LRT. One future work is thus to design another fusion rule that uses less CSI or just the statistic of the channel, while still performing better than the proposed fusion rule in this work.

Bibliography

[1] J. Chamberland, "Wireless sensors in distributed detection applications," *IEEE signal processing magazine*, vol. 24, no. 3, pp. 16-25, May 2007.

[2] A. Swami, Q. Zhao, Y.-W. Hong, and L. Tong, *Wireless Sensor Networks: Signal Processing and Communications Perspectives*, John Wiley & Sons Ltd., 2007.

[3] P. K. Varshney, *Distributed Detection and Data Fusion*, New York, Springer 1997.

[4] Z. Qing, A. Swami, and T. Lang, "The interplay between signal processing and networking in sensor networks," *IEEE Signal Processing Magazine*, vol. 23, no. 4, pp. 84-93, July 2006.

[5] C. Biao, J. Ruixiang, T. Kasetkasem, and P. K. Varshney, "Channel aware decision fusion in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 52, no. 12, pp. 3454-3458, Dec. 2004.

[6] C. Qi, C. Biao, and P. K. Varshney, "Detection performance limits for distributed sensor networks in the presence of nonideal channels," *IEEE Transactions on Wireless Communication*, vol. 5, no. 11, pp. 3034-3038, Nov. 2006.

[7] S. K. Jayaweera, "Large system decentralized detection performance under communication constraints," *IEEE Communications Letters*, vol. 9, no. 9, pp. 769-771, Sept. 2005.

[8] V. R. Kanchumarthy, R. Viswanathan, and M. Madishetty, "Impact of channel errors on decentralized detection performance of wireless sensor networks: A study of binary modulations, Rayleigh-fading and nonfading channels, and fusion-combiners," *IEEE Transactions on Signal Processing*, vol. 56, no. 5, pp. 1761-1769, May 2008.

[9] N. Ruixin, C. Biao, and P. K. Varshney, "Fusion of decisions transmitted over Rayleigh fading channels in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 54, no. 3, pp. 1018-1027, March 2006.

[10] C. Biao, T. Lang, and P. K. Varshney, "Channel-aware distributed detection in wireless sensor networks," *IEEE Signal Processing Magazine*, vol. 23, no. 4, pp. 16-26, July 2006.

[11] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume 2: Detection Theory*: Prentice Hall, 1998.

[12] R. Radner, "Team decision problems," *Ann. Math. Statist.* , vol. 33, no. 3, pp. 857-881, Sept. 1962.

[13] E. L. Lehmann and Springer, *Testing Statistical Hypotheses*, 2nd ed., New York, Wiley, 1986.

[14] R. T. Robert and R. S. Nils, "Detection with distributed sensors," *19th IEEE Conference on Decision and Control including the Symposium on Adaptive Processes*, vol. 19, pp. 433-437, Dec. 1980.

[15] Z. B. Tang, K. R. Pattipati, and D. L. Kleinman, "An algorithm for determining the decision thresholds in a distributed detection problem," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 21, pp. 231-237, Nov. 1991.

[16] Z. Chair and P. K. Varshney, "Optimal data fusion in multiple sensor detection systems," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-22, pp. 98-101, Sept. 1986.

[17] R. V. Hogg and A. Craig, *Introduction to Mathematical Statistics*, 6th ed., Prentice Hall, 2005.

[18] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*, Prentice Hall, 1993.

[19] W. Neng and S. D. Blostein, "Minimum BER power allocation for MIMO spatial multiplexing systems," *IEEE International Conference on Communications*, vol. 4, pp. 2282-2286, May 2005

[20] Y. Zhiyong, K. M. Wong, and L. Zhi-Quan, "Optimal diagonal precoder for multiantenna communication systems," *IEEE Transactions on Signal Processing*, vol. 53, pp. 2089-2100, June 2005.

APPENDIX A:

PROOF OF THEOREM 5.2

To prove the theorem, first define

$$S_k^A \triangleq \sum_{i=k}^{N} A_i,\ S_k^B \triangleq \sum_{i=k}^{N} B_i,\ S_k^C \triangleq \sum_{i=k}^{N} C_i,\ S_k^D \triangleq \sum_{i=k}^{N} D_i \tag{A.1}$$

where A_i and B_i are defined in (4.12), and C_i and D_i are defined in (4.14). The two technical lemmas shown next facilitate the proof of the theorem.

Lemma A.1: Assume that $\pi_0 < 0.5 < \pi_1$. The following results hold.

(1) Both $S_k^A > 0$ and $S_k^C > 0$ are monotonically decreasing in *k*.

(2) $S_k^B \le 0$ and $S_k^D \ge 0$ for all *k*. $\square$

Proof of Lemma A.1:

Because $A_i > 0$ and $C_i > 0$, (1) follows immediately by definitions.

To prove (2). Let's write

$$B_k = \underbrace{\pi_1^{k-1}(1-2\pi_1)(1-\pi_1)^{N-k} C_{k-1}^{N-1}}_{:=Q_k} - \underbrace{\pi_1^{k}(1-2\pi_1)(1-\pi_1)^{N-k-1} C_{k}^{N-1}}_{:=R_k}. \tag{A.2}$$

Note that $Q_0 = R_N = 0$, due to the fact $C_{-1}^{N-1} = 0$ and $C_N^{N-1} = 0$. Since

$$\begin{aligned} \sum_{k=1}^{N} Q_k &= (1-2\pi_1)\sum_{k=1}^{N} \pi_1^{k-1}(1-\pi_1)^{N-k} C_{k-1}^{N-1} \\ &= (1-2\pi_1)(\pi_1 + 1 - \pi_1)^{N-1} = (1-2\pi_1) \end{aligned} \tag{A.3}$$

and

$$\begin{aligned} \sum_{k=0}^{N-1} R_k &= (1-2\pi_1)\sum_{k=0}^{N-1} \pi_1^{k}(1-\pi_1)^{N-k-1} C_{k}^{N-1} \\ &= (1-2\pi_1)(\pi_1 + 1 - \pi_1)^{N-1} = (1-2\pi_1) \end{aligned}, \tag{A.4}$$

it means

$$S_0^B = \sum_{i=0}^{N} B_i = \sum_{i=0}^{N} Q_i - \sum_{i=0}^{N} R_i = \sum_{i=1}^{N} Q_i - \sum_{i=0}^{N-1} R_i = 0\,. \tag{A.5}$$

Furthermore, as $\pi_1 > 0.5$,

$$S_N^B = \pi_1^{N-1}\left(1-2\pi_1\right) \to 0^- \quad \text{as } N \text{ gets large.} \tag{A.6}$$

Also observe that

$$\begin{aligned} C_{k-1}^{N-1}\left(1-\pi_1\right) - C_k^{N-1}\pi_1 &= \frac{(1-\pi_1)(N-1)!}{(k-1)!(N-k)!} - \frac{\pi_1(N-1)!}{(k)!(N-k-1)!} \\ &= (N-1)!\left[\frac{k-N\pi_1}{(k)!(N-k)!}\right] \end{aligned}. \tag{A.7}$$

From (A.7) and definition of B_k in (4.12), it follows immediately that

$$B_k > 0 \ \text{ for } \ k < N\pi_1\text{, and } \ B_k < 0 \ \text{ for } \ k > N\pi_1\,. \tag{A.8}$$

From (A.8), S_k^B decreases for $0 \le k < N\pi_1$, and increases for $N\pi_1 < k \le N$. This result, together with (A.5) and (A.6), imply $S_k^B \le 0$. Using the similar techniques, it can be verified that $S_k^D \ge 0$. □

Lemma A.2: The following results hold.

(1) If $\sum_{i=1}^{N} \varepsilon_i \le \frac{\pi_1}{2\pi_1 - 1}$, then $S_k^A + S_k^B\left(\sum_{i=1}^{N}\varepsilon_i\right)$ is monotonically decreasing.

(2) If $\sum_{i=1}^{N} \varepsilon_i \le \frac{\pi_0}{1-2\pi_0}$, then $S_k^C + S_k^D\left(\sum_{i=1}^{N}\varepsilon_i\right)$ is monotonically decreasing. □

Proof of Lemma A.2:

We only proved (1), since (2) can be similarly verified. To proceed, first focus on these k's such that $k > N\pi_1$. By assumption,

$$\begin{aligned} \sum_{i=1}^{N}\varepsilon_i \le \frac{\pi_1}{2\pi_1 - 1} &= \frac{\pi_1}{2\pi_1 - 1} \times \frac{C_k^N(1-\pi_1)}{C_k^N(1-\pi_1)} \\ &= \frac{C_k^N}{2\pi_1 - 1} \times \frac{\pi_1(1-\pi_1)}{C_k^N - \pi_1 C_k^N} \le \frac{C_k^N}{2\pi_1 - 1} \times \frac{\pi_1(1-\pi_1)}{C_{k-1}^{N-1} - \pi_1 C_k^N} \end{aligned}, \tag{A.9}$$

where the last equality follows due to $C_{k-1}^{N-1} < C_k^N$ and $k > N\pi_1$. From (A.9) we immediately have

$$\frac{C_{k-1}^{N-1} - \pi_1 C_k^N}{\pi_1(1-\pi_1)} \cdot \left(\sum_{i=1}^{N} \varepsilon_i\right) \le \frac{C_k^N}{2\pi_1 - 1}. \tag{A.10}$$

Since $C_k^N = C_{k-1}^{N-1} + C_k^{N-1}$, (A.10) can be rewritten as

$$\begin{aligned} &\frac{C_{k-1}^{N-1} - \pi_1\left(C_{k-1}^{N-1} + C_k^{N-1}\right)}{\pi_1(1-\pi_1)} \cdot \left(\sum_{i=1}^{N} \varepsilon_i\right) = \frac{(1-\pi_1)C_{k-1}^{N-1} - \pi_1 C_k^{N-1}}{\pi_1(1-\pi_1)} \cdot \left(\sum_{i=1}^{N} \varepsilon_i\right) \\ &= \left[\frac{C_{k-1}^{N-1}}{\pi_1} - \frac{C_k^{N-1}}{(1-\pi_1)}\right] \cdot \left(\sum_{i=1}^{N} \varepsilon_i\right) \le \frac{C_k^N}{2\pi_1 - 1} \end{aligned} \tag{A.11}$$

The last inequality in (A.11) equals to

$$C_k^N + \left(C_{k-1}^{N-1}\frac{1-2\pi_1}{\pi_1} - C_k^{N-1}\frac{1-2\pi_1}{1-\pi_1}\right)\left(\sum_{i=1}^{N} \varepsilon_i\right) \ge 0. \tag{A.12}$$

Multiply both sides of (A.12) by $(1-\pi_1)^{N-k}\pi_1^k$ and by rearrangement, we obtain

$$\left[\begin{aligned} &C_k^N (1-\pi_1)^{N-k}\pi_1^k \\ &+\pi_1^{k-1}(1-2\pi_1)(1-\pi_1)^{N-k-1}\left(C_{k-1}^{N-1}(1-\pi_1) - C_k^{N-1}\pi_1\right)\left(\sum_{i=1}^{N} \varepsilon_i\right) \end{aligned}\right] \ge 0. \tag{A.13}$$

By definition of the sequences A_k and B_k in (5.2), inequality (A.13) essentially asserts

$$A_k + B_k\left(\sum_{i=1}^{N} \varepsilon_i\right) \ge 0. \tag{A.14}$$

Since $S_k^A - S_{k+1}^A = A_k$ and $S_k^B - S_{k+1}^B = B_k$, (A.14) thus implies

$$S_k^A + S_k^B\left(\sum_{i=1}^{N} \varepsilon_i\right) \ge S_{k+1}^A + S_{k+1}^B\left(\sum_{i=1}^{N} \varepsilon_i\right), \tag{A.15}$$

which proves (1) for $k > N\pi_1$. If $k < N\pi_1$ then $\left[\frac{C_{k-1}^{N-1}}{\pi_1} - \frac{C_k^{N-1}}{1-\pi_1}\right] < 0$, and hence the last inequality in (A.11) still holds. By repeating the procedures as in (A.12)~(A.14), the relation (A.15) can also be obtained. The proof is thus completed.

□

Proof of Theorem 5.2:

Associated with total error rate E, let $(S_k^A,\ S_k^B,\ S_k^C,\ S_k^D)$ be accordingly defined as in (A.1). For a given threshold γ and with the given E, we can then express the performance bounds in (5.1) and (5.3) as

$$P_d^{(L)} = S_{k_l+1}^A + S_{k_l+1}^B E \ \text{ and } \ P_f^{(L)} = S_{k_l}^C + S_{k_l}^D E\,, \tag{A.16}$$

where k_l is some positive integer. Now if E is reduced to $E' < E$, it follows from part (2) of Lemma A.1 that

$$S_{k_l}^A + S_{k_l}^B E < S_{k_l}^A + S_{k_l}^B E' \ \text{ and } \ S_{k_l}^C + S_{k_l}^D E > S_{k_l}^C + S_{k_l}^D E'\,. \tag{A.17}$$

Since $\pi_0 < 0.5 < \pi_1$, we have $\dfrac{\pi_1}{2\pi_1 - 1} > 0$ and $\dfrac{\pi_0}{1 - 2\pi_0} > 0$. Under the assumptions of Lemma A.2, we have $S_k^C + S_k^D E'$ is monotonically decreasing. Let $k_l' < k_l$ be such that

$$k_l' = \min\{k \mid S_k^C + S_k^D E' \le S_{k_l}^C + S_{k_l}^D E\}\,. \tag{A.18}$$

For such k_l', it follows that $P_f^{(L)}(k_l') := S_{k_l'}^C + S_{k_l'}^D E' \le S_{k_l}^C + S_{k_l}^D E = P_f^{(L)} \le P_f$, and the corresponding detection probability lower bound shall satisfy

$$P_d^{(L)}(k_l') := S_{k_l'}^A + S_{k_l'}^B E' \overset{(a)}{>} S_{k_l}^A + S_{k_l}^B E' \overset{(b)}{>} S_{k_l}^A + S_{k_l}^B E = P_d^{(L)}\,, \tag{A.19}$$

where (a) holds since $S_k^A + S_k^B E$ is also monotonically decreasing (see Lemma A.2), and (b) follows from the first inequality in (A.17). Hence, as E is reduced to E', we have $P_d^{(L)}(k_l') > P_d^{(L)}$ whenever $P_f^{(L)}(k_l') \le P_f$. This implies that the detection probability lower bound $P_d^{(L)\prime}$ corresponding to P_f must exceed $P_d^{(L)}$.

$\square$