Distributed Inference over Multiple-Access Channels with Wireless Sensor

Networks

by

Mahesh Krishna Banavar

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved November 2010 by the
Graduate Supervisory Committee:

Cihan Tepedelenlioglu, Co-Chair
Andreas Spanias, Co-Chair
Antonia Papandreou-Suppappola
Tolga Duman
Junshan Zhang

ARIZONA STATE UNIVERSITY

December 2010

ABSTRACT

Distributed inference has applications in fields as varied as source localization, evaluation of network quality, and remote monitoring of wildlife habitats. In this dissertation, distributed inference algorithms over multiple-access channels are considered. The performance of these algorithms and the effects of wireless communication channels on the performance are studied.

In a first class of problems, distributed inference over fading Gaussian multiple-access channels with amplify-and-forward is considered. Sensors observe a phenomenon and transmit their observations using the amplify-and-forward scheme to a fusion center (FC). Distributed estimation is considered with a single antenna at the FC, where the performance is evaluated using the asymptotic variance of the estimator. The loss in performance due to varying assumptions on the limited amounts of channel information at the sensors is quantified. With multiple antennas at the FC, a distributed detection problem is also considered, where the error exponent is used to evaluate performance. It is shown that for zero-mean channels between the sensors and the FC when there is no channel information at the sensors, arbitrarily large gains in the error exponent can be obtained with sufficient increase in the number of antennas at the FC. In stark contrast, when there is channel information at the sensors, the gain in error exponent due to having multiple antennas at the FC is shown to be no more than a factor of $8/\pi$ for Rayleigh fading channels between the sensors and the FC, independent of the number of antennas at the FC, or correlation among noise samples across sensors.

In a second class of problems, sensor observations are transmitted to the FC using constant-modulus phase modulation over Gaussian multiple-access-channels. The phase modulation scheme allows for constant transmit power and estimation of moments other than the mean with a single transmission from the sensors. Estimators are

developed for the mean, variance and signal-to-noise ratio (SNR) of the sensor observations. The performance of these estimators is studied for different distributions of the observations. It is proved that the estimator of the mean is asymptotically efficient if and only if the distribution of the sensor observations is Gaussian.

To my parents,

who taught me by example to be

"... strong in will

To strive, to seek, to find, and not to yield."

ACKNOWLEDGEMENTS

Just as Frodo and the Fellowship and Lews Therin and the Hundred Companions, I too have had teachers, guides and friends who have helped me on my quest. I will use this space to convey my gratitude to them, for supporting me on this journey.

First of all, I would like to thank Dr. Cihan Tepedelenlioğlu and Dr. Andreas Spanias for being ideal advisors and mentors. Their help during my graduate studies has been instrumental in keeping me motivated throughout the process. Through their example, I have learnt what it is to be a researcher, a writer and a teacher. Their willingness to critique my work and attention to detail have made these past few years extremely interesting and enjoyable.

I am also grateful to Dr. Tolga Duman, Dr. Antonia Papandreou Suppappola and Dr. Junshan Zhang for agreeing to serve on my dissertation committee. Their feedback and advice during the process has been insightful and helpful. I cannot forget to thank my undergraduate mentor Dr. H. N. Shankar, for all the help and encouragement I have received from him over the years. I would also like to thank Dr. Joseph Palais for giving me an opportunity to teach undergraduate labs for most of my graduate studies, providing me with a most rewarding and enjoyable experience. Thanks also to Ms. Darleen Mandt for helping me with paperwork at all the different stages of my graduate studies.

I have received a lot of assistance from my friends and colleagues in the Signal Processing and Communications research groups. Special thanks to Dr. Venkatraman Atti, Dr. Adarsh Narasimhamurthy, Harish Krishnamoorthi, Robert Santucci, Lakshminarayan Ravichandran, N R. Karthikeyan and J. T. Jayaraman for all their help with reviewing my work, and providing valuable feedback. For helping me navigate the administrative minutiae, I would like to thank Ms. Cynthia Moayedpardazi, Ms. Donna Rosenlof, Ms. Jenna Marturano and Ms. Karen Anderson.

Most importantly, I would like to thank my parents for supporting me through-

out this endeavor. Thanks also to Tootle (my brother, Adithya), who, like his namesake, refuses to stay on the straight-and-narrow. At times, the diversions have helped me with my sanity.

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

1.1   Sensor Networks

Sensor networks provide a safe and low-cost sensing alternative to monitoring environmental conditions or physical phenomena where it may be otherwise difficult or impossible to do so. Typical examples involve identification of certain signal sources at a remote or unreachable area. These sensing tasks may be monitoring characteristics of hazardous materials, chemicals near a volcano, temperatures in a furnace, shifts in undersea tectonic plates, animal activities in dense forests, hormones in the blood or detecting toxins or explosives in the air, to name a few [1–12].

Sensor networks consist of infrastructure that allows the observation and collection of information of interest by using autonomous nodes deployed in space [1, 9, 10, 13]. These autonomous nodes contain sensing, processing and communication capabilities that allow us to observe and if required, to act on certain occurrences and events. Depending on the types of sensors available on the nodes, a single sensor network can be specialized to observe a single type of physical phenomenon, or a single network can be used to collate information from various physical conditions. Advances in hardware technology have allowed for the development of small, low power sensing nodes that have the capability to perform sophisticated sensing along with being outfitted with transceivers for wireless communications [2, 14–20].

The sensor nodes themselves can range from being extremely small (smartdust [14]) to large platforms collecting telemetry on tanks or aircrafts [9]. Depending on how sensor nodes are used and deployed, their capabilities can vary widely. Extremely small sensor nodes cannot have very sophisticated hardware or large computing capacities. Larger nodes that are supported by a more complex infrastructure can have more sophisticated sensors with larger memories, transceiver capabilities, higher computing power and different types of sensing abilities. These can be further supported by larger

1

computers with better computing capabilities, at the cost of higher power requirements and loss of mobility, in addition to being more difficult to deploy.

A major constraint that is faced when dealing with autonomous nodes is that these nodes are severely power limited [1, 9, 10]. In most cases, the nodes are supplied with power (charged batteries) and then deployed. In realistic situations, the batteries cannot be recharged or replaced once the sensors are deployed. A node that loses power will have to be discarded. Therefore, whenever nodes are autonomously deployed, the algorithms used on the nodes are designed for minimal power consumption and hence, to maximize the battery life of the node. It should be noted that while computing operations do consume power, most of the power is consumed by their transceivers [1, 9, 10, 20]. Hence, there is a need for efficient communication schemes to maximize transfer of information while consuming limited power.

Due to hardware and power limitations, when deployed individually, disposable sensors are capable of only simple computations and tend to perform poorly with some sensing tasks. However, when deployed in large numbers, sensors can be used to form intelligent networks and sensor data can be accumulated at a central location to obtain better results, using a process called data fusion. When connected to centralized computers, more complicated computations can be performed on the data gathered. The topology of sensor networks can be classified broadly into three types based on the presence or absence of a fusion center (FC) and the organization of the sensors.

In network literature, ad-hoc networks (Figure 1.1) refer to devices placed to form a network without a controlling base station. These devices discover each other and cooperate intelligently in order to function as a network.

When applied to sensors, ad-hoc sensor networks are constructed using the same principles [21–51]. Low-power sensors are placed in an observation field without a fusion center. Algorithms are developed for diverse applications such as data routing, collaborative inference and distributed signal processing, all subject to power
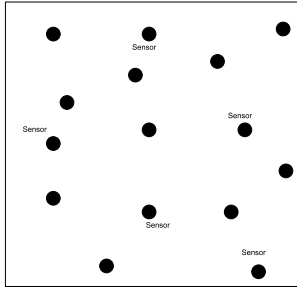
Figure 1.1: An example of an ad-hoc network with no fusion center.

constraints.

Data-transmission between sensors in an ad-hoc network is typically achieved using multi-hop routing, i.e., sensors in between the source and destination are used to route the data between the transmitter and the receiver. These sensors behave as relays in addition to their functions as sensors. When the messages are passed on by the relays, the data can be passed on digitally (for example, decode and forward) or using analog methods such as the amplify-and-forward technique.

With no fusion center, connectivity between all sensors may not be guaranteed. The transmit power radiated by each sensor must be such that the connection between neighboring nodes is guaranteed, without interfering with other communications. Conditions and degrees of connectivity are described in [21, 26, 29, 31, 34]. In these papers, the authors consider an ad-hoc network in a fixed area and compute the minimum power required or the minimum number of neighboring nodes to guarantee connectivity in the network. It is shown that the introduction of even a few base stations significantly improves the connectivity of a sparse network. The amount of data transfer that occurs between a given set of transmitters and receivers within a unit area of an ad-hoc network in unit time is defined as the capacity of a wireless network. Capacity of wireless ad-hoc networks for different conditions are analyzed in [27, 28, 30, 35, 40, 41].

Another configuration for sensors is called the hierarchical configuration (See Figure 1.2). In this setting, sensors, in addition to observing data, collect decisions

3
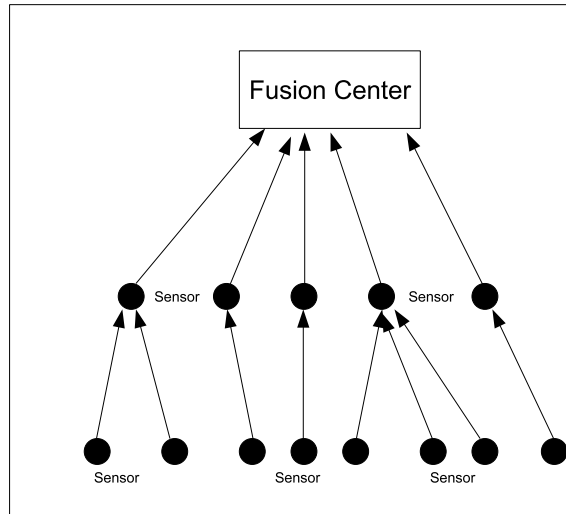
Figure 1.2: Hierarchical model - Data passes through multiple sensors.

from other sensors. The sensors use all this information to arrive at their own decisions and pass along their decision to subsequent sensors [5,52–54]. Typical applications are in sequential detection and sequential estimation.

In other sensor networks, sensors gather data and transmit them to a fusion center (in Figure 1.3), which processes the data. The transmissions over the channels between the sensors and the fusion center may be additive or orthogonal. When the transmissions are orthogonal, the transmissions from each sensor reaches the fusion center individually. The transmissions from the sensors do not interfere with each other. Therefore, the fusion center can choose to select each transmission independent of the other sensor transmissions. On the other hand, when the channels are additive (also called multiple-access channels [55, pp. 378]), the transmissions of the sensors add incoherently in noise before the fusion center has access to the data. The fusion center cannot select individual sensor transmissions. The bandwidth requirements of sensor networks with orthogonal channels scale linearly with the number of sensors, whereas, when the channels are multiple-access, transmissions are simultaneous and in the same frequency band, keeping the utilized bandwidth independent of the number of sensors

4

in the sensor network. For this multiple access channel model, it has been shown in [56] that a simple amplify-and-forward scheme for analog signals is asymptotically optimal over AWGN channels. It has also been shown in a distributed estimation context, that if the fading channels are zero-mean, having no channel state information at the sensors results in poor performance [57].

Transmissions from the sensors to the FC can be analog or digital. The digital method consists of quantizing the sensed data and then transmitting the data digitally over a rate-constrained channel [58–61]. In these cases, the required channel bandwidth is quantified by the number of bits being transmitted between the sensors and the fusion center. One such analog method consists of amplifying and then forwarding the sensed data to the FC, while respecting a power constraint [57]. The transmissions can be appropriately pulse-shaped and amplitude modulated to consume finite bandwidth. The major drawback of the amplify-and-forward scheme is that the transmit power depends on the sensing noise realizations and therefore may not be bounded. A solution to this problem is the use of phase modulation techniques with constant modulus transmissions from the sensors. Distributed estimation and detection algorithms with this transmit scheme are studied in [62–65].

Sensor networks that use this architecture are typically used for collaborative signal processing applications such as joint estimation, distributed detection, histogram estimation, etc. Due to the presence of multiple sensors, statistical methods perform very well since the number of observed data points can be very large. Histogram estimation using type based multiple access (TBMA) is introduced in [66]. Distributed detection is described in [64, 67–77]. Work and results in distributed estimation are in [57, 62, 63, 65, 78–84].

1.2 Applications of Sensor Networks

A few popular applications of sensor networks are described in this section.

Sensor networks can be used for traffic control [85], to warn drivers of areas
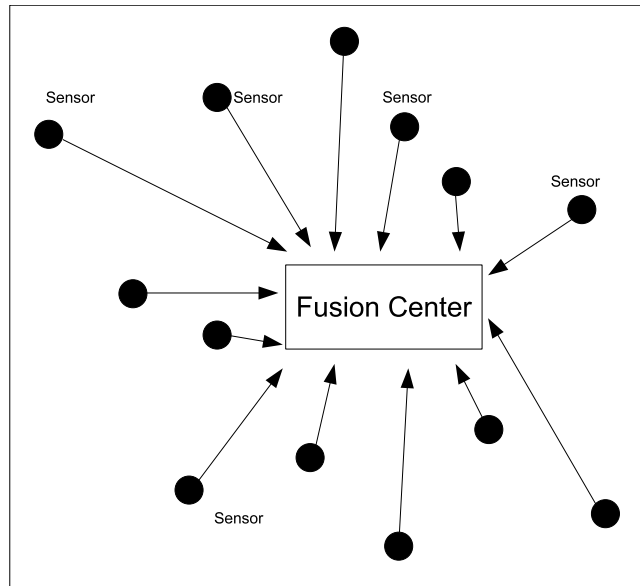
Figure 1.3: Sensor networks with a fusion center.

of congestion, to divert traffic to increase the efficiency of the roadways, and also to monitor roads for accidents and stoppages [86]. Sensor networks can be deployed to manage parking areas and to detect illegal use of parking areas. In addition, sensors can also be used to alert emergency services when required. These networks can be used to detect forest fires, toxic gas leaks in occupied mines, etc. Sensor networks can also be used for monitoring vital signs for medical purposes [11].

When deployed in an area to monitor sources such as climate changes, animal behavior, bird migration patterns, etc., the application is known as habitat monitoring. Such sensor deployment is used in sanctuaries and other protected areas. In [3,5] sensor networks are used for habitat monitoring in remote islands for data collection. Sensors developed for this applications need to be inconspicuous in order not to interfere with the natural behavior of wildlife. The Smart Dust system [14] is another example where sensors are made inconspicuous, in this case by reducing their size.

In a related application, the authors in [2] have developed system that is used for localization tasks. Such systems are also used in applications such as identifying the location of a sound source [6–8]. Once data is acquired, sensors can also be used

6

for more complicated tasks such as classification and tracking [4, 12]

In the next section, a literature review of distributed detection (Section 1.3) is presented, followed by a literature review of distributed estimation in Section 1.4, both for centralized sensor networks.

1.3 Distributed Detection

A detection problem that is solved with the help of multiple observations that are aggregated at a fusion center is called distributed detection. During hypothesis testing, when the a-priori probabilities of the hypotheses are not known, the Neyman-Pearson (NP) formulation is used and when the a-priori probabilities are known, the Bayesian risk approach is used [87]. The typical distributed detection problem involves local algorithms on the sensors and a central algorithm at the fusion center. Depending on the hypotheses, the likelihood ratio tests (LRTs) can be locally optimized at the sensors.

When a large number of sensors are deployed, asymptotic results indicate that the performance of the detector at the fusion center depends on the receiver operating characteristics (ROC) of the detectors at the individual sensors [88]. In addition, an LRT is performed at the fusion center as well. Since the performance depends on the LRTs at the sensors as well as at the fusion center, the algorithms have to be jointly optimized. This optimization can be done as a single one-shot solution, or iteratively, progressively improving the performance of the algorithms at the sensors and the fusion center [89]. For analysis with a finite number of sensors, the minimum number of sensors required to attain a certain performance is shown in [90].

Various metrics are used to characterize the performance of systems engaged in distributed detection. The most common techniques used are the probability of detection [91], the shape of the ROC curve [87], the J-divergence [92] and the error exponent [64, 77, 93, 94].

Distributed detection problems have been mainly studied assuming a single re-

ceive antenna at the FC. It is possible that introducing multiple antennas at a receiver may overcome the degradations caused by multi-path fading and noise. Inspired by conventional MIMO systems, a natural question is how much performance gain can be expected from adding multiple antennas at the FC in a distributed detection problem. However, this question cannot be directly answered by the studies in the MIMO literature. Adding multiple antennas to the FC for distributed detection problems is different when compared to the analysis of conventional MIMO systems for two reasons: (i) the presence of sensing noise (the parameter of interest is corrupted *before* transmission); and (ii) a large number of sensors enable *asymptotic* analysis.

In [95], a decision fusion problem with binary symmetric channels between the users and the FC is considered where the data are quantized at the sensors, transmitted over parallel channels, and processed after being received by three antennas. In [92], the authors consider multiple antennas at the FC. However, they consider a set of deterministic gains for the orthogonal channels, known at the sensors. They do not consider multiple-access channels, or characterize the performance benefits of adding antennas at the FC in the presence of fading. The system models in [72, 74, 96–99] are similar to adding multiple antennas at the FC, where the authors consider other forms of diversity, such as independent frequencies, CDMA codewords or several time intervals over fast-time-varying channels. When asymptotic techniques are used to investigate the benefits of adding multiple antennas at the fusion center, it can be shown that the gain on the error exponent by adding antennas to the FC when there is no channel state information (CSI) at the sensors grows linearly with the number of antennas. In stark contrast, when there is CSI at the sensors, only limited gains are possible by adding antennas at the FC [77, 100, 101]. This is unlike what is seen in traditional MIMO wireless communications, where adding antennas at the FC will result either in diversity gain or array gain, for asymptotically large SNRs.

1.4 Distributed Estimation

Distributed estimation deals with estimating the value of a random parameter by using a large number of observations that are provided by geographically separated sensors, whose observations are aggregated at a fusion center.

The authors in [59,60,78] consider quantized transmissions between the sensors and the fusion center. In these cases, the bandwidth is quantified by the number of bits being transmitted between the sensors and the fusion center. Furthermore, the system model used in [59, 60, 78] assumes that the channels between the sensors and the FC are orthogonal. In [102], the authors use the transmission model introduced in [59, 60, 78], and consider the effects of sending one-bit from each sensor through orthogonal binary symmetric channels (BSC). Similarly, in [103], the authors consider an imperfect channel modeled as a BSC. However, the channels are not fading in either of the cases. In the detection problems considered in [91] and [104], the authors use a transmission model where local detection decisions are transmitted over orthogonal fading channels to the fusion center.

Distributed estimation over multiple access channels with deterministic coefficients is considered in [81], where optimal sensor gains are derived for a finite number of sensors with perfect channel knowledge at the sensors. It is well-known (see, for example, [66]) that if the multiple-access channel between the sensors and the FC is fading with a zero-mean, and the sensors have no channel knowledge, the performance of the estimator is poor because the signals at the FC add incoherently over fading channels. A solution to this problem is to provide channel information to the sensors with feedback from the FC. In [84], orthogonal Rayleigh fading channels are considered between the sensors and FC. Performance is analyzed when perfect channel information is available at the sensors.

In [57, 105, 106], performance over multiple access fading channels is examined, and asymptotic results for variance are derived. Using the amplify-and-forward

scheme, the variance of the estimate is computed. The performance for different degrees of channel state information at the sensors (CSIS) for Rayleigh faded channels, when there is no CSIS, partial CSIS and full CSIS are investigated. It is shown that the feedback of only channel phase, even when quantized, leads to a surprisingly small performance loss. Also, the effect of errors in feedback on the performance are characterized. Furthermore, the effects of multiple antennas at the FC are characterized in [100, 107].

When constant modulus phase-modulation schemes are used at the sensors, information about the data is stored in the empirical characteristic function of the data. Using this, it is possible to estimate the location parameter, the scale parameter and the SNR of the data. SNR estimation finds applications in diverse areas in signal processing and communications, such as signal strength estimation for cognitive radio, in diversity combining and in bit-synchronization applications. SNR estimation for signals embedded in Gaussian noise are considered in [108, 109]. In the case of non-Gaussian noise, scale and location parameters are estimated simultaneously, and then combined to estimate the SNR, as reported in [110–114].

In a sensor network situation, sensors phase modulate the observations using a constant-modulus scheme and transmit these signals to a fusion center (FC) over a Gaussian multiple-access channel [55]. Due to the additive nature of the multiple-access channel, the signals transmitted from the sensors add and approximate the characteristic function of the signal and noise, as the number of sensors increases. At the FC, a noisy version of this empirical characteristic function is received in Gaussian noise, and the location and scale parameter are estimated from this value. A single transmission from each sensor to the FC is used for the estimation of the location parameter and the scale parameter. A single snapshot in time is sufficient for the estimation [62, 63, 65].

1.5 Organization of the Dissertation

The rest of this dissertation is organized as follows. A distributed estimation problem is discussed in Chapter 2. A single antenna is present at the FC. The asymptotic performance of the estimator is evaluated when the channels between the FC and the sensors are AWGN or fading, and when the sensors have full, partial and no channel information. In addition, speed of convergence is also characterized. With multiple antennas at the FC, a distributed detection problem is considered in Chapter 3. The channels between the sensors and the FC can be AWGN, Rayleigh fading or Ricean fading. Furthermore, differing amounts of channel information are considered at the FC. In each case, the performance is characterized in terms of the number of antennas at the FC.

Constant-modulus phase modulated transmissions from the sensors are considered in Chapter 4 and Chapter 5. In Chapter 4, the location parameter of a signal embedded in noise is estimated. The performance for different sensing noise distributions is considered and asymptotic efficiency is evaluated in each case. Both the location parameter and the scale parameter are estimated in Chapter 5. These estimates are then combined to form an estimate for the SNR of the signal in noise. Performance is evaluated for different cases of sensing noise.

Concluding remarks and future work are presented in Chapter 6 .

Chapter 2

DISTRIBUTED ESTIMATION OVER FADING MULTIPLE-ACCESS CHANNELS

2.1   Introduction

In this chapter, the effect of different channel fading models on the performance of the system is characterized. Partial channel feedback is considered and the asymptotic variance expressions for large number of sensors for different fading channel models and feedback scenarios are derived. Due to the asymptotic analysis used, the dependence of performance on the specific channel realizations can be removed, and the individual effects of feedback of channel phase only, imperfect channel phase feedback, and noisy feedback channels, on the performance can be decoupled. With correlated channels it is shown that for the M-dependent channel correlation model, the asymptotic results continue to hold. Also the speed of convergence is investigated as well as the effects of power, observation noise and channel correlation on the speed of convergence. The asymptotic analysis also allows comparison with the AWGN benchmark by revealing the factor by which the number of sensors should be increased to attain AWGN performance over fading channels with limited feedback.

2.2   System Model

Figure 2.1 shows our wireless sensor network setup with L sensors, which transmit observations to an estimator at the FC. The lth sensor amplifies its observation by a
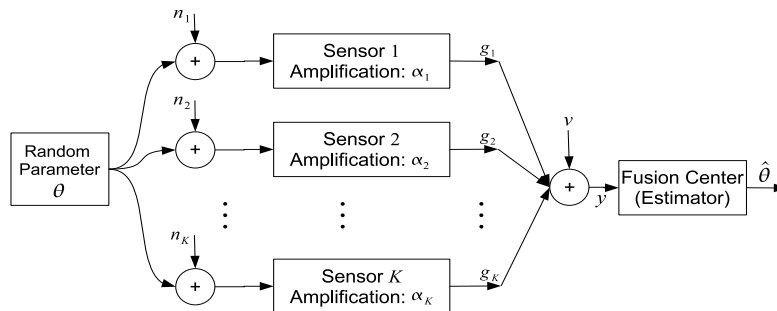
Figure 2.1: System Model.

factor α_l. The sensors transmit the amplified observations over L independent channels to the FC where the estimate $\hat{\theta}$ is produced. The flat fading channel h_l, between the lth sensor and the fusion center, is normalized to ensure $E[|h_l|^2] = 1, \forall l$, since when the sensors are placed close to each other and the far away from the FC, the distances between the sensors and the FC, in each case will be approximately the same, and the assumption of $E[|h_i|^2]$ is valid. The observation noise added at the lth sensor is given by $n_l \sim \mathscr{CN}(0, \sigma_n^2), \forall l$, and the channel noise with normalized variance is $v \sim \mathscr{CN}(0,1)$. The parameter being estimated, θ, has a variance of σ_θ^2. It is assumed that all these random variables are mutually independent of each other. The received signal at the FC is given by

$$y = \sum_{i=1}^{L} (\theta + n_i)\alpha_i h_i + v, \tag{2.1}$$

where the time index is dropped since the estimation is done in a single time snapshot.

Power Constraint

A total power constraint is imposed on the sensors. The signal transmitted by the lth sensor is $(\theta + n_l)\alpha_l$. The total transmitted power averaged over the parameter and noise distribution is given by

$$P_T = E\left[\sum_{l=1}^{L} |\alpha_l(\theta + n_l)|^2\right] = (\sigma_\theta^2 + \sigma_n^2)\sum_{l=1}^{L} |\alpha_l|^2. \tag{2.2}$$

In terms of the total power, P_T, the sensor gains, $\{\alpha_l\}$, are constrained by

$$P := \sum_{l=1}^{L} |\alpha_l|^2 = \frac{P_T}{\sigma_\theta^2 + \sigma_n^2}. \tag{2.3}$$

Estimation of θ

It is assumed that the FC has complete knowledge of the channels and sensor gains but only statistical information about the noise sources. Given the received signal in (2.1), the minimum variance linear unbiased estimate for θ is given as follows:

$$\hat{\theta} = \frac{y}{\sum_{i=1}^{L} \alpha_i h_i} = \theta + \frac{\sum_{i=1}^{L} n_i \alpha_i h_i}{\sum_{i=1}^{L} \alpha_i h_i} + \frac{v}{\sum_{i=1}^{L} \alpha_i h_i}. \tag{2.4}$$

13

Its variance, conditioned on the channel coefficients, is given by

$$\text{var}\left(\hat{\theta}\,\middle|\,\mathbf{h}\right) = E\left[|\theta - \hat{\theta}|^2\,\middle|\,\mathbf{h}\right] = \frac{\sigma_n^2 \sum_{i=1}^{L}\left(|\alpha_i|^2\,|h_i|^2\right) + 1}{\left|\sum_{i=1}^{L} \alpha_i h_i\right|^2}, \tag{2.5}$$

where $\mathbf{h} = [h_1 \quad h_2 \quad \ldots \quad h_L]^T$.

Performance over AWGN channels

First, the performance of the system over AWGN channels is examined, which will serve as a benchmark for the fading channel case. For AWGN channels, $h_l = 1, \forall l$. Due to symmetry, and to respect the power constraint, the gain on each sensor is set to $\alpha_l = \sqrt{P/L}, \forall l$. Substituting in (2.5), we obtain

$$\text{var}\left(\hat{\theta}\,\middle|\,\mathbf{h}\right) = \frac{\sigma_n^2 P + 1}{PL}. \tag{2.6}$$

Note that the variance in (2.6) goes to zero like $O(L^{-1})$ in the number of sensors. Scaling the variance in (2.6) with L, and defining

$$C_{AWGN} := \frac{\sigma_n^2 P + 1}{P} \tag{2.7}$$

as a benchmark against which to compare the asymptotic variances of other schemes, which will be addressed next.

2.3 Asymptotic Analysis of Performance

When the channels, h_l, are fading and random, the conditional variance in (2.5) is also random. When the variance in (2.5) goes to zero in such a way that

$$\lim_{L \to \infty} L \text{var}\left(\hat{\theta}\,\middle|\,\mathbf{h}\right) = C \tag{2.8}$$

in probability, where C is a deterministic constant, (2.8) is called the asymptotic variance. It has already been seen that for the AWGN case, C is given in (2.7). Different channel models and feedback schemes considered subsequently will have an associated value of asymptotic variance. Much of the remainder of the chapter will be devoted to calculating and interpreting the asymptotic variance under different assumptions on the channel and feedback schemes.

The following theorem will often prove useful towards evaluating the asymptotic variance over fading channels.

Theorem 2.3.1 *Let X_L and Y_L be two random sequences that converge in probability to deterministic constants x_0 and y_0, respectively. Let $f(x,y)$ be a scalar function of x and y. Then, $f(X_L,Y_L) \to f(x_0,y_0)$ in probability, if $f(\cdot,\cdot)$ is continuous at (x_0,y_0).*

Proof The proof follows directly from [115, Theorem C.1, pp. 422].

∎

2.4 Performance over Fading channels

Flat fading channels are considered between the sensors and the FC. It will be shown that whether the sensors have channel state information will greatly influence performance. The no channel state information at the sensors (CSIS) case will be used to motivate the need for some channel knowledge at the sensor side.

No Channel State Information at the Sensors

In the simplest case, the sensors have no channel information. Therefore, due to the i.i.d. channel statistics, the sensor gains are each set to $\alpha_l = \sqrt{P/L}, \forall l$, in order to satisfy the power constraint in (2.3). Substituting into (2.5), we get

$$\text{var}\left(\hat{\theta}\,\middle|\,\mathbf{h}\right) = \frac{1}{L}\frac{\sigma_n^2 \frac{P}{L}\sum_{l=1}^{L}|h_l|^2 + 1}{P\left|\frac{1}{L}\sum_{l=1}^{L}h_l\right|^2}. \tag{2.9}$$

Using the law of large numbers, substituting (2.9) in the definition of asymptotic variance in (2.8), and using Theorem 2.3.1 with $f(x,y) = (\sigma_n^2 Px + 1)/(P|y|^2)$, $X_L = L^{-1}\sum_{l=1}^{L}|h_l|^2$, $Y_L = L^{-1}\sum_{l=1}^{L}h_l$, evaluated at $x_0 = E[|h_l|^2] = 1$, $y_0 = E[h_l]$,

$$C_{NoCSIS} = \frac{\sigma_n^2 P + 1}{P\,|E\,[h_l]|^2}, \tag{2.10}$$

provided $E[h_l] \neq 0$.

For zero-mean channels, the signals received at the FC from different sensors combine incoherently, resulting in poor performance as seen in (2.10), which is unde-

15

fined for $E[h_l] = 0$, suggesting that $L\text{var}\left(\hat{\theta}|\mathbf{h}\right)$ does not converge for zero-mean channels.

In fact, this result, that for Rayleigh fading channels, the value of $L\text{var}\left(\hat{\theta}|\mathbf{h}\right)$ does not converge in probability for the no CSIS case, can be shown to be true for any deterministic or random set of α_i's independent of $\mathbf{h}$. To see this, consider

$$\text{var}\left(\hat{\theta}\Big|\mathbf{h},\alpha\right) = \frac{\sigma_n^2 \sum_{i=1}^{L}\left(|\alpha_i|^2 |h_i|^2\right) + 1}{\left|\sum_{i=1}^{L}\alpha_i h_i\right|^2} \geq \frac{1}{\left|\sum_{i=1}^{L}\alpha_i h_i\right|^2}, \qquad (2.11)$$

because $\sigma_n^2 \geq 0$ and $\sum_{i=1}^{L}\left(|\alpha_i|^2 |h_i|^2\right) \geq 0$. For any set of channel gains that satisfy the power constraint with equality $\sum_{l=1}^{L}|\alpha_l|^2 = P$, the denominator on the right hand side of (2.11) is an exponential random variable with mean P. Since the expected value of the inverse of an exponential random variable does not exist, $E\left[\text{var}\left(\hat{\theta}\Big|\mathbf{h}\right)\right]$ with respect to the channel distribution does not exist.

The well-known Ricean channel model is now considered as an example of the nonzero-mean scenario.

Example – Ricean channels

A Ricean channel can be represented as [116]

$$h_l = \sqrt{\frac{1}{K+1}}h_l^{diff} + \sqrt{\frac{K}{K+1}}e^{j\omega}, \qquad (2.12)$$

where $h_l^{diff} \sim \mathscr{CN}(0,1)$ is the diffuse component, ω is the phase of the specular component, and $K > 0$ is the ratio of the specular power to the power of the diffuse component. Using (2.12), the value of C_{NoCSIS} in (2.10) is

$$C_{NoCSIS} = \frac{\sigma_n^2 P + 1}{P}\frac{K+1}{K}. \qquad (2.13)$$

Clearly (2.13) is worse than C_{AWGN} in (2.7) by a factor of $(K+1)/K > 1$. As K increases, the channels have less fading, and C_{NoCSIS} approaches C_{AWGN}. On the other extreme, as $K \to 0$, only the diffuse component remains with Rayleigh amplitude, resulting in the value of C_{NoCSIS} growing without bound. Since the variances under

16

both scenarios are $O(L^{-1})$, the ratio of asymptotic variances, $(K+1)/K$, in (2.13), can be interpreted as the factor by which the number of sensors should be increased by a system with no CSIS over fading channels, to get AWGN performance. Throughout the manuscript, the ratio of asymptotic variances of any two schemes can be interpreted similarly.

Perfect Channel State Information at the Sensors

In rich scattering environments, the non-zero-mean assumption on the channel does not always hold. When the channel is zero-mean, an incoherent sum of faded signals are received at the FC, leading to unacceptable performance as seen in (2.10). One solution to the zero mean channel problem is to provide channel information to the sensors, which can be obtained through feedback, or by exploiting reciprocity on some systems [117]. The gains α_l used at the sensors may then depend on the channel, in order to make the effective channels $\{\alpha_l h_l\}$ non-zero mean.

To obtain a benchmark for our subsequent results with partial CSI, the optimal set of gains, derived in [81], are used, which require full-CSIS. Consider the sensor gains that minimize the variance of $\hat{\theta}$ in (2.5) subject to the power constraint on the sensors in (2.3):

$$
\underset{\{\alpha_i\}}{\text{minimize}} \frac{\sigma_n^2 \sum_{l=1}^{L} |\alpha_l|^2 |h_l|^2 + 1}{\left|\sum_{l=1}^{L} \alpha_l h_l\right|^2},
$$

$$
\text{subject to} \sum_{l=1}^{L} |\alpha_l|^2 \leq P. \tag{2.14}
$$

Let $\psi_l := \angle h_l$ be the phase of the lth sensor's channel. Applying the Cauchy-Schwarz inequality to the denominator of the objective function in (2.14), it is clear that the phase of the sensor gain that provides the best performance is given by $\angle \alpha_l = -\psi_l$, which means that only $\{|\alpha_l|\}$ need to be optimized.

Substituting $\angle \alpha_l = -\psi_l$, swapping the objective function and the constraint, and introducing a new variable, $s = \sum_{l=1}^{L} (|\alpha_l||h_l|)$, (2.14) becomes a (convex) second

17

order cone programming problem [118] in the variables $\{|\alpha_l|\}$ and s [81]:

$$\underset{\{|\alpha_i|\},s}{\text{minimize}} \quad \sum_{k=1}^{L} |\alpha_k|^2,$$

$$\text{subject to} \quad \sigma_n^2 \sum_{l=1}^{L} |\alpha_l|^2 |h_l|^2 + 1 \leq v_t s^2,$$

$$\sum_{l=1}^{L} (|\alpha_l||h_l|) - s = 0, \tag{2.15}$$

where v_t is a constant value below which the variance of the estimate must be constrained.

Using the Karush-Kuhn-Tucker conditions [118], the solution is given as:

$$\alpha_l = \sqrt{\frac{P}{\sum_{i=1}^{L} \left(\frac{|h_i|}{1 + P|h_i|^2 \sigma_n^2} \right)^2}} \left(\frac{|h_l|}{1 + P|h_l|^2 \sigma_n^2} \right) e^{-j\psi_l}, \tag{2.16}$$

and the optimum conditional variance is given by

$$\text{var}\left(\hat{\theta}\Big|\mathbf{h}\right) = \left(\sum_{l=1}^{L} \frac{1}{\sigma_n^2 + \frac{1}{P|h_l|^2}} \right)^{-1}. \tag{2.17}$$

Note that α_l in (2.16) can be computed at the FC and fed back to the sensors. The conditional variance in (2.17) is an achievable best-case benchmark for the conditional variance over fading channels.

The asymptotic variance for this optimized case can be calculated using (2.8), (2.17) and Theorem 2.3.1, with $f(x,y) = 1/x$, $X_L = L^{-1} \sum_{l=1}^{L} (\sigma_n^2 + (P|h_l|^2)^{-1})^{-1}$, evaluated at $x_0 = E[(\sigma_n^2 + (P|h_l|^2)^{-1})^{-1}]$, to obtain

$$C_{OPT} = \left(E\left[\frac{1}{\sigma_n^2 + \frac{1}{P|h_l|^2}} \right] \right)^{-1}. \tag{2.18}$$

For Rayleigh fading channels, it is straightforward to calculate (2.18), which can be expressed in closed form as

$$C_{OPT} = \frac{2\sigma_n^4 P}{2\sigma_n^2 P - \exp\left(\frac{1}{2\sigma_n^2 P} \right) E_1\left(\frac{1}{2\sigma_n^2 P} \right)}, \tag{2.19}$$

where $E_1(\cdot)$ is an exponential integral function [119, pp. 228].

Now to compare C_{OPT} in (2.18) with C_{AWGN}. Since C_{OPT} is obtained over fading channels, one would conjecture that $C_{AWGN} \leq C_{OPT}$. This is indeed the case by noting that $(\sigma_n^2 + 1/Px)^{-1}$ is a concave function of x, and using Jensen's inequality. This establishes the expected result that the performance over fading channels cannot be better than that over AWGN channels.

By examining (2.7) and (2.18), it is also clear that for small σ_n^2, both C_{AWGN} and C_{OPT} approach P^{-1}. On the other hand, the asymptotic variance expressions for large P yield

$$\lim_{P \to \infty} C_{AWGN} = \lim_{P \to \infty} C_{OPT} = \sigma_n^2. \tag{2.20}$$

In conclusion, C_{OPT} can be obtained in closed form for Rayleigh fading channels, and it is always no less than C_{AWGN}. They coincide when σ_n^2 is small, or when P is large. Therefore, for large L, when the sensing noise is small, or when the transmit power is large, it is possible to obtain near-AWGN performance over fading channels.

Phase-Only (PO) CSIS

For the scheme described in Section 2.4, calculation of α_l requires computing (2.16) at the FC for each sensor. Also, the amplification factors, α_l, have a large dynamic range that depend on the channel coefficients, which is undesirable due to the need for having inexpensive power amplifiers. This motivates the consideration of a constant gain at each sensor, so that each sensor compensates only for the phase of its channel. The sensors need only phase information in this case, implying less feedback, and provide a constant magnitude gain, requiring low-cost amplifiers. Also the loss in performance, with this choice of equal magnitudes for α_l when compared to the full CSIS case, is sufficiently small, which is another reason why the equal $|\alpha_l|$ case is important.

Theorem 2.4.1 *When the sensors have only knowledge of channel phase, the asymp-*

totic variance is given by

$$C_{PO} = \frac{\sigma_n^2 P + 1}{P} \frac{1}{(E[|h_l|])^2}.$$ (2.21)

Proof As in Section 2.4, $\angle \alpha_l = -\psi_l$ is the choice of phase at each sensor that minimizes the variance. In order to use phase-only feedback, and to respect the power constraint, $|\alpha_l|^2 = P/L, \forall l$. Therefore

$$\alpha_l = \sqrt{\frac{P}{L}} e^{-j\psi_l}.$$ (2.22)

Substituting in (2.5),

$$\text{var}\left(\hat{\theta} \big| \mathbf{h}\right) = \frac{1}{L} \frac{\sigma_n^2 \frac{P}{L} \sum_{l=1}^{L} |h_l|^2 + 1}{P \left[\frac{1}{L} \sum_{l=1}^{L} |h_l|\right]^2}.$$ (2.23)

The asymptotic variance for this phase-only (PO) case is now computed. From (2.8), (2.23) and using Theorem 2.3.1 with $f(x,y) = (\sigma_n^2 P x + 1)/(P y^2)$, $X_L = L^{-1} \sum_{l=1}^{L} |h_l|^2$, $Y_L = L^{-1} \sum_{l=1}^{L} |h_l|$, evaluated at $x_0 = E[|h_l|^2] = 1$, and $y_0 = E[|h_l|]$, (2.21) yields the proof.

∎

Notice that the first term in the right hand side of (2.21) is C_{AWGN} given in (2.7). The second term satisfies $(E[|h_l|])^{-2} \geq 1$ due to Cauchy Schwarz inequality and the fact that $E[|h_l|^2] = 1$. This implies that $C_{PO} \geq C_{AWGN}$, as expected. Indeed, $C_{PO} \geq C_{OPT} \geq C_{AWGN}$, since the optimized choice of the gains will outperform the phase-only case. However, (2.21) provides the further insight that C_{PO} is a constant multiple of C_{AWGN} for any value of P or σ_n^2. In (2.20), it is shown that as P increases, C_{OPT} and C_{AWGN} converge to σ_n^2. As C_{PO} is compared with C_{OPT}, as P increases, the ratio of asymptotic variances gets arbitrarily close to a factor of $4/\pi$. This is also seen in the simulations in Figure 2.4.

The phase-only system over fading channels will have approximately the same performance as a system over AWGN channels if its number of sensors is larger by a

factor given by $(E[|h_l|])^{-2} \geq 1$. Rayleigh, Ricean and Nakagami fading examples are now considered to see what this constant factor is for these cases.

Example 1 – Rayleigh Fading

For Rayleigh fading channels, $h_l \sim \mathscr{CN}(0,1)$, and $|h_l|$ is Rayleigh distributed. In this case, (2.21) yields

$$C_{PO} = \frac{\sigma_n^2 P + 1}{P} \frac{4}{\pi}. \tag{2.24}$$

Since $4/\pi < 1.3$, over Rayleigh fading channels, one can obtain AWGN performance asymptotically if the number of sensors for the phase-only scenario is about 30% larger than that of the AWGN scenario, because the variance is $O(L^{-1})$. The value of $(E[|h_l|])^{-2}$ is less over Ricean channels and depends on the Ricean factor as seen below.

Example 2 – Ricean Fading

Substituting the first moment of a Ricean random variable [116, Equation (4)] into (2.21),

$$C_{PO} = \frac{\sigma_n^2 P + 1}{P} \frac{1}{(K+1)\Gamma^2(3/2)e^{-2K} {}_1F_1^2(3/2;1;K)}, \tag{2.25}$$

where ${}_1F_1(\cdot;\cdot;\cdot)$ is the confluent hypergeometric function [119, pp. 504]. As expected, the value of $(E[|h_l|])^{-2}$ for the Ricean case lies between $4/\pi$ and 1 when K varies between 0 and ∞, respectively. Therefore, (2.25) has equations (2.7) and (2.24) as special cases corresponding to $K \to \infty$, and $K = 0$, respectively.

Example 3 – Nakagami Fading

For fading channels with Nakagami distributed envelopes, with parameter $m \in [1/2, \infty)$,

$$E[|h_l|] = \frac{\Gamma\left(m + \frac{1}{2}\right)}{\Gamma(m)} \frac{1}{\sqrt{m}},$$

where $\Gamma(\cdot)$ is the gamma function [119, pp. 225].

Substituting in (2.21),

$$C_{PO} = \frac{\sigma_n^2 P + 1}{P} \frac{m\Gamma^2(m)}{\Gamma^2\left(m + \frac{1}{2}\right)}. \tag{2.26}$$

21

Similar to the Ricean case, as $m \to \infty$ (AWGN channels), the value of C_{PO} for Nakagami channels converges to C_{AWGN}; when $m = 1$, the value of C_{PO} for Nakagami channels converges to C_{PO} for Rayleigh fading channels in (2.24). When $m = 1/2$, the Nakagami distribution is a one-sided Gaussian and represents a more severe fading scenario than the Rayleigh case. In this case, (2.26) becomes $(\pi/2)C_{AWGN}$, which is worse than the Rayleigh case, $(4/\pi)C_{AWGN}$, since $\pi/2 > 4/\pi$.

<div align="center">Continuous Channel Feedback with Phase Error</div>

When the channel phase feedback is not correct, the performance of the estimator will deteriorate. Let $\hat{\psi}_l$ denote the estimated phase being fed back to sensor l. The error in feedback is given by $\tilde{\psi}_l = \hat{\psi}_l - \psi_l$. A common model for phase error is the von Mises random variable [120], whose pdf can be described by

$$f_{\tilde{\Psi}}(\tilde{\psi}) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos \tilde{\psi}}, \quad \tilde{\psi} \in [-\pi, \pi), \tag{2.27}$$

where κ denotes the inverse variance of the random variable and $I_n(\cdot)$ is the nth order modified Bessel function of the first kind. When $\kappa = 0$, the distribution collapses to a uniform random variable, and approximates a Gaussian random variable with variance $1/\kappa$ for large κ. The parameter κ quantifies the accuracy of the feedback phase. When there is no error, $\kappa \to \infty$, and for large error, $\kappa \to 0$. The phase is known at the fusion center without any error, and this correct phase is used at the estimator. The value of the phase is corrupted on the feedback channel. Since the magnitude of the gain at each sensor is fixed, only the phase estimated for each channel is fed to the corresponding sensors.

Theorem 2.4.2 *When the sensors have noisy estimates of the channel phase, with the error in feedback whose p.d.f. is given as in (2.27), the asymptotic variance is given by*

$$C_{PO}(\kappa) = \left[\frac{I_1(\kappa)}{I_0(\kappa)} - \frac{\mathbf{L}_1(\kappa)}{I_0(\kappa)} - \frac{2}{\pi I_0(\kappa)} \right]^{-2} C_{PO}, \tag{2.28}$$

where $\mathbf{L}_m(\cdot)$ is the mth order modified Struve function [119, pp. 498].

Proof With the phase feedback of $\hat{\psi}_l$, the sensor gains are set to

$$\alpha_l = \sqrt{\frac{P}{L}} e^{-j\hat{\psi}_l}, \quad l = 1, \ldots, L, \tag{2.29}$$

and the conditional variance is given by

$$\mathrm{var}\left(\hat{\theta}\,\Big|\,\mathbf{h}\right) = \frac{1}{L} \frac{\sigma_n^2 P \frac{1}{L} \sum_{l=1}^{L} |h_l|^2 + 1}{P \left|\frac{1}{L} \sum_{l=1}^{L} |h_l| e^{-j(\hat{\psi}_l - \psi_l)}\right|^2}. \tag{2.30}$$

Using (2.8) and (2.30) and Theorem 2.3.1 with $f(x, y) = (\sigma_n^2 P x + 1)/(P|y|^2)$, $X_L = L^{-1} \sum_{l=1}^{L} |h_l|^2$, and $Y_L = L^{-1} \sum_{l=1}^{L} |h_l| e^{-j(\hat{\psi}_l - \psi_l)}$, evaluated at $x_0 = E[|h_l|^2] = 1$ and $y_0 = E[|h_l| e^{-j(\hat{\psi}_l - \psi_l)}]$, and recalling that $|h_l|$ is statistically independent of ψ_l and therefore also of $e^{-j(\hat{\psi}_l - \psi_l)}$, for Rayleigh fading channels, the following asymptotic variance in the presence of continuous feedback error is obtained:

$$C_{PO}(\kappa) = C_{PO} \left| E\left(e^{-j\tilde{\psi}_l}\right)\right|^{-2}. \tag{2.31}$$

To calculate the expectation in (2.31), the distribution of $\tilde{\psi}_l$ in (2.27) and [121, §3.365, pp. 345] is used to obtain $E(e^{-j\tilde{\psi}_l})$ and substitute into (2.31) to get the desired result in (2.28).

∎

Quantized Channel Phase Feedback

Since the channel phase cannot be fed back with infinite precision, it is natural to investigate the effects of quantization. A constant amplitude transmission is assumed for each sensor, and the channel phases are uniformly quantized for feedback to the sensors. This is optimal for the Rayleigh fading channel model, which has uniformly distributed phase. For the Ricean and Nakagami models, the phase of the channels may be non-uniform. The Rayleigh case are selected here to get a simple framework within which to evaluate the effect of feedback quantization.

For Rayleigh fading channels, $\psi_l := \angle h_l$ are uniformly distributed over $[0, 2\pi)$. With Q bits of quantization, $[0, 2\pi)$ is divided equally into 2^Q sectors, each constituting

23
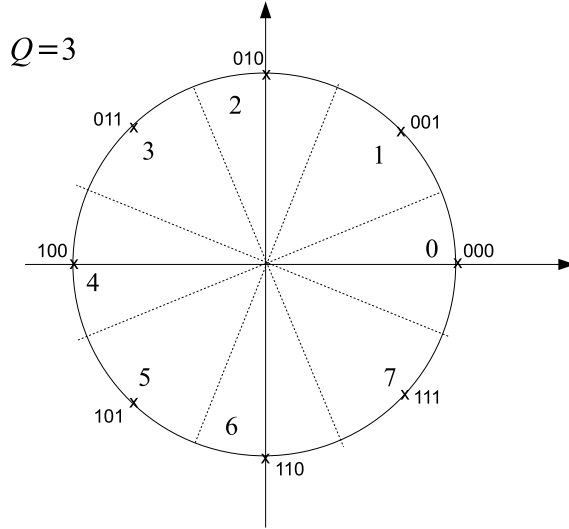
Figure 2.2: Phase to bits mapping for quantized feedback.

of $2\pi/2^Q$ radians. The center of each sector is chosen as $\{\exp(j2\pi k/2^Q)\}_{k=0}^{2^Q-1}$ so that the quantization points yield error magnitudes of at most $\pi/2^Q$ radians. To send the appropriate phase feedback, each sector is mapped to a unique Q-bit sequence, as shown in Figure 2.2, where as an example, $Q = 3$ is assumed.

Let

$$\alpha_l = \sqrt{\frac{P}{L}}e^{-jf_Q(\psi_l)}, \tag{2.32}$$

where $f_Q(\psi_l)$ is the quantized phase given by the element $x \in \left\{\frac{2\pi k}{2^Q}\right\}_{k=0}^{2^Q-1}$ which minimizes $(|\psi_l - x|)_{\mathrm{mod}2\pi}$.

Following from (2.29) - (2.31), the expectation in (2.31) is computed using the facts that ψ_l is uniformly distributed over $[0, 2\pi)$ and $\phi = (f_Q(\psi_l) - \psi_l)$ is uniformly distributed on $[-\pi/2^Q, \pi/2^Q)$. From this:

$$E\left[e^{-j\phi}\right] = \frac{2^{Q-1}}{\pi}\int_{-\frac{\pi}{2^Q}}^{\frac{\pi}{2^Q}}e^{-j\phi}d\phi = \frac{2^Q}{\pi}\sin\left(\frac{\pi}{2^Q}\right). \tag{2.33}$$

It follows that the asymptotic variance in the presence of Q-bit quantization is

24

Q	1	2	3	4	5
$\frac{C_{PO}[Q]}{C_{PO}}$	2.4674	1.2337	1.0530	1.0130	1.0032

Table 2.1: Degree of deterioration due to quantization.

given by

$$C_{PO}[Q] = \left[\text{sinc}(2^{-Q})\right]^{-2} C_{PO}, \tag{2.34}$$

where C_{PO} is as in (2.24), and $\text{sinc}(x) := \sin(\pi x)/(\pi x)$.

The loss in performance caused due to quantization is $[\text{sinc}(2^{-Q})]^{-2}$, which takes the value of 2.4674 for $Q = 1$ and goes to 1 as $Q \to \infty$. Table 2.1 contains the deterioration in asymptotic performance due to quantization ($C_{PO}[Q]/C_{PO}$) for different values of Q. Notice that by using three bits of quantization, there is an increase in variance of only about 5%. Therefore, a system with perfect phase feedback will perform similarly to a system with three-bit quantized phase feedback, if the latter system has 5% more sensors.

<center>Error in Quantized Feedback</center>

Suppose that each bit that is fed back could be received in error equally likely with probability p. Since p is often much less than one, the single-bit error events will dominate the performance. The error in phase that is committed with each single bit error is evaluated. This clearly depends on the bit assignment. To get an analytically tractable setting, a natural bit assignment is assigned to each sector as in Figure 2.2. Note that in this case, a single bit error will cause a phase error of $\pm\frac{2\pi}{2^Q}(2^k)$ for $k = 0,\ldots,2^{Q-1}$, with the minus sign used if the error is $1 \to 0$ and the plus sign is used otherwise.

In order to evaluate the performance of the system, the expectation in (2.31), the only factor affected by errors in feedback, is recalculated. To calculate this expected value in the presence of errors, the event is conditioned so that it contains all bit vectors with i errors. Since the single error case is the main interest, the expectation is

<center>25</center>

expressed as:

$$E\left[e^{-j\phi}\right] = \sum_{i=0}^{Q} A_i(p) = A_0(p) + A_1(p) + \sum_{i=2}^{Q} A_i(p), \tag{2.35}$$

where for notational convenience, $A_i(p) := E\left[e^{-j\phi}\big|i\text{errors}\right] Pr\left[i\text{errors}\right]$ for $i = 0, \ldots, Q$.

Evaluating the $i = 0$ term:

$$A_0(p) = (1-p)^Q \text{sinc}(2^{-Q}). \tag{2.36}$$

To evaluate the single-error case, recall that $\phi = \pm\frac{2\pi}{2^Q}(2^k)$, where $k \in \{0, \ldots, Q-1\}$ denotes the bit that is toggled and the sign is determined by the value of the bit. Therefore,

$$A_1(p) = \frac{(1-p)^{Q-1}p}{2}\left[\sum_{k=0}^{Q-1} e^{-j\frac{2\pi}{2^Q}(2^k)} + \sum_{k=0}^{Q-1} e^{-j\frac{2\pi}{2^Q}(-2^k)}\right]$$

$$= (1-p)^{Q-1}p \sum_{k=0}^{Q-1} \cos\left(\frac{2\pi}{2^Q}2^k\right). \tag{2.37}$$

Noting that $A_i(p)$ for $i \geq 2$ are $o(p)$ as $p \to 0$ [1],

$$E\left[e^{-j\phi}\right] = (1-p)^Q \text{sinc}(2^{-Q})$$

$$+ (1-p)^{Q-1}p \sum_{k=0}^{Q-1} \cos\left(\frac{2\pi}{2^Q}2^k\right)$$

$$+ o(p). \tag{2.38}$$

The asymptotic variance in the presence of Q-bit quantization and feedback errors, $C_{PO}[Q,p)$, is obtained by finding the ratio $C_{PO}/|E\left[e^{-j\phi}\right]|^2$ using (2.38):

$$C_{PO}[Q,p) = \frac{C_{PO}}{|A_0(p) + A_1(p) + o(p)|^2}. \tag{2.39}$$

It is straightforward from (2.39) that $C_{PO}[Q,0) = C_{PO}[Q]$. Table 2.2 shows the effect of errors on the feedback channel. Even with only five bits ($Q = 5$) and $p = 10^{-3}$, the deterioration from C_{PO} is about only about 1%, compared to perfect phase feedback. When the value of p reduces, predictably, the loss in performance also reduces.

---

[1] A function $A(p) = o(p)$ as $p \to 0$ means $A(p)/p \to 0$ as $p \to 0$

26

Q	$p = 10^{-1}$	$p = 10^{-2}$	$p = 10^{-3}$	$p = 10^{-4}$
1	4.4705	2.5993	2.4801	2.4687
2	2.4471	1.3136	1.2414	1.2345
3	2.1207	1.1253	1.0600	1.0537
4	2.0532	1.0838	1.0198	1.0136
5	2.0416	1.0740	1.0100	1.0039

Table 2.2: $C_{PO}[Q,p]/C_{PO}$ for different values of p and Q.

The performance in (2.39) is the performance of the estimator over Rayleigh fading channels when the sensors are provided with quantized phase-only feedback with errors on the feedback channel. As mentioned earlier, when $p = 0$, the performance reduces to the quantized, phase only result, $C_{PO}[Q]$ from (2.34). When $p = 0$ and $Q \to \infty$, $C_{PO}[Q,p)$ reduces to C_{PO} from (2.21), which is the performance with phase-only feedback with Rayleigh fading channels. This in turn is a factor $(4/\pi)$ worse than the performance over AWGN channels. Remarkably, this chain of relationships between the asymptotic variances can be decoupled and seen individually in our framework.

2.5   Effects of Fading Correlation

For the value of $\lim_{L \to \infty} L\mathrm{var}\left(\hat{\theta}\big|\mathbf{h}\right)$ to converge to a constant, using Theorem 2.3.1, the sample means, X_L and Y_L, need to separately converge in probability. It is well known that the weak law of large numbers holds for a wide range of correlation models [122]. For simplicity, it is assumed that the channels are M-dependent, i.e., if $s - r > M$, then the two vectors $\begin{bmatrix} h_1 & h_2 & \ldots & h_r \end{bmatrix}$ and $\begin{bmatrix} h_s & h_{s+1} & \ldots & h_L \end{bmatrix}$ are independent. Under this M-dependent model, the following theorem can be stated.

Theorem 2.5.1 *The asymptotic variance results in (2.7), (2.13), (2.18), (2.19), (2.21) - (2.26), (2.31), (2.28), (2.34) and (2.39) hold when the channels, $\{h_l\}$, are M-dependent.*

Proof Using the terminology, $f(\cdot,\cdot), X_L, Y_L, x_0$ and y_0 introduced in Theorem 2.3.1. There exist choices for $f(\cdot,\cdot), X_L, Y_L, x_0$ and y_0 for each of the cases considered in

Sections 2.2 and 2.4. For example, in the case of phase-only CSIS (Section 2.4), $f(x,y) = (\sigma_n^2 Px + 1)/(Py^2)$, $X_L = L^{-1}\sum_{l=1}^{L} |h_l|^2$, $Y_L = L^{-1}\sum_{l=1}^{L} |h_l|$, $x_0 = E[|h_l|^2] = 1$ and $y_0 = E[|h_l|]$. The partial sums X_L and Y_L for each of the cases mentioned in the statement of Theorem 2.5.1 converge in probability due to the law of large numbers for M-dependent sequences [122, Theorem 27.4] and the corresponding $f(\cdot,\cdot)$ satisfies $f(X_L,Y_L) \to f(x_0,y_0)$ in probability, based on the result of Theorem 2.3.1.

∎

Though correlation does not affect the asymptotic variance to which $L\mathrm{var}\left(\hat{\theta}\,\middle|\,\mathbf{h}\right)$ converges, the correlation will affect the speed of convergence. The speed of convergence is now quantified.

Speed of Convergence

It has been shown that $\lim_{L\to\infty} L\mathrm{var}\left(\hat{\theta}\,\middle|\,\mathbf{h}\right)$ converges in probability to a value C, under various conditions. Since $L\mathrm{var}\left(\hat{\theta}\,\middle|\,\mathbf{h}\right) - C$ goes to zero, it will be appropriately normalized with $\sqrt{L}/C$ to ensure its convergence in distribution to a nondegenerate random variable. Toward this goal, consider

$$A[L] := \sqrt{L}\left[\frac{L\mathrm{var}\left(\hat{\theta}\,\middle|\,\mathbf{h}\right) - C}{C}\right] \tag{2.40}$$

The sequence $A[L]$ approaches a Gaussian random variable with zero mean and variance σ_A^2 as $L \to \infty$. This will establish that the normalized difference between $L\mathrm{var}\left(\hat{\theta}\,\middle|\,\mathbf{h}\right)$ and its asymptotic value C scales as $L^{-1/2}$ with σ_A^2 quantifying the size of the discrepancy between $L\mathrm{var}\left(\hat{\theta}\,\middle|\,\mathbf{h}\right)$ and C as a measure of the speed of convergence. Clearly, a small σ_A^2 implies faster convergence than a large σ_A^2. In order to calculate the value of σ_A^2, the behavior of $A[L]$ needs to be analyzed. This approach can be used for all the channel and feedback models considered. As an illustrative example, the case of the phase-only feedback case is examined, and the value of σ_A^2 is derived.

28

$$\sigma_A^2 = 16\pi \left[1 - 2M - \frac{\pi}{4} + \pi \sum_{l=1}^{M} {}_2F_1\left(-\frac{1}{2}, -\frac{1}{2}; 1; |r[l]|^2\right) \right]$$
$$+ \frac{\pi}{4} \left(\frac{\sigma_n^2 P}{\sigma_n^2 P + 1} \right)^2 \left(1 + 2 \sum_{l=1}^{M} |r[l]|^2 \right). \qquad (2.41)$$

Theorem 2.5.2 *For the phase-only case with Rayleigh fading channels, the conditional variance of the estimate is given in (2.23) and C_{PO} is in (2.24). Then, $A[L]$ is asymptotically Gaussian with variance given in (2.41), where $r[l] := E[h_i h_{i-l}^*]$ and ${}_2F_1(\cdot, \cdot; \cdot; \cdot)$ is the Gauss hypergeometric series [119, pp. 556].*

Proof The proof is shown in Section 2.7.

■

Note that ${}_2F_1(-0.5, -0.5; 1; z)$ ranges from 1 to $4/\pi$ so that $\sigma_A^2 > 0$ as expected. Note also that the value of σ_A^2 in (2.41) is a monotonically increasing function of $|r[l]|^2$ for each l. Therefore, if the correlation between any pair of sensors is increased, the convergence is slower, which is expected. It is also monotonically increasing with $\sigma_n^2 P$. Recall that due to the amplify and forward scheme employed at the sensors, P multiplies observation noise. Therefore, increase in P increases observation noise, and one can consider the effect of the product $\sigma_n^2 P$ without loss of generality. Therefore, as either σ_n^2 or P increases, σ_A^2 increases and convergence slows down.

The asymptotic results derived for independent channels between the sensors and the FC continue to hold if the channels are M-dependent. However, σ_A^2 is affected by the degree correlation, and also depends on the values of P, σ_n^2 and M. It should also be noted here that the M-dependent correlation model is adopted for simplicity and more elaborate correlation models can also be used. In fact, any correlation model that satisfies the conditions of the central limit theorem for α-mixing random variables [122, Thm. 27.4, pp. 364] can be used to obtain similar results. Hence, inspecting the proof
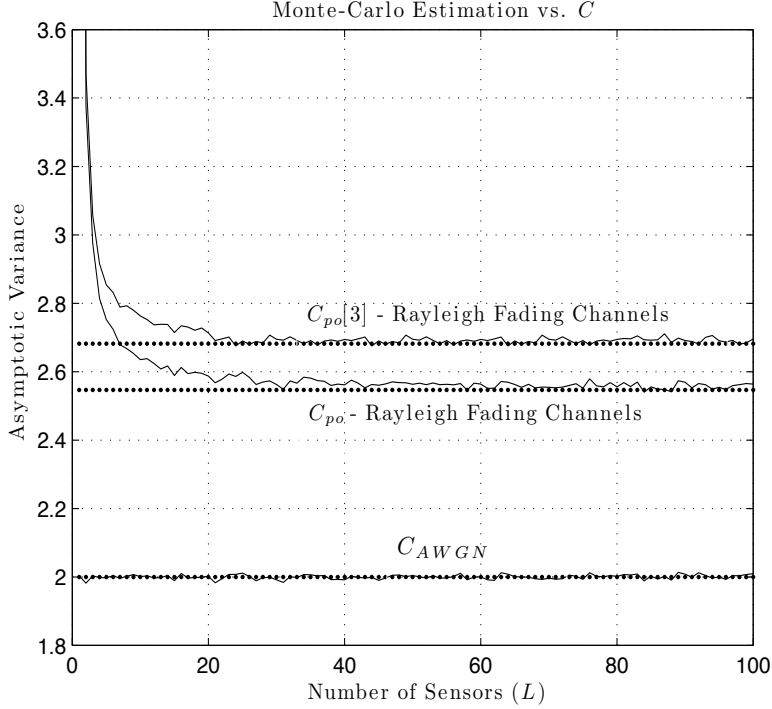
Figure 2.3: The theoretical values (dots) match the Monte-Carlo estimates (solid lines) versus *L*; about 50 sensors are needed for convergence.

2.6 Numerical Results

The results obtained are verified using simulations. Simulations determine how many sensors are required for the asymptotic results to hold. The asymptotic results are then compared against each other. The behavior of $\sigma_{A_C}^2$ from (2.41) is also studied.

Figure 2.3 compares the Monte-Carlo estimates of the asymptotic variances against the values of C_{AWGN}, and C_{PO} and $C_{PO}[Q]$, for $Q = 3$ bits of feedback, over Rayleigh fading channels, when $\sigma_n^2 = 1$, $\theta \sim \mathscr{CN}(0,1)$ and $P = 1$, versus the number of sensors *L*. All the Monte-Carlo estimates are obtained by averaging over 10^5 realizations. It can be seen that the best performance is obtained when the channels are AWGN, and the ratio between the phase only case and AWGN case is exactly a factor of $4/\pi$. There is further loss due to quantization of channel phase feedback. The

30
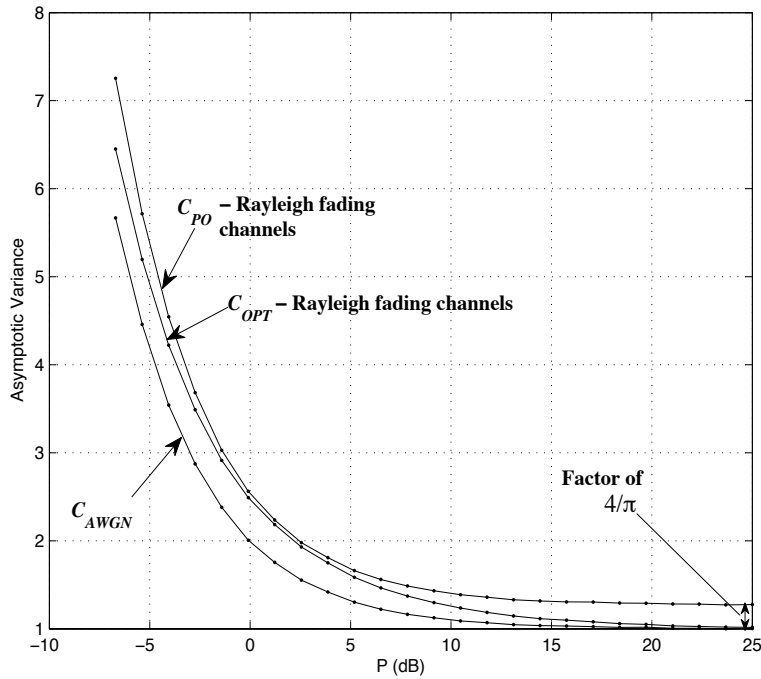
Figure 2.4: Performance of the System vs. P. For large P, AWGN and optimal performance over Rayleigh fading channels is identical.

Monte-Carlo estimates and the theoretical values converge as L increases. As few as $L = 20$ sensors are sufficient to come within 2% of the asymptotic value, and at most $L = 50$ is needed for convergence.

In all subsequent simulations, the values of the asymptotic variance are compared against P. Parameters are set to $\sigma_n^2 = 1$, $\theta \sim \mathscr{C}\mathscr{N}(0,1)$, and $L = 100$ for Monte Carlo simulations.

Figure 2.4 shows the effect of power on performance. Note that C_{PO} closely approximates C_{OPT} for medium amounts of power. For large power, the performance for AWGN channels and perfect CSIS for Rayleigh fading channels is the same, verifying (2.20), whereas the phase-only case performs worse with a deterioration upper-bounded by $4/\pi$, verifying (2.24).

Figure 2.5 shows the effect of quantization on the performance of the system. For two bits of quantization, there is a loss in performance by a factor of about 1 dB
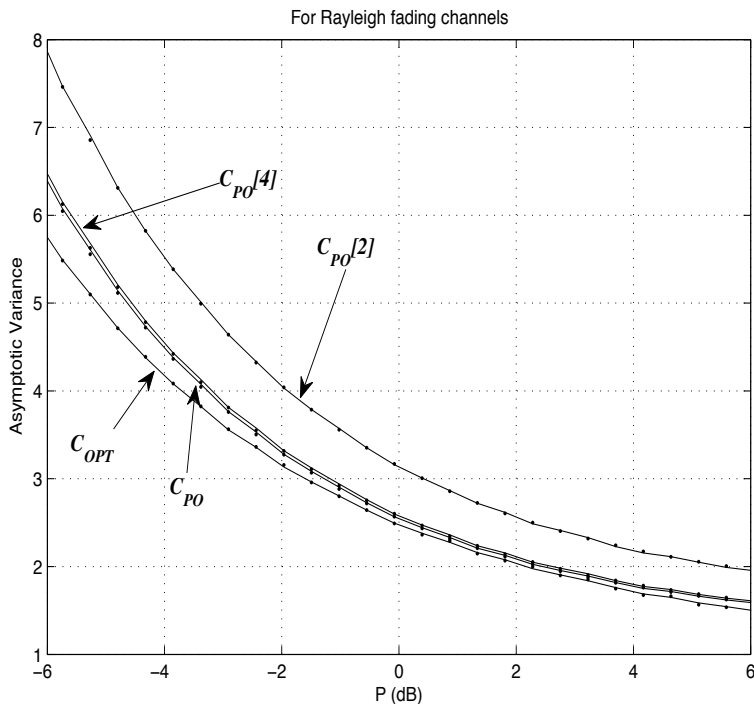
Figure 2.5: Effect of quantization on asymptotic variance - Rayleigh fading channels. As few as four bits of quantization causes negligible loss in performance compared to the phase-only case.

compared to C_{PO}. The loss incurred due to quantization is negligible for $Q = 4$ bits.

When the error in feedback is continuous, the loss in performance can be seen in Figure 2.6. Since the error in feedback is modeled as a von Mises random variable, the performance loss is characterized by the κ parameter. A lower value of κ indicates larger error, and the error goes to zero as $\kappa \to \infty$. The curves in Figure 2.6 also indicate that $\kappa = 50$ is large enough for negligible error in the system.

Figure 2.7 shows the effects of error on the feedback channel. The natural bit mapping analyzed in Section 2.4 is not the only choice. In fact, if Gray coding is used, the performance is marginally better than the natural bit mapping for low powers, and the difference is more clearly visible at high values of p, such as $p = 10^{-1}$ (not shown). In Figure 2.7, $p = 2 \times 10^{-2}$, and the performance of the natural bit-mapping scheme is almost identical to the Gray code, and the approximation, $C_{PO}[Q, p)$ from (2.39), is a
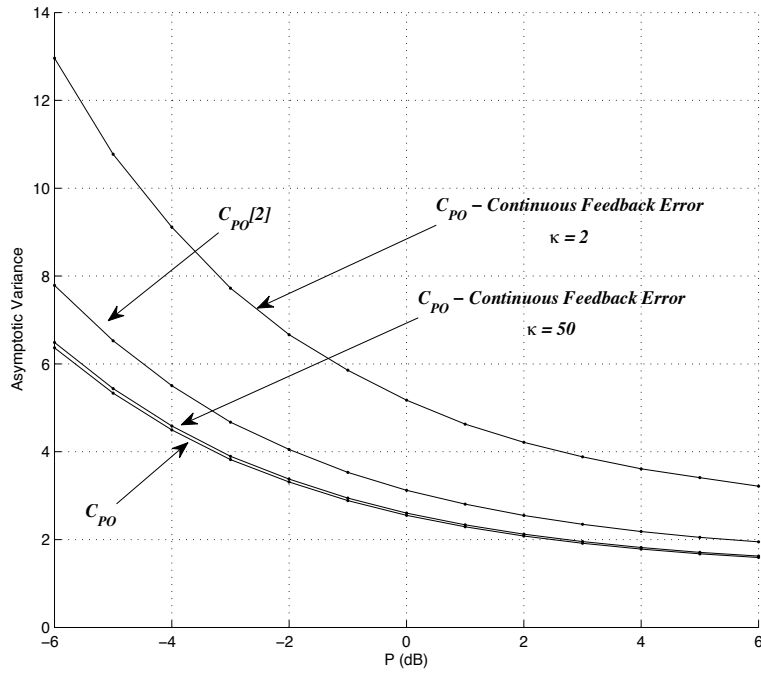
Figure 2.6: Effect of error on asymptotic variance - Rayleigh fading channels. Comparison of the phase-only performance with performance with two bits of channel phase feedback, and continuous error with $\kappa = 2$ and $\kappa = 50$.

very good match to the simulation results.

Figures 2.8 and 2.9 study the performance over Ricean channels. The AWGN case is shown as a benchmark. Figure 2.8 shows the performance of the system over Ricean fading channels with no CSIS. For (large) $K = 20$, the performance is close to the AWGN performance and for (small) $K = 0.1$, the performance is poor. For the partial CSIS case (Figure 2.9), for small K, the performance is close to C_{PO} for Rayleigh fading channels and for large K, the performance is close to C_{AWGN}.

In Fig. 2.10 the joint effect of increasing L and P is considered to compensate for a loss of asymptotic variance due to fading and limited feedback. It is possible to get AWGN performance over fading channels provided that the number of sensors are increased by the correct amount determined by the ratio of the asymptotic variances. A similar idea might be to compensate for this loss by increasing power as well. Figure
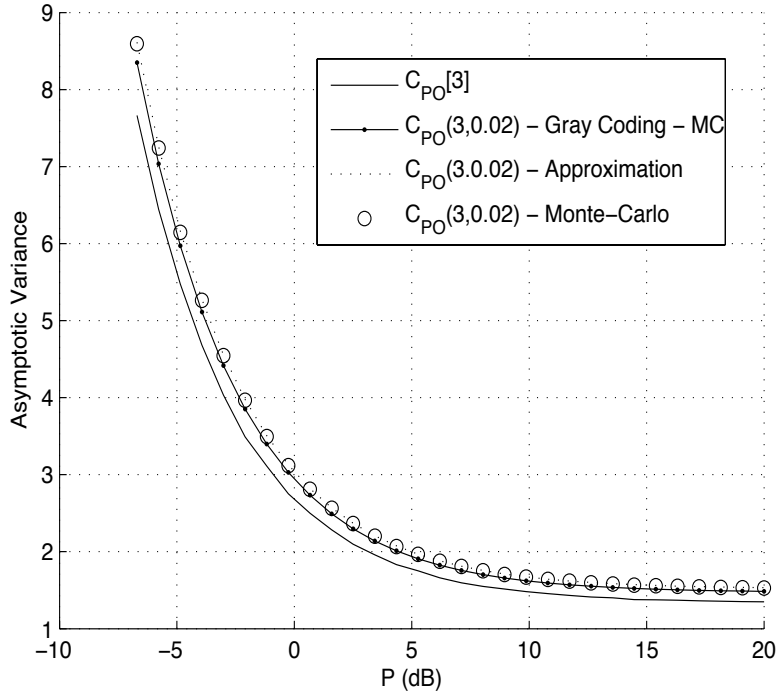
Figure 2.7: Effect of error on feedback channel - Rayleigh fading models. The plot demonstrates the effect of errors on the feedback channel.

2.10 compares the AWGN performance with the phase-only case, as an example. A point in this plot indicates that the phase-only scheme can achieve AWGN performance if its power and number of sensors is larger by a factor indicated by that point. For example, if the power of the phase-only feedback scheme over fading channels is 3 dB above that of the scheme over AWGN channels, then about a 10% increase in the number of sensors is needed to get AWGN performance if $P_{PO} = 5$ dB and $\sigma_n^2 = 1$. It is clear that the penalty paid by increasing the power is substantial. In particular, if we insist that the schemes have the same number of sensors, anywhere between $4 - 15$ dB of increased power is necessary to get the same variance. This indicates that for most practical applications one would opt for increasing the number of deployed sensors rather than increasing power.

Figure 2.11 shows the effect of M and $\sigma_n^2 P$ on σ_A^2, which quantifies the speed of convergence of the asymptotic variance, for the phase-only case discussed in Theorem
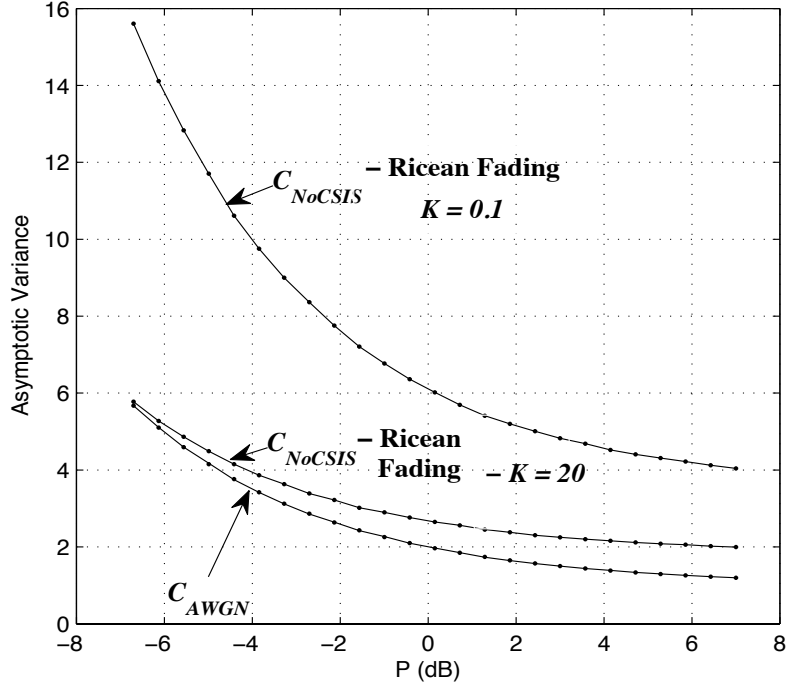
Figure 2.8: No CSIS - Ricean fading channels with large and small K. Performance with large values of K approximates AWGN performance.

2.5.2. The parameter σ_A^2 from (2.41) is compared for $M = 1, 2, \ldots, 8$ with two correlation models, $r[l] = 1, l = 0, 1, \ldots, M$, which implies equal channels, and $r[l] = e^{-0.1l}, l = 0, 1, 2, \ldots, M$, or an exponentially correlated model. For each of the correlation models, the value of σ_A^2 increases as the number of correlated channels increases. Further, as the value of $\sigma_n^2 P$ increases, the value of σ_A^2 increases and convergence slows down. However, the effect of $\sigma_n^2 P$ on σ_A^2 is not as pronounced as the effect of correlation on σ_A^2.

2.7   Proof of Theorem 2.5.2

It is to be shown that $A[L]$ in (2.40) is asymptotically Gaussian with variance given in (2.41) for the phase-only case where the conditional variance, $\text{var}\left(\hat{\theta}\middle|\mathbf{h}\right)$, is given by (2.23). Substituting (2.23) in (2.40),

$$A[L] = a_L \tilde{X}_L - b_L \tilde{Y}_L, \tag{2.42}$$
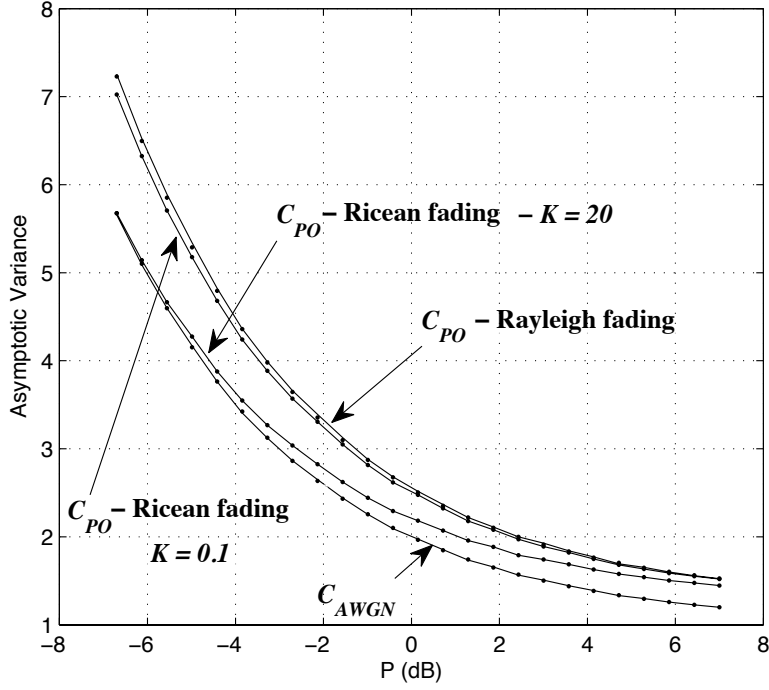
Figure 2.9: Comparison of partial CSIS schemes for Rayleigh fading and Ricean fading channels with small and large K. Performance with large K approximates AWGN performance, and small K performance is similar to performance over Rayleigh fading channels.

where $x_0 := \lim_{L\to\infty} L^{-1} \sum_{i-1}^{L} |h_i|^2 = 1$ and $y_0 := \lim_{L\to\infty} L^{-1} \sum_{i=1}^{L} |h_i| = \sqrt{\pi}/2$. Also, $\tilde{X}_L := \sqrt{L} \left(L^{-1} \sum_{i=1}^{L} |h_i|^2 - x_0 \right)$ and $\tilde{Y}_L := \sqrt{L} \left(L^{-1} \sum_{i=1}^{L} |h_i| - y_0 \right)$, with the definitions, $b_L := (\sigma_n^2 P x_0 + 1) \left([L^{-1} \sum_{l=1}^{L} |h_i|] + |y_0| \right) / \left(P[L^{-1} \sum_{i=1}^{L} |h_i|]^2 |y_0|^2 \right)$ and further, $a_L := \sigma_n^2 \left[L^{-1} \sum_{i=1}^{L} |h_i| \right]^{-2}$.

Due to the weak law of large numbers and Theorem 2.3.1, $a_0 := \lim_{L\to\infty} a_L = 4\sigma_n^2/\pi$ and $b_0 := \lim_{L\to\infty} b_L = 16(\sigma_n^2 P + 1)/(\pi^{3/2} P)$ in probability. Moreover, by invoking the central limit theorem for M-dependent random variables [123], the vector $\tilde{Z}_L := [\tilde{X}_L \quad \tilde{Y}_L]$ is asymptotically Gaussian with a 2×2 covariance matrix Σ whose elements are given by $\Sigma_{1,1} = \lim_{L\to\infty} \text{var}(\tilde{X}_L)$, $\Sigma_{2,2} = \lim_{L\to\infty} \text{var}(\tilde{Y}_L)$ and $\Sigma_{1,2} = \Sigma_{2,1} = \lim_{L\to\infty} \text{cov}(\tilde{X}_L, \tilde{Y}_L)$, where $\text{cov}(X,Y)$ is the covariance between X and Y. Using the
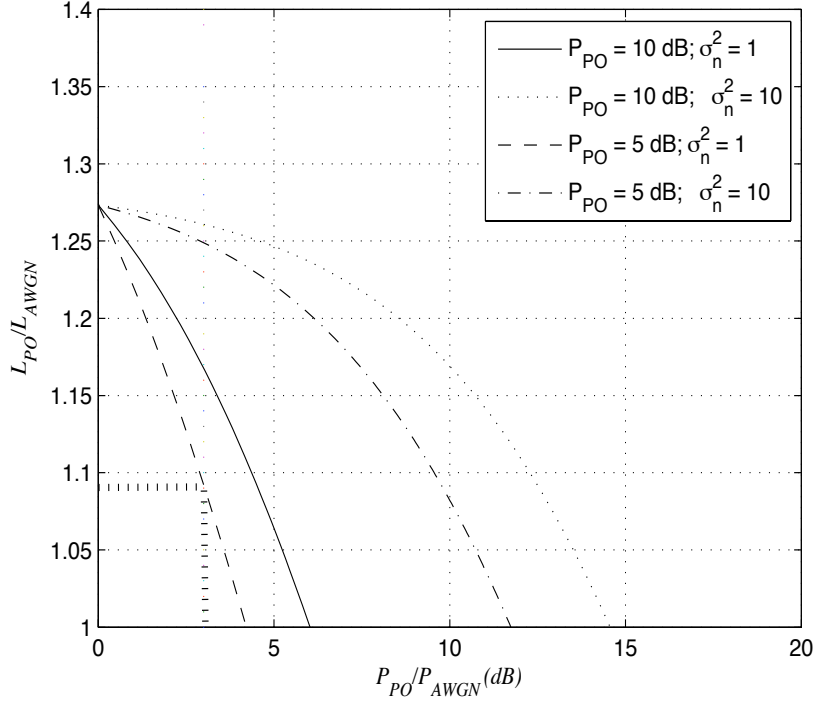
Figure 2.10: Power/sensor penalty for equal variances - AWGN channel case vs. Rayleigh channel with phase-only feedback.

M-dependence of h_i and the fact that it is complex Gaussian,

$$\Sigma_{1,1} = 1 + 2\sum_{l=1}^{M} |r[l]|^2, \tag{2.43}$$

$$\Sigma_{2,2} = 1 - 2M - \frac{\pi}{4} + \pi \sum_{l=1}^{M} {}_2F_1\left(-\frac{1}{2}, -\frac{1}{2}; 1; |r[l]|^2\right), \tag{2.44}$$

and $\Sigma_{1,2} = \Sigma_{2,1} = 0$, where $r[l] = E[h_i h_{i-l}^*]$.

It is now established that $A[L]$ in (2.42) is a linear combination of two asymptotically normal sequences $\tilde{X}_L$ and $\tilde{Y}_L$ where the combining coefficients are sequences that converge in probability. Using [115, Theorem C.4], $A[L]$ is asymptotically normal with zero mean and variance given by $\Sigma_{1,1} + \Sigma_{2,2} - 2\Sigma_{1,2}a_0b_0$. Substituting (2.43) and (2.44), (2.41) is obtained.

■

37
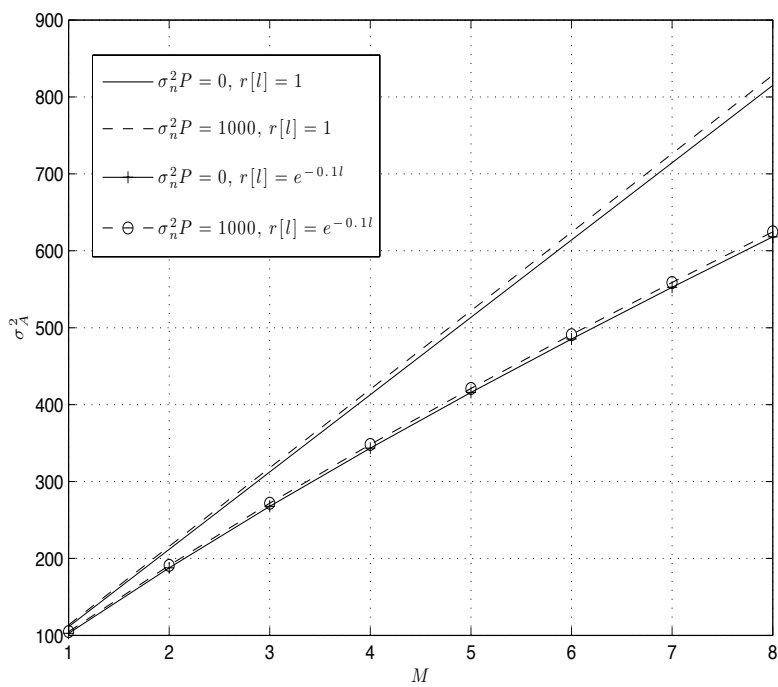
Figure 2.11: Effect of number of correlated channels on σ_A^2.

Chapter 3

DISTRIBUTED DETECTION WITH MULTIPLE ANTENNAS AT THE FUSION

CENTER

3.1 Problem Summary

In this chapter, a distributed detection problem over a multiple access channel, where

the FC has multiple antennas is considered (Figure 3.1). The data collected by the

sensors are transmitted to the FC using the amplify and forward scheme, with a total

power constraint on the sensor gains. Performance is evaluated when the sensors have

no channel information, have full channel information and partial channel information

in the presence of fading, both with zero and non-zero mean. Analysis is performed for

two cases: (a) large number of sensors and a fixed number of antennas, and (b) large

number of antennas *and* sensors with a fixed ratio. In each case, the error exponent is

used as the metric to quantify performance through the effect of channel statistics and

the number of antennas. It is shown that the system performance depends on the chan-

nel distribution through its first and second order moments. This information is used

to address *our main objective*, which is to quantify the gain possible by adding mul-

tiple antennas at the FC over fading multiple-access channels for distributed detection

problems.
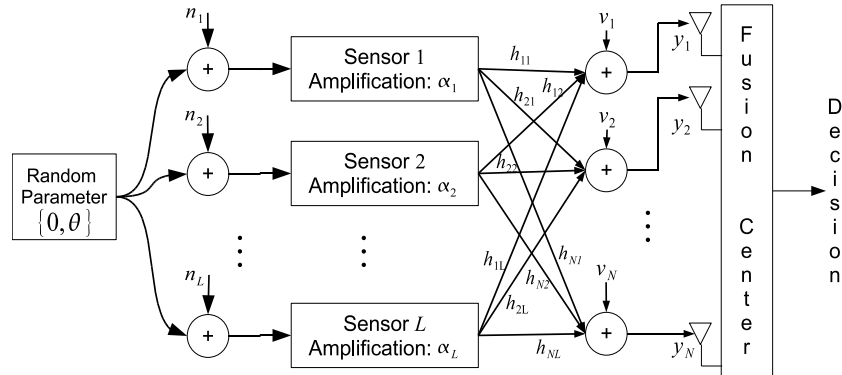
Figure 3.1: System Model: A random parameter is sensed by *L* sensors. Each sensor
transmits amplified observations over fading multiple access channels to a fusion center
with *N* antennas.

3.2 System Model

A sensor network, illustrated in Figure 3.1, consisting of L sensors and a fusion center with N antennas is considered. The sensors are used to observe a parameter $\Theta \in \{0, \theta\}$. The value, x_l, observed at the lth sensor is

$$
x_l = \begin{cases} \eta_l & \text{under} H_0 \\ \theta + \eta_l & \text{under} H_1 \end{cases} \tag{3.1}
$$

for $l = 1, ..., L$. It is assumed that $\eta_l \sim \mathscr{CN}(0, \sigma_\eta^2)$ are iid, the hypothesis H_1 occurs with *a priori* probability, $0 < p_1 < 1$, and the hypothesis H_0 with probability $p_0 = 1 - p_1$. The lth sensor applies a complex gain, α_l, to the observed value, x_l. This amplified signal is transmitted from sensor l to antenna n over a fading channel, h_{nl}, $n = 1, ..., N$, and $l = 1, ..., L$, which are iid and satisfy $E[|h_{nl}|^2] = 1$. Unless otherwise specified, no other assumptions are made on the channel distribution. The nth antenna receives a superposition of all sensor transmissions in the presence of iid channel noise, $v_n \sim \mathscr{CN}(0, \sigma_v^2)$, such that

$$
y_n = \sum_{i=1}^{L} h_{in} \alpha_i (\Theta + \eta_i) + v_n, \tag{3.2}
$$

where $\{\eta_i\}_{i=1}^{L}$ and $\{v_n\}_{n=1}^{N}$ are independent.

Defining α as an $L \times 1$ vector containing $\{\alpha_i\}_{i=1}^{L}$, $\mathbf{D}(\alpha)$ an $L \times L$ diagonal matrix with the components of α along the diagonal, the received signal is expressed in vector form as

$$
\mathbf{y} = \mathbf{H}\alpha\Theta + \mathbf{H}\mathbf{D}(\alpha)\eta + v, \tag{3.3}
$$

where $\mathbf{H}$ is an $N \times L$ matrix containing the elements h_{nl} in the nth row and l^{th} column, η is an $L \times 1$ vector containing $\{\eta_i\}_{i=1}^{L}$, and v is an $N \times 1$ vector containing $\{v_n\}_{n=1}^{N}$. Based on the received signal, $\mathbf{y}$ (from (3.3)), the FC decides on one of the two hypotheses H_0 or H_1. Since the FC has full knowledge of $\mathbf{H}$ and α, $\mathbf{y}$ is Gaussian distributed

under both hypotheses:

$$H_0 : \mathbf{y} \sim \mathscr{C}\mathscr{N}(\mathbf{0}_N, \mathbf{R}(\alpha))$$

$$H_1 : \mathbf{y} \sim \mathscr{C}\mathscr{N}(\theta\mathbf{H}\alpha, \mathbf{R}(\alpha)) \tag{3.4}$$

where $\mathbf{0}_N$ is an $N \times 1$ vector of zeros and $\mathbf{R}(\alpha)$ is the $N \times N$ covariance matrix of the received signal given by

$$\mathbf{R}(\alpha) = \sigma_\eta^2 \mathbf{H}\mathbf{D}(\alpha)\mathbf{D}(\alpha)^H\mathbf{H}^H + \sigma_v^2\mathbf{I}_N. \tag{3.5}$$

We consider detection at a single snapshot in time, and therefore, we do not have a time index.

Power Constraint

The ith sensor transmits $\alpha_i(\Theta + \eta_i)$. The total transmitted power is given by

$$P_T = E\left[\sum_{i=1}^{L} |\alpha_i(\Theta + \eta_i)|^2\right] = \left(p_1\theta^2 + \sigma_\eta^2\right)\sum_{i=1}^{L} |\alpha_i|^2. \tag{3.6}$$

It should also be noted here that the instantaneous transmit power from the sensors is $|\alpha_i(\theta + \eta_i)|^2$. This is a function of the actual realizations of sensing noise, making it difficult to predict and constrain. Therefore, we constrain α_i's, which allows imposing an average (over sensing noise) power constraint. The sensor gains, $\{\alpha_i\}$, are constrained by

$$P := \sum_{i=1}^{L} |\alpha_i|^2 = \frac{P_T}{p_1\theta^2 + \sigma_\eta^2}. \tag{3.7}$$

The Detection Algorithm and its Performance

Given the received data, $\mathbf{y}$, the FC selects the appropriate hypothesis according to

$$\mathfrak{R}\{\theta\mathbf{y}^H\mathbf{R}(\alpha)^{-1}\mathbf{H}\alpha\} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{1}{2}\theta^2\alpha^H\mathbf{H}^H\mathbf{R}(\alpha)^{-1}\mathbf{H}\alpha + \tau, \tag{3.8}$$

where τ is a threshold that can be selected using the Neyman-Pearson or the Bayesian approach. Using (3.4) and (3.8), and the Bayesian test with the detection threshold, $\tau =$

41

$(1/2)\ln(p_0/p_1)$, the probability of error conditioned on the channel can be calculated as

$$P_{e|\mathbf{H}}(N) = p_0 Q\left(\omega + \tau/\omega\right) + p_1 Q\left(\omega - \tau/\omega\right), \tag{3.9}$$

where $\omega := \theta\sqrt{\alpha^H \mathbf{H}^H \mathbf{R}(\alpha)^{-1}\mathbf{H}\alpha/2}$ for brevity, N is the number of antennas at the FC and $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}}e^{-y^2/2}dy$. The error exponent is defined in terms of the conditional error probability for the FC with N antennas as [72, 73, 94]

$$\mathscr{E}(N) = \lim_{L\to\infty} -\frac{1}{L}\log P_{e|\mathbf{H}}(N). \tag{3.10}$$

Note that even though $P_{e|\mathbf{H}}(N)$ in (3.9) is a channel-dependent random variable, we will show that the limit in (3.10) converges in probability to a deterministic constant for the cases of interest to us. Substituting (3.9) into (3.10), using L'Hôpital's rule, and the Leibniz Integral rule for differentiating under the integral sign,

$$\mathscr{E}(N) = \lim_{L\to\infty} \frac{1}{8}\frac{1}{L}\theta^2 \alpha^H \mathbf{H}^H \mathbf{R}(\alpha)^{-1}\mathbf{H}\alpha \tag{3.11}$$

in probability, which does not depend on p_0 and p_1. Since $\mathscr{E}(N)$ is the negative exponent of the probability of error, a larger value represents better performance. The error exponent in (3.11) is a deterministic performance metric over fading channels and depends on fading statistics. It can also be viewed as a "generalized SNR" expression in this system with multiple sensor and channel noise sources. We follow [72, 73, 94] in our definition of the error exponent in (3.10). Alternatively, one can consider the unconditional error exponent, $E_{\mathbf{H}}[P_{e|\mathbf{H}}(N)]$, which would depend on the distribution of $\mathbf{H}$ in (3.10), in place of $P_{e|\mathbf{H}}(N)$. We will not pursue this approach herein.

Our primary focus throughout this paper is the dependence of (3.11) on

(i) the number of antennas, N, for different fading-channel distributions;

(ii) different assumptions about the dependence of the sensor gains, α, on the channel, $\mathbf{H}$.

42

With the Neyman-Pearson test, rather than the Bayesian test, it can be shown that the error exponent is given by $\lim_{L\to\infty} 0.5L^{-1}\theta^2\alpha^H\mathbf{H}^H\mathbf{R}(\alpha)^{-1}\mathbf{H}\alpha$, which does not depend on the false alarm probability and is a factor of four greater than the error exponent derived in the Bayesian case. Since the two cases differ only by a fixed constant, the Bayesian approach will be used throughout.

3.3 Performance over AWGN channels

The error exponent with AWGN channels is computed to establish a benchmark for the fading case of the next section, which is our main focus. For AWGN channels, $h_{nl} = 1$. Due to symmetry and to respect the power constraint, $\alpha_i = \sqrt{P/L}, \forall i$. Defining $\mathbf{1}_L$ as an $L \times 1$ vector of ones, and $\mathbf{1}_{N \times L}$ as an $N \times L$ matrix of ones, we have $\alpha = \sqrt{P/L}\mathbf{1}_L$ and $\mathbf{H} = \mathbf{1}_{N \times L}$. Substituting these in (3.5),

$$\mathbf{R} := \mathbf{R}(\sqrt{P/L}\mathbf{1}_L) = \sigma_\eta^2 P\mathbf{1}_{N \times N} + \sigma_v^2\mathbf{I}_N. \tag{3.12}$$

The inverse of (3.12) can be expressed using the Sherman-Morrison-Woodbury formula for matrix inversion and substituted into (3.11) to yield

$$\mathscr{E}_{\mathrm{AWGN}}(N) := \frac{1}{8}\frac{N\gamma_s\gamma_c}{N\gamma_c + p_1\gamma_s + 1}, \tag{3.13}$$

where the sensing SNR is defined as $\gamma_s := \theta^2/\sigma_\eta^2$, and the channel SNR, $\gamma_c := P_T/\sigma_v^2$. Since the partial derivative $\partial\mathscr{E}_{\mathrm{AWGN}}(N)/\partial N > 0$, for the AWGN case, having multiple antennas improves the error exponent which can be interpreted as array gain on the channel SNR γ_c. As a special case, consider $N = 1$, to get the result for the single antenna case:

$$\mathscr{E}_{\mathrm{AWGN}}(1) = \frac{1}{8}\frac{\gamma_c\gamma_s}{\gamma_c + p_1\gamma_s + 1}. \tag{3.14}$$

With $p_1 = 0.5$, $\gamma_c = 1$ and $\gamma_s = 1$, adding a second antenna at the FC provides a gain of 3.1dB. Adding a third antenna provides a further gain of 1.34dB, indicating diminishing returns. To study the benefits of having multiple antennas, we compare the error exponent in each case with $\mathscr{E}_{\mathrm{AWGN}}(1)$. The multiple antenna gain for the AWGN case

43

is given by

$$G_{AWGN}(N) := \frac{\mathcal{E}_{AWGN}(N)}{\mathcal{E}_{AWGN}(1)} = \frac{N\gamma_c + Np_1\gamma_s + N}{N\gamma_c + p_1\gamma_s + 1}. \tag{3.15}$$

It can be seen from (3.15) that by making N sufficiently large, and γ_c sufficiently small, (3.15) can be made arbitrarily large. In contrast, it will be seen in Section 3.4 that when the channels are fading and known at the sensors, the corresponding gain expression will be bounded for all parameter values, indicating limited gains due to antennas.

3.4 Performance over Fading Channels

Suppose that the elements of the channel matrix, $\mathbf{H}$, are non-zero-mean, that is, $h_{nl} = \sqrt{K/(K+1)} + (1/\sqrt{K+1})h_{nl}^{\text{diff}}$, where the first term is the line-of-sight (LOS) component, h_{nl}^{diff} is the zero-mean diffuse component, and the parameter K is the ratio of the LOS power to the power of the diffuse component, chosen so that the channel satisfies $E[|h_{nl}|^2] = E[|h_{nl}^{\text{diff}}|^2] = 1$.

In what follows, different cases of channel state information at the sensors (CSIS) are considered.

No Channel State Information at the Sensors

When the sensors have no channel knowledge, then the sensor gains are set to $\alpha = \sqrt{P/L}\mathbf{1}_L$ due to the i.i.d. nature of the channels and to respect the power constraint in (3.7). Substituting in (3.5),

$$\mathbf{R} := \mathbf{R}(\sqrt{P/L}\mathbf{1}_L) = \sigma_\eta^2 P\frac{1}{L}\mathbf{H}\mathbf{H}^H + \sigma_v^2\mathbf{I}_N. \tag{3.16}$$

Since the elements of $\mathbf{H}$ are i.i.d., from the weak law of large numbers,

$$\lim_{L\to\infty} \mathbf{R} = \sigma_\eta^2 P\frac{K}{K+1}\mathbf{1}_{N\times N} + \frac{\sigma_\eta^2 P + \sigma_v^2(K+1)}{K+1}\mathbf{I}_N, \tag{3.17}$$

in probability. Since the right-hand-side of (3.17) is non-singular, it can be seen that $\lim_{L\to\infty}\mathbf{R}^{-1} = (\lim_{L\to\infty}\mathbf{R})^{-1}$ [124, Thm. 2.3.4]. Using the matrix inversion lemma on

(3.17) and substituting into (3.11),

$$\mathcal{E}_{\text{NoCSIS}}(N,K) = \frac{\theta^2}{8} \frac{P(K+1)}{\sigma_\eta^2 P + \sigma_v^2(K+1)} \lim_{L \to \infty} \sum_{n=1}^{N} \left| \frac{1}{L} \sum_{l=1}^{L} h_{nl} \right|^2$$

$$- \frac{\theta^2}{8} \frac{\sigma_\eta^2 P^2 K(K+1)}{\left[\sigma_\eta^2 P + \sigma_v^2(K+1) \right] \left[\sigma_\eta^2 PNK + \sigma_\eta^2 P + \sigma_v^2(K+1) \right]} \lim_{L \to \infty} \left| \frac{1}{L} \sum_{n=1}^{N} \sum_{l=1}^{L} h_{nl} \right|^2 . \quad (3.18)$$

Using the weak law of large numbers and (3.7), the error exponent can be expressed in terms of γ_c and γ_s as

$$\mathcal{E}_{\text{NoCSIS}}(N,K) := \frac{1}{8} \frac{NK\gamma_c\gamma_s}{\gamma_c(NK+1) + (p_1\gamma_s+1)(K+1)}, \quad (3.19)$$

which can be shown to be a monotonically increasing function of N, K, γ_s and γ_c, as expected. For the single antenna case, using (3.14) it can be seen that $\mathcal{E}_{\text{NoCSIS}}(1,K) = \mathcal{E}_{AWGN}(1)K/(K+1)$, which is a factor $K/(K+1)$ worse than $\mathcal{E}_{AWGN}(1)$.

As the antennas increase, $\lim_{N \to \infty} \mathcal{E}_{\text{NoCSIS}}(N,K) = \gamma_s/8$, which is the same as $\lim_{N \to \infty} \mathcal{E}_{AWGN}(N)$. That is, so long as there is some non-zero LOS component, as the number of antennas at the FC increases, the performance approaches the AWGN performance even in the absence of CSI at the sensors. Furthermore, it can be seen that $\lim_{K \to \infty} \mathcal{E}_{\text{NoCSIS}}(N,K) = \mathcal{E}_{AWGN}(N)$, which matches the AWGN result, as expected.

To characterize the gain due to having multiple antennas at the FC, we define

$$G_{\text{NoCSIS}}(N,K) := \frac{\mathcal{E}_{\text{NoCSIS}}(N,K)}{\mathcal{E}_{\text{NoCSIS}}(1,K)} = \frac{N(K+1)(\gamma_c + p_1\gamma_s + 1)}{\gamma_c(NK+1) + (p_1\gamma_s+1)(K+1)}. \quad (3.20)$$

When the channel noise is large, ($\gamma_c \to 0$), we have $G_{\text{NoCSIS}}(N,K) = N$ and the gain increases with the number of antennas at the FC. However, when $\gamma_c \to 0$, the absolute performance of the system is poor, as can be verified by substituting in (3.19). Conversely, when the channel SNR grows, the maximum gain in (3.20) is given by $(K+1)/K$. This leads to the conclusion that when the channels between the sensors and the FC are relatively noise-free, there is little advantage in having multiple antennas at the FC when K is large. When the channel is zero-mean ($K=0$), the error exponent in (3.19) is zero for any N, indicating that the probability of error does not decrease

45

exponentially with L for any N, confirming results from [57, 66, 75]. However, from (3.20), it is clear that the *gain* satisfies $\lim_{K\to 0} G_{\text{N}oCSIS}(N,K) = N$, which shows that when the channel is zero-mean, gain in the error exponent due to antennas is linear and can be made arbitrarily large. We have thus established the following:

Theorem 3.4.1 *For zero-mean channels, with no CSI at the sensors, the error exponent in (3.19) is zero and therefore, the error probability does not decrease exponentially with L for any number of antennas, N. The antenna gain, defined in (3.20) satisfies $\lim_{K\to 0} G_{NoCSIS}(N,K) = N$, implying unlimited gains from multiple antennas for zero-mean channels when CSI is unavailable at the sensors.*

In what follows, it will be seen that when CSI is available at the sensors, the antenna gain is bounded over all parameter values for zero-mean channels.

Channel State Information at the Sensors

We have just seen that when the non-zero-mean channel assumption does not hold, the incoherent sum of signals at each each antenna leads to poor performance at the FC, which results in a zero error exponent. If channel information is available at the sensors, the sensor gains can be adjusted in such a way that the signals are combined coherently. It should be noted here that full CSI at the sensors implies full CSI of the network, $\mathbf{H}$, at the sensors. In such a case, α is chosen as a function of the channels, $\mathbf{H}$.

As a benchmark result for fading channels, the sensor gains are selected in such a way as to maximize the error exponent of the system given in (3.11), subject to the power constraint in (3.7):

$$\alpha_{OPT} = \underset{\alpha}{\text{argmax}} \left[\alpha^H \mathbf{H}^H \mathbf{R}(\alpha)^{-1} \mathbf{H}\alpha \right] \quad \text{subject to } \|\alpha\|^2 \leq P, \qquad (3.21)$$

to obtain the error exponent in the presence of CSIS,

$$\mathscr{E}_{\text{CSIS}}(N) = \lim_{L\to\infty} \frac{\theta^2}{8} \frac{1}{L} \alpha_{OPT}^H \mathbf{H}^H \left(\sigma_\eta^2 \mathbf{H}\mathbf{D}(\alpha_{OPT})\mathbf{D}(\alpha_{OPT})^H \mathbf{H}^H + \sigma_v^2 \mathbf{I}_N \right)^{-1} \mathbf{H}\alpha_{OPT}. \qquad (3.22)$$

The optimization problem in (3.21) is not tractable when $N > 1$ since $\mathbf{R}(\alpha)$ depends on $\mathbf{H}$ and α. In order to assess the effect of number of antennas, the solution for (3.22) with $N = 1$, and two upper bounds on (3.22) are derived for $N > 1$.

Solution for Single Antenna at the FC

When $N = 1$, the channel matrix reduces to a column vector, given by $[h_1 h_2 \ldots h_L]^T$, where h_i is the channel between the i-th sensor and the FC. The maximization problem in (3.21) reduces to

$$\alpha_{OPT} = \underset{\alpha}{\text{argmax}} \frac{\left| \sum_{i=1}^{L} \alpha_i h_i \right|^2}{\sigma_\eta^2 \sum_{i=1}^{L} |\alpha_i|^2 |h_i|^2 + \sigma_v^2} \qquad \text{subjectto} \sum_{i=1}^{L} |\alpha_i|^2 \leq P. \qquad (3.23)$$

A similar problem was formulated in [76] and in a distributed estimation framework in [57, 81]. We recognize that the best value for the phase of the sensor gain is $\angle \alpha_l = -\psi_l$ where $\psi_l = \angle h_l$. Therefore, we set $\angle \alpha_l = -\psi_l, \forall l$. We then define $s := \sum_{i=1}^{L} \alpha_i h_i$ and swap the objective function with the constraint so we can rewrite the optimization problem as

$$\alpha_{OPT} = \underset{\{|\alpha_i|\}, s}{\text{argmin}} \sum_{k=1}^{L} |\alpha_k|^2 \qquad \text{subjectto} \quad \sigma_\eta^2 \sum_{l=1}^{L} |\alpha_l|^2 |h_l|^2 + 1 \leq v_t s^2$$

$$\sum_{l=1}^{L} (|\alpha_l||h_l|) - s = 0, \qquad (3.24)$$

where v_t is an auxiliary variable. The optimization problem in (3.24) is now a (convex) second-order-cone problem [118]. Using the Karush-Kuhn-Tucker conditions [118], the optimal solution is given by

$$\alpha_i = \sqrt{\frac{P}{\sum_{l=1}^{L} \left(\frac{|h_l|}{P|h_l|^2 \sigma_\eta^2 + \sigma_v^2} \right)^2}} \left(\frac{|h_i|}{\sigma_\eta^2 P |h_i|^2 + \sigma_v^2} \right) e^{-j\angle h_i}. \qquad (3.25)$$

The error exponent can be obtained by substituting (3.25) in (3.22) with $N = 1$:

$$\mathscr{E}_{CSIS}(1) = \lim_{L \to \infty} \frac{\theta^2}{8} \frac{1}{L} \sum_{l=1}^{L} \frac{1}{\sigma_\eta^2 + \frac{\sigma_v^2}{P|h_l|^2}} = \frac{\theta^2}{8} E \left[\frac{1}{\sigma_\eta^2 + \frac{\sigma_v^2}{P|h_l|^2}} \right] \qquad (3.26)$$

from the weak law of large numbers, where the expectation is with respect to $\{h_l\}$. As an example, for Rayleigh fading channels (3.26) yields [121, §3.353]

$$\mathscr{E}_{CSIS}(1) = \frac{1}{32}\gamma_s \left[2 - \frac{p_1\gamma_s + 1}{\gamma_c} \exp\left(\frac{p_1\gamma_s + 1}{2\gamma_c}\right) E_1\left(\frac{p_1\gamma_s + 1}{2\gamma_c}\right) \right], \tag{3.27}$$

where $E_1(\cdot)$ is an exponential integral function [119, pp. 228]. The expression for $\mathscr{E}_{CSIS}(1)$ is obtained when the channels between the sensors and the FC are fading. To compare with the AWGN case, note that $Px/(\sigma_\eta^2 Px + \sigma_v^2)$ in (3.26) is a concave function of x, and from Jensen's inequality, $\mathscr{E}_{AWGN}(1) \geq \mathscr{E}_{CSIS}(1)$, as expected.

Since (3.26) is rather complicated, it is desirable to find a simpler expression as a lower bound to (3.26). Any choice of $\|\alpha\|^2 = P$ will yield such a lower bound, since α_{OPT} is optimal. Considering phase-only correction at the sensors, $\alpha_i = \sqrt{P/L}\exp(-j\angle h_i)$ is substituted in (3.11) with $N = 1$ to yield the error exponent for phase-only CSIS for $N = 1$:

$$\mathscr{E}_{PO}(1) = \lim_{L\to\infty} \frac{\theta^2}{8} \frac{P\left[\frac{1}{L}\sum_{l=1}^{L}|h_l|\right]^2}{\sigma_\eta^2 P\frac{1}{L}\sum_{l=1}^{L}|h_l|^2 + \sigma_v^2}. \tag{3.28}$$

From the weak law of large numbers, the random sequences in the numerator and denominator converge separately. However, since the expression for $\mathscr{E}_{PO}(1)$ is a continuous function of these sequences, the value of $\mathscr{E}_{PO}(1)$ converges to [115, Thm. C.1]

$$\mathscr{E}_{PO}(1) = (E[|h_l|])^2 \mathscr{E}_{AWGN}(1) \tag{3.29}$$

in probability, since $E[|h_l|^2] = 1$. The expression in (3.29) serves as a lower bound to $\mathscr{E}_{CSIS}(1)$ as follows:

$$\frac{1}{\zeta}\mathscr{E}_{AWGN}(1) \leq \mathscr{E}_{CSIS}(1) \leq \mathscr{E}_{AWGN}(1), \tag{3.30}$$

where $\zeta = (E[|h_l|])^{-2}$.

Upper Bound (AWGN channels)

Since (3.21) cannot be solved in closed form when $N > 1$, one cannot evaluate the error exponent in (3.22) by substitution as it was done for $N = 1$. Two upper bounds on (3.22) will be convenient at this stage. Since the AWGN performance is a benchmark for fading channels, the error exponent of the system over AWGN channels is an upper bound on that of fading channels, even in the case of full CSIS. Therefore, the first upper bound to (3.22) is given in (3.13):

$$\mathscr{E}_{CSIS}(N) \leq \mathscr{E}_{AWGN}(N) = \frac{1}{8} \frac{N\gamma_s\gamma_c}{N\gamma_c + p_1\gamma_s + 1}. \tag{3.31}$$

Upper Bound (No Sensing Noise)

Clearly, (3.22) is a monotonically decreasing function of the sensing noise variance, σ_η^2. The second benchmark is obtained by setting $\sigma_\eta^2 = 0$, which also affects α_{OPT} in (3.21), since $\mathbf{R}(\alpha)$ no longer depends on α when $\sigma_\eta^2 = 0$. Substituting this in (3.21), the optimal value of α when $\sigma_\eta^2 = 0$ is

$$\underset{\alpha}{\text{argmax}} \left(\alpha^H \mathbf{H}^H \mathbf{H} \alpha \right) \quad \text{subject to} \, \|\alpha\|^2 \leq P. \tag{3.32}$$

The solution to (3.32) is the eigenvector corresponding to the maximum eigenvalue of $\mathbf{H}^H\mathbf{H}$, scaled in a way to satisfy the constraint with equality. Substituting into (3.22) with $\sigma_\eta^2 = 0$, we have the second upper bound to $\mathscr{E}_{CSIS}(N)$:

$$B(N,K) = \frac{\theta^2}{8} \frac{P}{\sigma_v^2} \lim_{L\to\infty} \lambda_{max}\left(\frac{1}{L}\mathbf{H}^H\mathbf{H}\right), \tag{3.33}$$

where $\lambda_{max}(\cdot)$ denotes the maximum eigenvalue function. Since it can be seen that $\lambda_{max}(\mathbf{H}^H\mathbf{H}) = \lambda_{max}(\mathbf{H}\mathbf{H}^H)$, and $\lambda_{max}(\cdot)$ is a continuous function of the matrix elements [124, Thm. 8.1.5], one can interchange the limit with the maximum eigenvalue function [115, pp. 422, Thm. C.1] to yield

$$B(N,K) = \frac{\theta^2}{8} \frac{P}{\sigma_v^2} \lambda_{max}\left(\lim_{L\to\infty}\frac{1}{L}\mathbf{H}\mathbf{H}^H\right). \tag{3.34}$$

From the weak law of large numbers,

$$\lim_{L\to\infty}\frac{1}{L}\mathbf{H}\mathbf{H}^H = \frac{K}{K+1}\mathbf{1}_{N\times N} + \frac{1}{K+1}\mathbf{I}_{N\times N}, \tag{3.35}$$

49

in probability, so that with the substitutions $\sigma_\eta^2 = 0$ and $\theta^2 P / \sigma_v^2 = \gamma_c / p_1$, we have the bound:

$$\mathscr{E}_{\text{CSIS}}(N) \le B(N, K) = \frac{1}{8} \frac{\gamma_c}{p_1} \frac{NK + 1}{K + 1}. \tag{3.36}$$

In (3.36), $B(N, K)$ is an upper bound when there is sensing noise in the system. When there is no sensing noise, it is the actual error exponent of the system with full CSIS. Furthermore, $\lim_{K \to \infty} B(N, K) = \lim_{\gamma_s \to \infty} \mathscr{E}_{AWGN}(N)$, verifying that as $K \to \infty$, $B(N, K)$ converges to the AWGN error exponent with no sensing noise. In addition, if $K = 0$, there is no advantage to having multiple antennas at the FC, for asymptotically large number of sensors, since the right hand side of (3.36) is independent of N in that case.

Since both $\mathscr{E}_{AWGN}(N)$ and $B(N, K)$ are upper bounds to $\mathscr{E}_{CSIS}(N)$, a combination of the two bounds, $\min[\mathscr{E}_{AWGN}(N), B(N, K)]$, provides a single, tighter upper bound. Equating the right hand sides of (3.31) and (3.36), it can be shown that this combined upper bound is given by

$$C(N, K) = \begin{cases} \mathscr{E}_{AWGN}(N) & \text{if} \quad \sigma_\eta^2 \ge \frac{N-1}{N(NK+1)} \\ B(N, K) & \text{if} \quad \sigma_\eta^2 \le \frac{N-1}{N(NK+1)} \end{cases}. \tag{3.37}$$

Combining the upper and lower bounds,

$$\frac{1}{\zeta} \mathscr{E}_{AWGN}(1) \le \mathscr{E}_{CSIS}(1) \le \mathscr{E}_{CSIS}(N) \le C(N, K), \tag{3.38}$$

obtained from (3.27), (3.30) and (3.37). The bounds in (3.38) will be used to further examine the effect of N on $\mathscr{E}_{CSIS}(N)$.

The value of $\mathscr{E}_{CSIS}(N)$ from (3.22) is the best achievable performance for fading channels. Defining the gain due to multiple antennas in the case of full CSI at the sensors as $G_{CSIS}(N) := \mathscr{E}_{CSIS}(N) / \mathscr{E}_{CSIS}(1)$, the following theorem can be stated:

Theorem 3.4.2 *When the channels have full CSI at the sensors, the gain due to multiple antennas at the FC can be upper bounded as*

$$G_{CSIS}(N) \le \zeta \frac{\mathscr{E}_{CSIS}(N)}{\mathscr{E}_{AWGN}(1)} \le \zeta \min \left[\frac{N(z+1)}{Nz+1}, (z+1) \frac{NK+1}{K+1} \right], \tag{3.39}$$

where $z := \gamma_c/(p_1\gamma_s + 1)$.

Proof The first inequality in (3.39) follows from the first inequality in (3.38). The second inequality in (3.39) follows from the last inequality in (3.38) and dividing the terms of (3.37) by (3.14).

∎

With $p_1 = 0.5$, $K = 1$, $\gamma_c = 1$ and $\gamma_s = 1$, for $N = 2$, $G_{CSIS}(2) \le 1.4286\zeta$. For $N = 3$, $G_{CSIS}(3) \le 1.6667\zeta$ and for $N = 4$, $G_{CSIS}(4) \le 1.8182\zeta$. These results indicate that there is diminishing returns in the multiple antenna gain.

Corollary 3.4.3 *$G_{CSIS}(N)$ can be bounded by an expression depending on N and K only:*

$$G_{CSIS}(N) \le \zeta\frac{N^2K + 2N - 1}{N(K+1)} \tag{3.40}$$

Proof The first argument of the $\min[\cdot,\cdot]$ function of the right hand side of (3.39) is a decreasing function in z and the second argument is an increasing function in z. Therefore, when the arguments are equal for fixed values of N and K, the maximum value of the $\min[\cdot,\cdot]$ function is obtained. This occurs when $z = N^{-1}(NK+1)^{-1}(N-1)$, allowing us to upper bound the $\min[\cdot,\cdot]$ function by the value in (3.40).

∎

Corollary 3.4.4 *When the channels have zero-mean, the maximum gain due to having multiple antennas at the FC is bounded by a constant independent of N and only dependent on $\zeta = (E[|h_l|])^{-2}$:*

$$G_{CSIS}(N) \le 2\zeta. \tag{3.41}$$

Proof Substituting $K = 0$, it is clear that (3.40) is monotonically increasing in N. Taking the limit as $N \to \infty$ yields the proof.

∎

As an example, in the case of Rayleigh fading, when full channel information is available at the sensors, the maximum gain that can be obtained by adding any number of antennas at the FC for any channel or sensing SNR is at most $2\zeta = 8/\pi$, which is less than 3.

The results in (3.39)-(3.41) have been derived for the case of iid sensing noise. We now address the correlated sensing noise case. To this end, we define $\mathbf{R}_\eta$ as the $L \times L$ covariance matrix of the sensing noise samples, $\{\eta_l\}_{l=1}^{L}$.

Theorem 3.4.5 *Suppose that the sensing noise samples are correlated and let λ_{min} be the minimum eigenvalue of $\mathbf{R}_\eta$. The gain due to multiple antennas in (3.39) holds with the change $z = \gamma_c / (p_1 \tilde{\gamma}_s + 1)$, where $\tilde{\gamma}_s := \theta^2 / \lambda_{min}$.*

Proof The proof is shown in Section 3.6.

∎

It can be seen from Theorem 3.4.5 that any full-rank sensing noise covariance matrix changes the conclusion in (3.39) only through a redefinition of z. By maximizing over z, the same upper-bound in (3.40) is obtained, and for zero-mean channels, the bound in (3.41) remains valid. This shows that the bounds in (3.40) and (3.41) are general, and hold even when the iid condition is relaxed to any arbitrary full-rank covariance matrix, $\mathbf{R}_\eta$. The gain due to adding multiple antennas is still upper-bounded by a factor of 2ζ, for zero-mean channels, when there is full CSI at the sensors.

Phase-only CSIS

One simplification to the full CSIS case is to provide only channel phase information to the sensors. For the single antenna case, and when the channels between the sensors

and the FC have zero-mean, the phase-only results have been presented in (3.29) and (3.30). What follows is an extension of those results to the multiple antenna case when $K = 0$.

Since there is only phase information at the sensors, the amplitudes of the sensor gains are selected such that $|\alpha_l| = \sqrt{P/L}, \forall l$, so that $\mathbf{D}(\alpha)\mathbf{D}(\alpha)^H = (P/L)\mathbf{I}_L$ and $\mathbf{R}(\alpha)$ is given by (3.16).

With phase-only information, one can constrain $|\alpha_i|$ to be constant to reformulate (3.21) as the following:

$$\alpha_{\text{PO}} = \underset{\alpha}{\text{argmax}} \, \alpha^H \mathbf{H}^H \mathbf{H} \alpha \quad \text{subject to} \quad |\alpha_i|^2 = \frac{P}{L}, i = 1, 2, \ldots, L. \tag{3.42}$$

In Section 3.4, a semidefinite relaxation approach will be presented to solve (3.42).

Asymptotically large sensors and *antennas*

When CSIS is available, (3.39 - 3.41) shows that only limited multiple antenna gains are available. It is interesting to see whether such limits would still be present if $N \to \infty$ *simultaneously* with L. A similar problem was considered, but in the context of CDMA transmissions in [74]. Note that this will in general yield results different than first sending $L \to \infty$ and then $N \to \infty$ as was done in Section 3.4. Such a situation can be interpreted as a case where a group of sensors is transmitting to another group, functioning as a virtual antenna array [125]. For such a system the scaling laws when L and N simultaneously increase [126, pp. 7], in such a way that

$$\lim_{L,N \to \infty} \frac{L}{N} = \beta, \tag{3.43}$$

are of interest. It should be noted that in spite of scaling the number of sensors and antennas, the power constraint is still maintained.

In this case, the error exponent is redefined as

$$\mathscr{E}^\infty(\beta) = \lim_{L,N \to \infty} -\frac{1}{L} \log P_{e|\mathbf{H}}(N), \tag{3.44}$$

53

with (3.43) satisfied. Similar to the upper bounds in (3.31) and (3.36), upper bounds on (3.44) are now derived. For the AWGN case,

$$\mathscr{E}^\infty(\beta) \le \mathscr{E}^\infty_{AWGN} := \lim_{L,N\to\infty} \mathscr{E}_{AWGN}(N) = \lim_{L,N\to\infty} \frac{1}{8} \frac{N\gamma_s\gamma_c}{N\gamma_c + p_1\gamma_s + 1} = \frac{1}{8}\gamma_s. \quad (3.45)$$

When there is no sensing noise, with $\sigma_\eta^2 = 0$, the second bound can be calculated as

$$\mathscr{E}^\infty(\beta) \le B^\infty(\beta) := \lim_{L,N\to\infty} \frac{\theta^2}{8} \frac{P}{\sigma_v^2} \lambda_{max}\left(\frac{1}{L}\mathbf{H}^H\mathbf{H}\right). \quad (3.46)$$

For fading channels with $K > 0$, it can be shown that the error exponent in (3.46) goes to infinity. Therefore, with any line-of-sight (LOS) and no sensing noise, increasing the number of sensors and the number of antennas to infinity provides very good performance. When $K = 0$, the Marčenko-Pastur Law [126, pp. 56] provides an empirical distribution of the eigenvalues of $N^{-1}\mathbf{H}^H\mathbf{H}$. From [127, 128], the maximum eigenvalue of $N^{-1}\mathbf{H}^H\mathbf{H}$ is shown to converge in such a way that

$$\lim_{L,N\to\infty} \lambda_{max}\left[\left(\frac{1}{\sqrt{N}}\mathbf{H}\right)^H\left(\frac{1}{\sqrt{N}}\mathbf{H}\right)\right] = \frac{\left(1+\sqrt{\beta}\right)^2}{\beta}, \quad (3.47)$$

in probability, which yields

$$B^\infty(\beta) = \frac{1}{8}\frac{\gamma_c}{p_1}\frac{(1+\sqrt{\beta})^2}{\beta}, \quad (3.48)$$

which is the optimum performance of the system in the absence of sensing noise. Similar to (3.37), the minimum of (3.45) and (3.48) yields

$$\mathscr{E}^\infty(\beta) \le \min\left[\mathscr{E}^\infty_{AWGN}, B^\infty(\beta)\right] = \begin{cases} \frac{1}{8}\gamma_s & \text{if } P\sigma_\eta^2 \ge \frac{\beta}{(1+\sqrt{\beta})^2} \\ \frac{1}{8}\frac{\gamma_c}{p_1}\frac{(1+\sqrt{\beta})^2}{\beta} & \text{if } P\sigma_\eta^2 \le \frac{\beta}{(1+\sqrt{\beta})^2} \end{cases}. \quad (3.49)$$

The gain due to antennas is expressed in terms of the ratio β in (3.43) and can be written as $G^\infty(\beta) := \mathscr{E}^\infty(\beta)/\mathscr{E}_{CSIS}(1)$. Using the bounds, we have the following:

Theorem 3.4.6 *With asymptotically large number of sensors and antennas, the gain due to having multiple antennas at the FC is bounded by*

$$G^\infty(\beta) \le \zeta\left(1 + \frac{\left(1+\sqrt{\beta}\right)^2}{\beta}\right). \quad (3.50)$$

54

Proof The relationship between $\mathscr{E}_{AWGN}(1)$ and $\mathscr{E}_{CSIS}(1)$ from (3.30) provides a lower bound on $\mathscr{E}_{CSIS}(1)$, and consequently an upper bound on $G^\infty(\beta)$, to yield the first inequality in (3.51) below. The expression in (3.49) provides an upper bound on $\mathscr{E}^\infty(\beta)$, and dividing by (3.14) yields the second inequality in

$$G^\infty(\beta) \le \zeta \frac{\mathscr{E}^\infty(\beta)}{\mathscr{E}_{AWGN}(1)} \le \zeta \min\left[1 + \frac{1}{w}, (1+w)\frac{(1+\sqrt{\beta})^2}{\beta}\right], \qquad (3.51)$$

where $w := \gamma_c/(p_1\gamma_s + 1)$. The first argument in the $\min[\cdot,\cdot]$ function decreases as w increases, while the second argument is an increasing function of w. Therefore, the $\min[\cdot,\cdot]$ function is maximized when arguments of the $\min[\cdot,\cdot]$ function are equal for a fixed value of β. This result is obtained when $w = (1 + \sqrt{\beta})^{-2}\beta$, to yield (3.50) and the proof.

∎

To interpret (3.50), cases corresponding to three values of β, are considered:

(i) $\beta \ll 1 (N \text{scalesfasterthan} L)$: When the number of antennas increases at a faster rate than the number of sensors, it can be seen that $B^\infty(\beta)$ is large. When there is no sensing noise, the performance obtained is exactly $B^\infty(\beta)$ as seen in (3.48). In this case, arbitrarily large gains are achievable. In case there is sensing noise in the system, $\mathscr{E}^\infty_{AWGN}$ and $B^\infty(\beta)$ become bounds, and the gain is bounded as shown in (3.50). As $\beta \to 0$ in this case, the bound goes to infinity, which indicates that there could be large gains possible.

(ii) $\beta = 1 (N \text{scalesasfastas} L)$: The number of antennas at the FC and the number of sensors scale at the same rate, the maximum possible gain can be calculated from (3.50) to yield $G^\infty(1) \le 5\zeta$.

(iii) $\beta \gg 1 (N \text{scaleslowerthan} L)$: When the number of sensors scales much faster than the number of antennas at the FC, it resembles the previous setting where

	$K > 0$	$K \to 0$
$G_{\mathrm{N}oCSIS}(N,K)$ from (3.20)	$O(N)$ when $\gamma_c = 0$; $O(1)$ when $\gamma_c > 0$	$O(N)$
$G_{CSIS}(N,K)$ from (3.40)	$O(N)$	$O(1)$
$G^\infty(\beta)$ from (3.50)	Undefined	$O(\beta^{-1})$ as $\beta \to 0$; $O(1)$ as $\beta \to \infty$

Table 3.1: Order of gain due to multiple antennas at the FC for large number of sensors, L.

$L \to \infty$, first, and N was scaled. Not surprisingly, when β is large in this case, $G^\infty(\beta) \le 2\zeta$, same as in Section 3.4.

It should be noted here that in cases (ii) and (iii), where both the number of sensors and antennas are scaled to infinity simultaneously, only limited gain is achievable, when the sensors have complete channel knowledge.

In Table 3.1 we summarize the rate at which the gain due to number of antennas increases, both when CSI is available and unavailable at the sensor side. Recalling that the gain is defined in terms of the ratio of error exponents relative to the single antenna case, all the results in the table apply when L is large, which is a major distinguishing factor between this study and standard analysis of multi-antenna systems. It is seen that when $K > 0$ the gain in error exponent grows like $O(N)$ depending on whether CSIS is available and whether $\gamma_c = 0$. More interestingly, when the channel is zero-mean ($K \to 0$), adding antennas improves the error exponent linearly when CSIS is not available. In stark contrast, when CSIS is available, the gain is bounded ($O(1)$) by 2ζ. Finally, the row on the bottom of Table 3.1 illustrates how the gain depends on the ratio $\beta = L/N$ as both N and L increase. The error exponents for $K > 0$ are infinite, yielding an undefined gain. For zero-mean channels, the dependence on β indicates an increasing gain when β is small ($L \ll N$), and bounded gain when β is large ($L \gg N$).

56

So far, we have provided bounds on the achievable gains due to antennas when CSI is available at the sensors, without providing a realizable scheme. This is because the calculation of α_{OPT} in (3.21) in closed form is intractable. Moreover, it is not clear how α should be chosen as a function of $\mathbf{H}$ when $N > 1$ to achieve a multiple-antenna gain. This is because each sensor sees N channel coefficients, corresponding to N antennas, and each channel coefficient has a different phase making the choices of $\angle\alpha_i$ non-trivial. We now present two sub-optimal schemes for the full CSIS case that are shown to provide gains over the single antenna case.

Method I: Optimizing Gains to Match the Best Antenna

In this method, the sensor gains, α, are selected in order to target the best receive antenna. However, the received signals at all of the other antennas are also combined at the FC, which uses the detection rule defined in (3.8).

Since L is finite for any practical scheme, (3.25) will be used to select α and (3.26) without the limit can be used to assess which antenna has the "best" channel coefficients. Therefore, using the channels from the sensors to all of the receive antennas,

$$n^* = \operatorname*{argmax}_{n} \frac{\theta^2}{8} \frac{1}{L} \sum_{l=1}^{L} \frac{1}{\sigma_\eta^2 + \frac{\sigma_v^2}{P|h_{nl}|^2}}, \tag{3.52}$$

is calculated and the sensor gains are set to (3.25) computed for the channels $\{h_{n^*i}\}_{i=1}^{L}$. The FC then uses all of the receive antennas for detection using (3.8). Since there are multiple antennas at the FC, for any realization of the channels between the sensors and the FC, the error exponent of this scheme is at least as good as the single antenna case.

Such an approach requires the calculation of (3.52) and the corresponding α from (3.25). Since these calculations require the complete knowledge of $\mathbf{H}$, they can be calculated at the FC, and fed back to the sensors.

Method II: Maximum Singular Value of the Channel Matrix

It was shown in Section 3.4 that when $\sigma_\eta^2 = 0$, the bound obtained in (3.36) is achievable. In this method, the values of α are selected as though there is no sensing noise. The sensor gains, α, are selected in such a way that they are a scaled version of the eigenvector corresponding to $\lambda_{max}\left(\mathbf{H}^H\mathbf{H}\right)$, such that $\|\alpha\|^2 = P$. In most practical cases, sensing noise is non-zero, and therefore, this method is sub-optimal. Similar to Method I, α can be calculated at the FC and fed back to the sensors.

Hybrid of Methods I and II

Since Method II is tuned to perform optimally when there is no sensing noise, it outperforms Method I when the sensing SNR, γ_s, is high. As the sensing SNR reduces, Method I begins to outperform Method II. These observations are illustrated and elaborated on in the simulations section (Section 3.5, Figure 3.8).

Since one of the schemes performs better than the other based on the value of γ_s, a hybrid scheme can be used: Method I for low values of γ_s, and Method II for high values. The exact value where the cross-over occurs depends on the parameters of the system, and can determined empirically. An example is shown in the simulation section in Figure 3.8, where it is also argued that an underestimation of the value of γ_s is tolerable, while an overestimation is not.

Semidefinite Relaxation

Following [81, 129] a semidefinite relaxation of the problem in (3.42) is obtained as follows:

$$\mathbf{X}_{\text{PO}} = \underset{\mathbf{X}}{\arg\max}\,\text{trace}(\mathbf{H}^H\mathbf{H}\mathbf{X}) \quad \text{subject to} \quad \mathbf{X} \succeq 0,$$

$$\mathbf{X}_{ii} = \frac{P}{L}, i = 1, 2, \ldots, L, \qquad (3.53)$$

where $\mathbf{X}$ is an $L \times L$ matrix. If $\mathbf{X}$ has a rank-1 decomposition, $\mathbf{X} := \alpha\alpha^H$, then α is a solution to (3.42) [81, 129]. In the more likely case where $\mathbf{X}$ does not have rank-1,

58
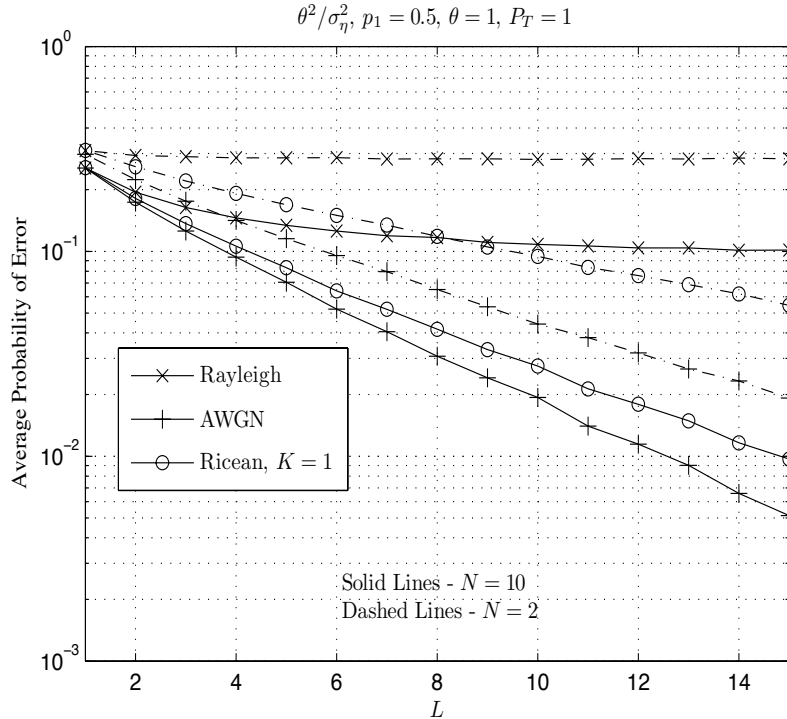
Figure 3.2: Monte-Carlo Simulation: $E[P_{e|\mathbf{H}}(N)]$ for AWGN channels, Rayleigh fading channels and Ricean channels with no CSIS.

then an approximation to the solution of (3.42) is obtained by choosing α as the vector consisting of the phases of the eigenvector corresponding to the maximum eigenvalue of $\mathbf{X}$. The semidefinite relaxation in (3.53) causes a loss of upto a factor of $\pi/4$ in the final answer of (3.42) [129]. The phases of eigenvector corresponding to the maximum eigenvalue of $\mathbf{X}_{\text{P}O}$ are extracted to constitute a possible set of values of α. In order to obtain the solution to the SDR problem, an eigenvalue decomposition of $\mathbf{X}_{\text{O}PT}$ is required, which is an $O(L^3)$ operation [124]. It is argued with the help of simulations (Figure 3.9) that the SDR outperforms the hybrid scheme when γ_s is small, at the expense of increased complexity.

3.5 Simulation Results

The theoretical results obtained are verified using simulations. The channels are generated as complex Gaussian (Rayleigh or Ricean) for the purposes of simulation, even though the results only depend on the first and second order moments of the channels.
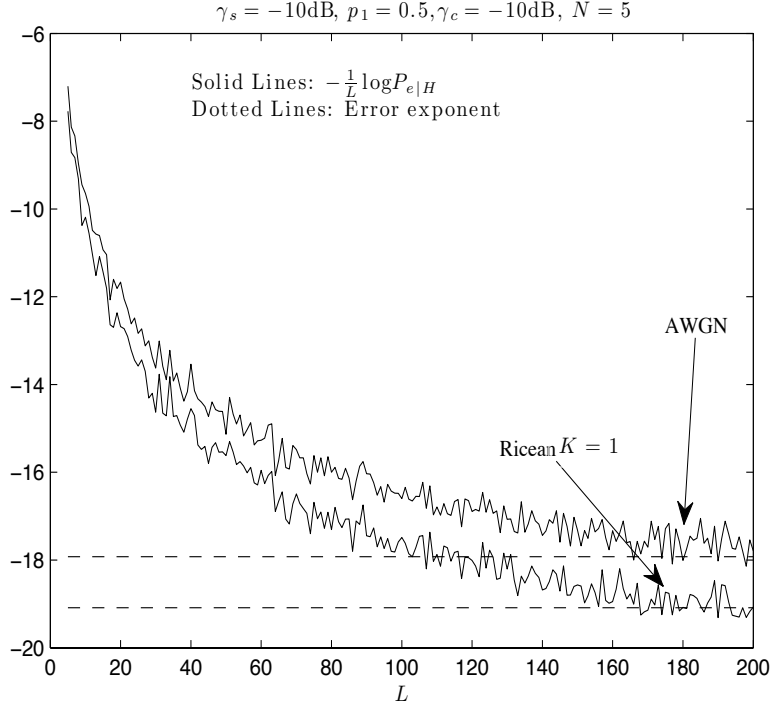
Figure 3.3: Monte-Carlo simulation - Error exponent for AWGN and Ricean Fading channels.

In Figure 3.2, it is verified that increasing the number of sensors improves the performance except when the channels are Rayleigh fading and there is no CSIS. Since the error exponent is zero for the Rayleigh fading case with no CSIS, the asymptotic average probability of error is computed and plotted. The Ricean case outperforms the Rayleigh fading case, and the AWGN channels provide the best performance. It can also be seen that the decay in probability of error is exponential in L, when the channels between the sensors and the FC are AWGN or Ricean fading. The decay is slower than exponential when the channels are Rayleigh fading. This confirms the observations in Section 3.4. In all cases, the performance improves as the number of antennas increases.

In Figure 3.3, the expression of error exponent is compared against the value of $L^{-1} \log P_{e|\mathbf{H}}(5)$ for increasing L, with AWGN channels and Ricean fading channels between the sensors and the FC. It can be seen that fewer than 200 sensors are required for

60
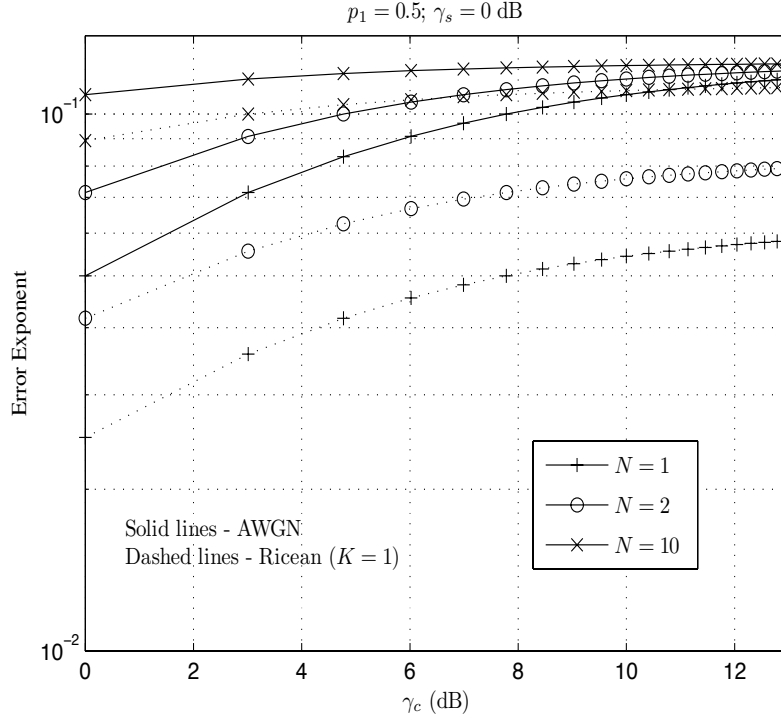
Figure 3.4: Error exponent vs γ_s for $N = 1, 2, 10$ for AWGN channels and Ricean channels and no CSIS.

the asymptotic results to hold. Therefore, in subsequent simulations, $L = 200$ sensors have been used.

The effect of increasing the number of antennas on the error exponent for the AWGN case and Ricean fading case with no CSIS is seen in Figure 3.4. As expected, increasing γ_c improves performance and there is an improvement in performance as the number of antennas at the FC increases. As predicted in Section 3.4, with an increase in N, the performance of $\mathscr{E}_{AWGN}(N)$ and $\mathscr{E}_{NoCSIS}(N,K)$ get closer to each other. There is a large performance gain between the $N = 1$ case and the $N = 2$ case, and almost the same gain between the $N = 2$ case and the $N = 10$ case, indicating diminishing returns, corroborating the results in Section 3.3.

In Figure 3.5, the error exponent is evaluated when there is a single antenna at the FC. The cases of AWGN channels, Ricean channels with no CSIS, Rayleigh fading channels with full CSIS and Rayleigh fading channels with phase-only CSIS are

61
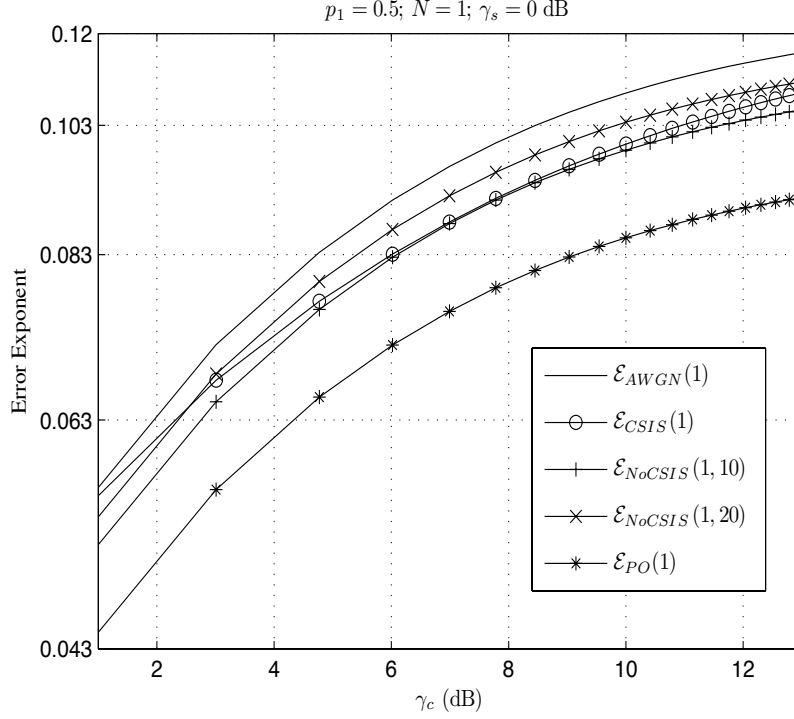
Figure 3.5: Optimal Rayleigh performance, AWGN performance and Ricean no CSIS performance with one antenna at the FC.

compared in Figure 3.5. It is seen that the AWGN performance is the best, and when the Ricean channels have larger line of sight, the performance improves, as expected. In fact, by increasing the amount of LOS, the no-CSIS Ricean case performs better than the full CSIS Rayleigh channel case, when γ_c is large. The performance of the Ricean no CSIS case is a constant factor $K/(K+1)$ worse than the AWGN case, corroborating the result of $\mathcal{E}_{\text{No}CSIS}(1,K)$. Similarly, the performance of the phase-only CSIS case confirms the result in (3.30). For Rayleigh fading channels, the phase-only CSIS case performs a constant $\pi/4$ worse than the AWGN case.

For the case of full CSIS, but with multiple antennas at the FC, bounds were derived on the error exponent of the system in Section 3.4 and Section 3.4, and combined to provide a single bound in (3.37). The value of $\mathcal{E}_{CSIS}(1)$ is set as a lower-bound on $\mathcal{E}_{CSIS}(N)$. In Figure 3.6, with $N=1$, the upper bound can be seen to be about 0.76 dB (in terms of error exponent) away from the actual value at $\gamma_c = 8$ dB. For small values
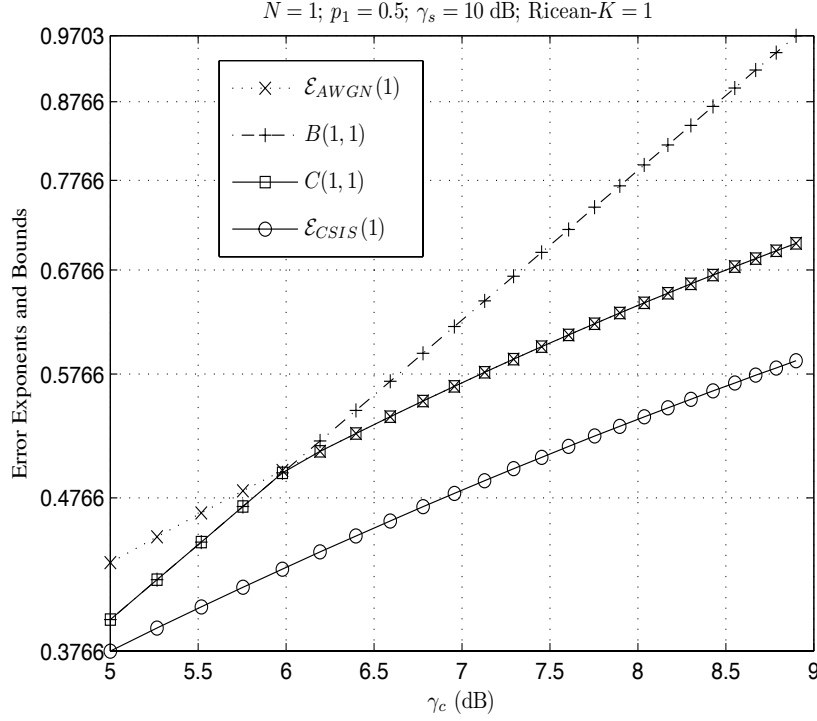
Figure 3.6: For a single antenna, optimal performance and performance bounds.

of γ_c, the AWGN bound is better, and as γ_c increases, the bound with the no sensing noise assumption is better, as expected.

Figure 3.7 shows the effect of increasing the number of antennas at the FC on the antenna gains of the different systems. Also, for the cases of partial CSIS and full CSIS, the upper bounds on the antenna gains are plotted. The actual error exponent for the AWGN case is larger than for the Ricean no-CSIS case. However, as seen in Figure 3.7, the gain for the Ricean no-CSIS case is larger than the gain for the AWGN channel case. The bound on the Ricean CSIS antenna gain grows rapidly with N, as predicted by (3.40). The maximum gains possible for the Rayleigh CSIS case and the Rayleigh no CSIS cases are also plotted. These results indicate that with full CSIS, there is not much to be gained by adding antennas at the FC, corroborating our results in Section 3.4.

The schemes introduced in Section 3.4 for the known CSIS case are simulated
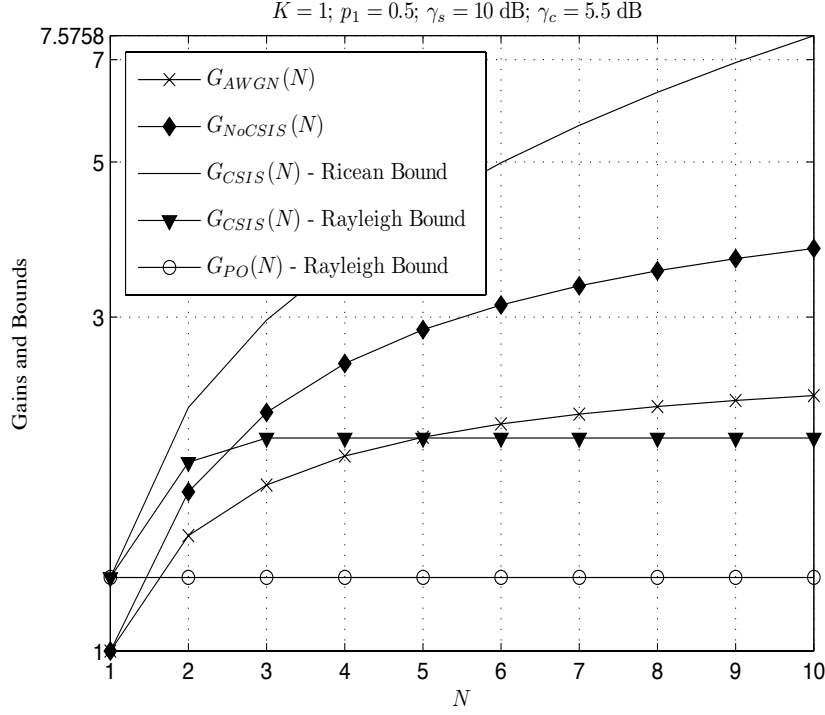
Figure 3.7: Comparison of antenna gains vs N.

in Figure 3.8. The performance of these schemes are evaluated for $N = 5$ and $N = 50$. The performance of these systems is compared against a lower bound given by $\mathscr{E}_{\mathrm{CSIS}}(1)$ from (3.27) and an upper-bound, $C(5,K)$ from (3.37). The hybrid scheme from Section 3.4 selects the better of the two practical methods depending on the value of γ_s. It can be seen that even with these simple sub-optimal practical schemes, the hybrid scheme is always better than $\mathscr{E}_{\mathrm{CSIS}}(1)$, indicating that it is possible to obtain multiple antenna gain. However, for each N, the hybrid scheme does not approach the upper-bound of $C(5,K)$. When $N = 5$, this is an expected result, since firstly, $C(N,K)$ is a bound that is not necessarily achievable, and secondly, the practical schemes are obtained as sub-optimal approximations to the optimal scheme with full CSIS. The hybrid scheme for $N = 50$ provides more gain over $\mathscr{E}_{\mathrm{CSIS}}(1)$ than the hybrid scheme for $N = 5$, but does not beat $C(5,K)$. This means that although gains are possible with the practical schemes, large gains are not possible, as predicted by the bounds in Section 3.4. For
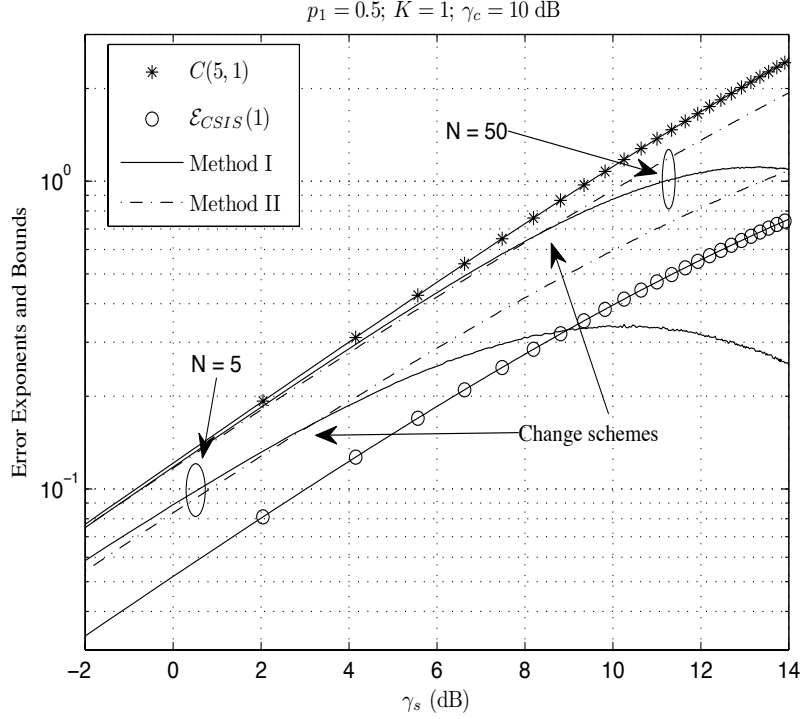
Figure 3.8: Practical Schemes for $N = 5$ and $N = 50$ vs. $\mathscr{E}_{CSIS}(1)$ and $C(5, 1)$.

the hybrid scheme, Method I is better at low values of γ_s and Method II is better at high values of γ_s. The value of γ_s at which the hybrid scheme changes methods can also be seen in the simulations. In Figure 3.8, the system has a channel SNR, $\gamma_c = 10$ dB, $p_1 = 0.5$ and the Ricean-K parameter is one. When there are five antennas at the FC, the hybrid scheme changes from Method I to Method II at $\gamma_s \approx 3$ dB, and when $N = 50$, the change occurs at $\gamma_s \approx 8.25$ dB. It can be seen that the hybrid scheme changes from Method I to Method II at different values of γ_s based on the system parameters. It can also be seen that when Method I is selected by the hybrid scheme, the error in performance between Method I and Method II is small. However, when Method II is selected by the hybrid scheme, the performance gap between Method I and Method II increases rapidly as γ_s increases. Therefore, an underestimation of the value of γ_s is tolerable, while an overestimation is not.

The semidefinite relaxation (SDR) approach in Section 3.4 is compared against
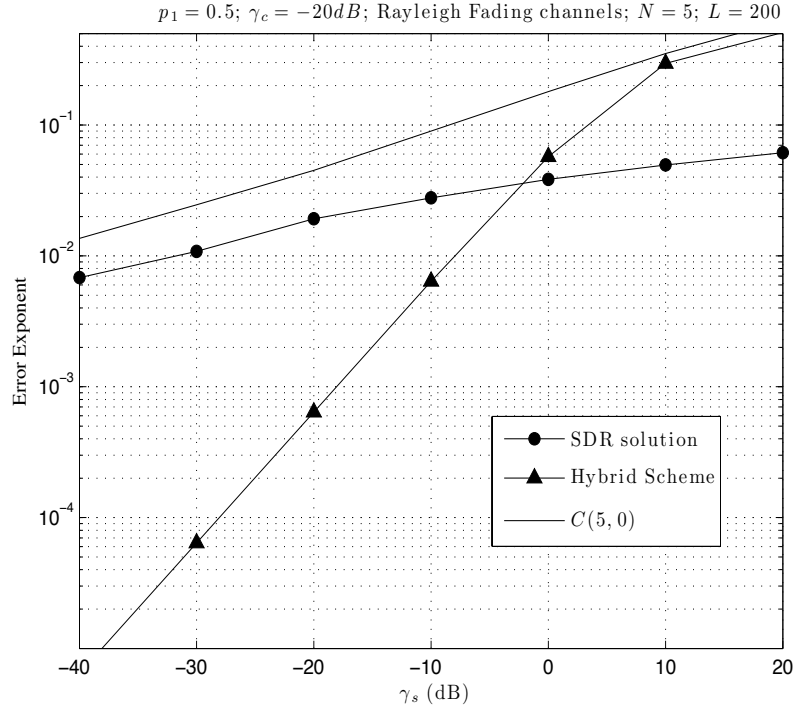
Figure 3.9: Hybrid realizable scheme, SDR relaxation and $C(N, K)$ vs γ_s.

the hybrid scheme (Section 3.4) in Fig 3.9. For the SDR solution, the value of $\mathbf{X}_{OPT}$ from (3.53) is calculated using CVX, a package for specifying and solving convex programs in MATLAB [130]. It can be seen from these simulations that for low values of sensing SNR, γ_s, the SDR solution outperforms the hybrid scheme. However, as the value of γ_s begins to increase, the hybrid scheme (which is designed to be optimal as $\gamma_s \to \infty$) outperforms the SDR solution. The comparison with the upper-bound on the optimal error exponent, $C(N, K)$ is tight with respect to the better of the hybrid and SDR approaches. In order to obtain the solution to the SDR problem, an eigenvalue decomposition of $\mathbf{X}_{OPT}$ is required, which is an $O(L^3)$ operation [124]. The SDR outperforms the hybrid scheme when γ_s is small, at the expense of increased complexity.

3.6 Proof of Theorem 3.4.5

We begin by noting that the presence of correlation in η_l affects the total average transmit power. Therefore, to prove Theorem 3.4.5, we need to reconsider the following in

66

presence of correlation: (i) the power constraint; (ii) the AWGN upper-bound in (3.31); (iii) the "no sensing noise" upper-bound in (3.36), which will then be used to redefine the combined upper-bound in (3.37).

(i) Power constraint: The total transmitted power is given by

$$P_T = E\left[\sum_{l=1}^{L}|\alpha_l\left(\Theta+\eta_l\right)|^2\right] = \alpha^H\left(p_1\theta^2\mathbf{I}_L+\mathbf{R}_\eta\right)\alpha, \tag{3.54}$$

and constrained as

$$\alpha^H\left(p_1\theta^2\mathbf{I}_L+\mathbf{R}_\eta\right)\alpha \leq P_T. \tag{3.55}$$

If (3.55) holds, then

$$\|\alpha\|^2 \leq \frac{P_T}{p_1\theta^2+\lambda_{\min}} := P, \tag{3.56}$$

also holds. Since (3.56) is less stringent than (3.55), if (3.56) is used instead of the original power constraint in (3.55), an upper-bound will be obtained in the subsequent derivation of the error exponent.

(ii) Upper-bound (AWGN channels): Recall that in this case, $\mathbf{H}=1_{N\times L}$. Since the sensing noise is not iid, α has to be selected in such a way that the error exponent is maximized:

$$\underset{\alpha}{\text{maximize}} \quad \alpha^H 1_{L\times N}\mathbf{R}(\alpha)^{-1}1_{N\times L}\alpha \quad \text{subject to} \quad \alpha^H\alpha \leq P, \tag{3.57}$$

to yield the error exponent in the AWGN case with correlated sensing noise:

$$\mathscr{E}_{\text{AWGN}}(N) \leq \frac{1}{L}\frac{\theta^2}{8}(\alpha_{\text{AWGN}}^{\text{opt}})^H 1_{L\times N}\mathbf{R}(\alpha_{\text{AWGN}}^{\text{opt}})^{-1}1_{N\times L}\alpha_{\text{AWGN}}^{\text{opt}}, \tag{3.58}$$

where $\alpha_{\text{AWGN}}^{\text{opt}}$ provides to solution to (3.57) and the inequality in (3.58) is due to (3.11) and the modified power constraint in (3.56). To fully compute an upper bound on the right hand side of (3.58), first, $\mathbf{R}(\alpha)$ is inverted and simplified. For the case of correlated noise, $\mathbf{R}(\alpha)$ is given by

$$\mathbf{R}(\alpha) = 1_{N\times L}\mathbf{D}(\alpha)\mathbf{R}_\eta\mathbf{D}(\alpha)^H 1_{L\times N}+\sigma_v^2\mathbf{I}_N$$

$$\succeq \lambda_{\min}1_{N\times L}\mathbf{D}(\alpha)^H\mathbf{D}(\alpha)1_{L\times N}+\sigma_v^2\mathbf{I}_N, \tag{3.59}$$

where $\mathbf{A} \succeq \mathbf{B}$ indicates that the matrix $(\mathbf{A} - \mathbf{B})$ is positive semi-definite. Using the Sherman-Morrison-Woodbury formula for matrix inversion,

$$\mathbf{R}(\alpha)^{-1} \preceq \frac{1}{\sigma_v^2}\mathbf{I}_N - \frac{1}{\sigma_v^2}\mathbf{1}_{N \times L}\left[\mathrm{diag}\left(\frac{1}{\lambda_{\min}|\alpha_i|^2}\right) + \frac{\mathbf{1}_{L \times N}\mathbf{1}_{N \times L}}{\sigma_v^2}\right]^{-1}\mathbf{1}_{L \times N}\frac{1}{\sigma_v^2}. \quad (3.60)$$

Invoking the Sherman-Morrison-Woodbury formula for matrix inversion once again,

$$\mathbf{R}(\alpha)^{-1} \preceq \frac{1}{\sigma_v^2}\mathbf{I}_N - \frac{1}{\sigma_v^2}M\mathbf{1}_{N \times N}, \quad (3.61)$$

where

$$M := \frac{\sum_{l=1}^{L}\frac{\lambda_{\min}}{\sigma_v^2}|\alpha_l|^2}{1 + N\sum_{i=1}^{L}\frac{\lambda_{\min}}{\sigma_v^2}|\alpha_i|^2} \leq \frac{\lambda_{\min}P}{N\lambda_{\min}P + \sigma_v^2}, \quad (3.62)$$

due to the fact that $\alpha^H\alpha \leq P$ from (3.56).

By substituting (3.61) in (3.57), the solution to (3.57) is upper-bounded by the solution to

$$\underset{\alpha}{\text{maximize}} \quad \alpha^H\mathbf{1}_{L \times L}\alpha \quad \text{subject to} \quad \alpha^H\alpha \leq P. \quad (3.63)$$

The value of α that maximizes (3.63) is the eigenvector corresponding to the maximum eigenvalue of $\mathbf{1}_{L \times L}$, scaled to satisfy the constraint with equality. Substituting this in (3.58), the bound in (3.31) obtained, with the substitution, $\gamma_s = \tilde{\gamma}_s$, where $\tilde{\gamma}_s = \theta^2/\lambda_{\min}$ and $P_T \leq P/(p_1\theta^2 + \lambda_{\min})$.

(iii) Upper-bound (no sensing noise): With no sensing noise, $\mathbf{R}_\eta = \mathbf{0}_{L \times L}$. The optimization problem to obtain the best error exponent is the same as in (3.32), to yield (3.36).

Combining the modified AWGN upper-bound and the no sensing noise upper-bound in (3.36), a joint upper-bound is obtained, which is identical to (3.37), except for the substitution $\sigma_\eta^2 = \lambda_{\min}$ and $\tilde{\gamma}_s = \theta^2/\lambda_{\min}$. It follows that (3.39) holds with $z = \gamma_c/(p_1\tilde{\gamma}_s + 1)$, to provide the proof.

■

Chapter 4

INEQUALITIES RELATING THE CHARACTERISTIC FUNCTION AND FISHER

INFORMATION

4.1 Problem Summary

We investigate the relationship between the Fisher information about a location param-

eter and the characteristic function of the additive noise by providing a new derivation

for two inequalities that involve the Fisher information and the characteristic function.

These inequalities were originally derived using a different approach and applied in a

quantum physics setting to estimate the survival probability of a quantum state in [131].

Conditions for equality are also delineated herein for the first time in the literature, and

used to investigate the asymptotic efficiency of a distributed estimation scheme over a

Gaussian multiple-access channel.

4.2 The Inequalities

Consider a model where a deterministic location parameter, θ, is related to observa-

tions $x_l = \theta + \eta_l$, $l = 1, \ldots, L$, where η_l are iid and real-valued random variables. Let

the characteristic function of η_l be $\varphi(\omega) := E[e^{j\omega\eta_l}]$ and let the Fisher information be

defined as [55, 132]

$$I(\eta) := \int_{-\infty}^{\infty} \frac{[p'(x)]^2}{p(x)} dx < \infty, \tag{4.1}$$

where $p(x)$ is the pdf of η_l, assumed to be continuously differentiable, and with support

$(-\infty, \infty)$. Note that $I(\eta)$ is the Fisher information in x_l about θ, and is a deterministic

value which does not depend on θ. In the following, η denotes a random variable with

the same distribution as any η_l.

We present the following theorem, which provides two bounds involving $I(\eta)$

and $\varphi(\omega)$. It was proved first in [131] using the Cramér-Rao inequality. We provide an

alternate proof which also delineates the condition for equality for the first time in the

literature. The condition for equality will be central in Section 4.3 to establish neces-

sary and sufficient conditions for the asymptotic efficiency of a distributed estimation

69

algorithm over a Gaussian multiple-access channel.

Theorem 4.2.1 *Let $\varphi_R(\omega)$ and $\varphi_I(\omega)$ be the real and the imaginary parts of $\varphi(\omega)$, respectively. We have*

$$\omega^2 \varphi_I^2(\omega) \leq I(\eta) \left[\frac{1}{2}[1 + \varphi_R(\omega)] - \varphi_R^2(\omega) \right], \tag{4.2}$$

$$\omega^2 \varphi_R^2(\omega) \leq I(\eta) \left[\frac{1}{2}[1 - \varphi_R(\omega)] - \varphi_I^2(\omega) \right], \tag{4.3}$$

with equality in both (4.2) and (4.3) if and only if $\omega = 0$.

Proof Let $s(x) := p'(x)/p(x)$ be the score function, where we recall that $p(x)$ is the pdf of η_l. Let $g(x)$ be a differentiable function satisfying $\lim_{x \to \pm\infty} g(x)p(x) = 0$. Using Stein's identity [133, Lemma 1.18], we have

$$E[g(\eta)s(\eta)] = -E[g'(\eta)]. \tag{4.4}$$

Applying the Cauchy-Schwarz inequality yields

$$E^2[g'(n)] \leq I(\eta)E[g^2(\eta)], \tag{4.5}$$

with equality if and only if $s(x) = \alpha g(x)$ for some α and all x. It can be seen that by substituting $g_1(x) := \cos(\omega x) - \varphi_R(\omega)$ for $g(x)$ in (4.5), equation (4.2) is obtained. Similarly, $g_2(x) := \sin(\omega x) - \varphi_I(\omega)$ substituted for $g(x)$ yields equation (4.3).

To examine when equality occurs, first note that if $\omega = 0$, since $\varphi_R(0) = 1$ and $\varphi_I(0) = 0$, equations (4.2) and (4.3) become equalities. Conversely, consider $\omega \neq 0$. The equality condition for (4.3) is $s(x) = \alpha g_2(x)$, which yields the first order differential equation

$$\frac{p'(x)}{p(x)} = \alpha [\sin(\omega x) - \varphi_I(\omega)], \tag{4.6}$$

which must provide a solution satisfying $p(x) \geq 0$ and $\int_{-\infty}^{\infty} p(x)dx = 1$. The solution to (4.6) is of the form $p(x) = Ce^{-\alpha x \varphi_I(\omega)}e^{-\frac{\alpha}{\omega}\cos(\omega x)}$, which is unbounded as $x \to -\infty$ when $\varphi_I(\omega) \neq 0$, and periodic when $\varphi_I(\omega) = 0$. In either case, $\int_{-\infty}^{\infty} p(x)dx = 1$ is not

70
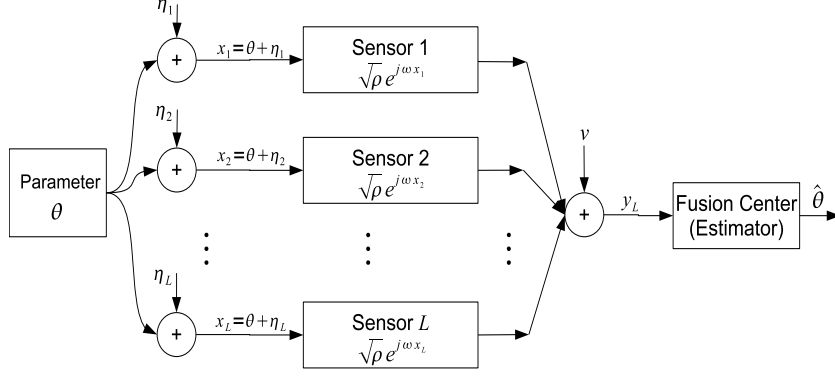
Figure 4.1: System model: Wireless sensor network. The estimator is located at the fusion center.

possible. This shows that there is no pdf satisfying (4.6) when $\omega \neq 0$, and therefore, equality in (4.3) cannot be attained for $\omega \neq 0$. The same conclusion can be drawn about equation (4.2), using a similar argument with $s(x) = \alpha g_1(x)$.

$\blacksquare$

4.3   Application to Distributed Estimation

A sensor network, illustrated in Figure 4.1, consisting of L sensors is considered. The value, x_l, observed at the lth sensor is

$$x_l = \theta + \eta_l \tag{4.7}$$

for $l = 1, ..., L$, where θ is a deterministic, real-valued, unknown parameter in a bounded interval of known length, $[0, \theta_R]$, where $\theta_R < \infty$, and η_l are iid real-valued random variables. We will assume that η_l has zero mean and variance σ_η^2, when the mean and variance exist. Due to constraints in the transmit power, we consider a scheme where the lth sensor transmits its measurement, x_l, using a constant modulus base-band equivalent signal, $\sqrt{\rho} e^{j\omega x_l}$, over a Gaussian multiple access channel so that the received signal at the fusion center is given by

$$y_L = \sqrt{\rho} \sum_{l=1}^{L} e^{j\omega x_l} + v, \tag{4.8}$$

where the transmitted signal at each sensor has per-sensor power of ρ, $\omega \in (0, 2\pi/\theta_R]$ is a design parameter to be optimized, and $v \sim \mathscr{CN}(0, \sigma_v^2)$ is independent of $\{\eta_l\}_{l=1}^{L}$.

71

Note that the restriction $\omega \in (0, 2\pi/\theta_R]$ is necessary even in the absence of sensing and channel noise $(y_L = \sqrt{\rho}e^{j\omega\theta})$ to uniquely determine θ from y_L.

In a centralized problem, θ is estimated from $\{x_l\}_{l=1}^L$. The Cramér-Rao bound is the well known benchmark on the variance of unbiased estimators with finite samples and is proportional to $[I(\eta)]^{-1}$ [134, pp. 120]. For large L, the asymptotic variance is an appropriate performance metric. Under certain regularity conditions, the benchmark on the asymptotic variance is given by $[I(\eta)]^{-1}$ [134, pp. 439]. Hence, the Fisher information has a central role to play in establishing benchmarks for the estimation of a location parameter for centralized estimation problems which address estimators of θ based on $\{x_l\}_{l=1}^L$.

For the distributed setting, based on (4.8), the estimators of θ rely on y_L. The desire to have constant modulus transmissions over a Gaussian multiple-access channel causes the fusion center in Figure 4.1 to have access to only y_L, rather than $\{x_l\}_{l=1}^L$. Clearly, y_L has less information about θ than $\{x_l\}_{l=1}^L$. In what follows, we quantify this loss by examining the efficiency of the minimum (asymptotic) variance estimator, and comparing it with the benchmark for the centralized problem, $[I(\eta)]^{-1}$, for different distributions on the sensing noise, η. Using Theorem 4.2.1, it is shown that there is no loss in efficiency if and only if η is Gaussian.

The Estimator

To estimate θ, we normalize y_L in (4.8) and define:

$$z_L := \frac{y_L}{L} = \sqrt{\rho}e^{j\omega\theta}\frac{1}{L}\sum_{l=1}^L e^{j\omega\eta_l} + \frac{v}{L}, \tag{4.9}$$

where $z_L = |z_L|\exp(j\angle z_L) = z_L^R + jz_L^I$, and z_L^R and z_L^I are the real and imaginary parts of z_L, respectively. Also $\mathbf{z}_L := [z_L^R z_L^I]^T$ and $\bar{\mathbf{z}}(\theta) := [E[z_L^R]E[z_L^I]]^T = \sqrt{\rho}[\varphi_R(\omega)\cos\omega\theta - \varphi_I(\omega)\sin\omega\theta \quad \varphi_R(\omega)\sin\omega\theta + \varphi_I(\omega)\sin\omega\theta]^T$.

Given y_L (or equivalently $\mathbf{z}_L$), the estimator with the smallest asymptotic vari-

72

ance is given by [115, (3.6.2), pp. 82]

$$\hat{\theta}_L = \underset{\theta}{\operatorname{argmin}}[\mathbf{z}_L - \bar{\mathbf{z}}(\theta)]\Sigma^{-1}(\theta)[\mathbf{z}_L - \bar{\mathbf{z}}(\theta)]^T, \tag{4.10}$$

where

$$\Sigma(\theta) = \begin{bmatrix} \Sigma_{11}(\theta) & \Sigma_{12}(\theta) \\ \Sigma_{21}(\theta) & \Sigma_{22}(\theta) \end{bmatrix} \tag{4.11}$$

is the 2×2 asymptotic covariance matrix of $\mathbf{z}_L$, satisfying $\lim_{L\to\infty}\sqrt{L}[\mathbf{z}_L - \bar{\mathbf{z}}(\theta)] = \mathcal{N}(\mathbf{0}, \Sigma(\theta))$. Its elements are given by

$$\Sigma_{11}(\theta) = \rho\left[v_c\cos^2(\omega\theta) + v_s\sin^2(\omega\theta)\right]$$

$$\Sigma_{22}(\theta) = \rho\left[v_s\cos^2(\omega\theta) + v_c\sin^2(\omega\theta)\right]$$

$$\Sigma_{12}(\theta) = \Sigma_{21}(\theta) = \rho(v_c - v_s)\sin(\omega\theta)\cos(\omega\theta),$$

where $v_c := \operatorname{var}[\cos(\omega\eta_l)] = 1/2 + \varphi_R(2\omega)/2 - \varphi_R^2(\omega)$ and $v_s := \operatorname{var}[\sin(\omega\eta_l)] = 1/2 - \varphi_R(2\omega)/2 - \varphi_I^2(\omega)$.

Estimators of the form in (4.10) have an asymptotic variance given by [115, Lemma 3.1]

$$\operatorname{AsV}(\omega) = \left[\left(\frac{\partial\bar{\mathbf{z}}(\theta)}{\partial\theta}\right)^T\Sigma^{-1}(\theta)\left(\frac{\partial\bar{\mathbf{z}}(\theta)}{\partial\theta}\right)\right]^{-1}. \tag{4.12}$$

It can be seen that by substituting the values of $\partial\bar{\mathbf{z}}(\theta)/\partial\theta = \sqrt{\rho}\omega[-\varphi_R(\omega)\sin\omega\theta - \varphi_I(\omega)\cos\omega\theta \quad \varphi_R(\omega)\cos\omega\theta - \varphi_I(\omega)\sin\omega\theta]^T$ and $\Sigma^{-1}(\theta)$, whose elements can be expressed in terms of $\Sigma_{11}(\theta)$, $\Sigma_{22}(\theta)$ and $\Sigma_{12}(\theta)$, the asymptotic variance is given by

$$\begin{aligned}
\operatorname{AsV}(\omega) &= \frac{2v_cv_s}{\omega^2\left[v_s\varphi_I^2(\omega) + v_c\varphi_R^2(\omega)\right]} \\
&= \frac{\left(1 + \varphi_R(2\omega) - 2\varphi_R^2(\omega)\right)\left(1 - \varphi_R(2\omega) - 2\varphi_I^2(\omega)\right)}{\omega^2\left[\varphi_R^2(\omega)\left(1 + \varphi_R(2\omega) - 2\varphi_R^2(\omega)\right) + \varphi_I^2(\omega)\left(1 - \varphi_R(2\omega) - 2\varphi_I^2(\omega)\right)\right]}.
\end{aligned} \tag{4.13}$$

Note that $\operatorname{AsV}(\omega)$ depends on the sensing noise through its characteristic function, and does not depend on the channel noise variance, σ_v^2, which washes out for large L.

73

We now address the asymptotic efficiency of $\hat{\theta}_L$ and characterize the condition under which $\mathrm{AsV}(\omega)$ can be made arbitrarily close to $[I(\eta)]^{-1}$:

Theorem 4.3.1 *The estimator in (4.10) can be arbitrarily close to being asymptotically efficient by the proper choice of ω, that is,*

$$\inf_{\omega \in (0, 2\pi/\theta_R]} \mathrm{AsV}(\omega) = \frac{1}{I(\eta)}, \tag{4.14}$$

if and only if η is Gaussian.

Proof We begin by showing that if (4.14) holds, then η is Gaussian. Using Theorem 4.2.1, the inequalities in (4.2) and (4.3) can be rewritten for $\omega > 0$ as

$$\frac{\omega^2 \varphi_I^2(\omega)}{\left(\frac{1}{2}[1 + \varphi_R(\omega)] - \varphi_R^2(\omega)\right)} < I(\eta), \tag{4.15}$$

$$\frac{\omega^2 \varphi_R^2(\omega)}{\left(\frac{1}{2}[1 - \varphi_R(\omega)] - \varphi_I^2(\omega)\right)} < I(\eta), \tag{4.16}$$

where we use that when $\omega \neq 0$, (4.2) and (4.3) are strict inequalities. Adding the inequalities in (4.15) and (4.16), rearranging the resulting inequality and recalling (4.13), we have

$$\frac{1}{I(\eta)} < \mathrm{AsV}(\omega), \qquad \omega \in (0, 2\pi/\theta_R]. \tag{4.17}$$

Equation (4.17) indicates that the infimum in (4.14) is not attained for any non-zero finite value of ω. Since ω is bounded above, the only way for (4.14) to hold is when $\lim_{\omega \to 0} \mathrm{AsV}(\omega) = [I(\eta)]^{-1}$. It is easy to verify, using L'Hospital's rule, that $\lim_{\omega \to 0} \mathrm{AsV}(\omega) = \sigma_\eta^2$, the variance of η_l. Therefore, for (4.14) to hold, we have $[I(\eta)]^{-1} = \sigma_\eta^2$. The only distribution that satisfies this is the Gaussian [133, Lemma 1.19]. This completes the proof of the first half.

To show that (4.14) holds when η_l is Gaussian, $\varphi(\omega) = e^{-\omega^2 \sigma_\eta^2/2}$ is substituted into (4.13) to yield:

$$\mathrm{AsV}(\omega) = \frac{1}{\omega^2} e^{-\sigma_\eta^2 \omega^2} \left(e^{2\sigma_\eta^2 \omega^2} - 1\right)^2, \tag{4.18}$$

74

which is non-decreasing in ω, since

$$\frac{\partial \text{AsV}(\omega)}{\partial \omega} = \frac{2e^{-2\sigma_\eta^2 \omega^2}}{\omega^3} \left(e^{2\sigma_\eta^2 \omega^2} - 1\right) \left[(1 - e^{2\sigma_\eta^2 \omega^2}) + 2\sigma_\eta^2 \omega^2 + 2\sigma_\eta^2 \omega^2 e^{2\sigma_\eta^2 \omega^2}\right] \geq 0,$$

(4.19)

for $\omega > 0$.

∎

The phase modulated scheme considered here has the advantage of constant modulus transmissions. Due to the use of phase modulation, the result in Theorem 4.3.1 is related to the efficiency of the estimator of a location parameter using the empirical characteristic function (ECF), defined as $\hat{\varphi}(\omega) := L^{-1} \sum_{l=1}^L e^{j\omega x_l}$. It can be seen from (4.9) that $z_L = \sqrt{\rho} e^{j\omega\theta} \hat{\varphi}(\omega) + v/L$ is related to the ECF through scaling and additive noise. The efficiency of empirical characteristic function based estimators has been considered for arbitrary parameters (that is, not just location parameters) in [112], but with a continuum of infinitely many values of the argument, ω, of the ECF. In the current distributed estimation application, the evaluation of $\hat{\varphi}(\omega)$ for many values of ω at the fusion center corresponds to many transmissions per sensor observation, requiring large bandwidth. In contrast, we consider a single value of ω for estimation, requiring a single transmission per sensor. The analog transmissions are assumed to be appropriately pulse-shaped and phase modulated to consume finite bandwidth.

When the sensing noise distribution is symmetric, the cost function on the right hand side of (4.10) that needs to be minimized can be expressed as

$$\begin{aligned}
c(\theta) =& [\mathbf{z}_L - \bar{\mathbf{z}}(\theta)]\Sigma^{-1}(\theta)[\mathbf{z}_L - \bar{\mathbf{z}}(\theta)]^T \\
=& \frac{1}{2\rho^2 v_c v_s} \Big[-4\rho^{3/2} v_s \varphi(\omega)[z_L^I \sin(\omega\theta) + z_L^R \cos(\omega\theta)] + 2\rho^2 v_s \varphi^2(\omega) \\
& + \rho(v_c - v_s)\left((z_L^I)^2 - (z_L^R)^2\right)\cos(2\omega\theta) - 2\rho(v_c + v_s)z_L^I z_L^R \sin(2\omega\theta) \\
& + \rho(v_c + v_s)\left((z_L^I)^2 + (z_L^R)^2\right) \Big].
\end{aligned}$$

(4.20)

Differentiating with respect to θ, we have

$$\frac{\partial c(\theta)}{\partial \theta} = \frac{2\omega z_L^R \cos(\omega\theta)}{\rho v_c v_s}\left[\frac{z_L^I}{z_L^R} - \tan(\omega\theta)\right]$$
$$\times \left[\left(1 + \frac{z_L^I}{z_L^R}\tan(\omega\theta)\right)v_c + \left(1 - \frac{\sqrt{\rho}\varphi(\omega)}{z_L^R \cos(\omega\theta)}\frac{z_L^I}{z_L^R}\tan(\omega\theta)\right)v_s\right]. \qquad (4.21)$$

The values of θ at which (4.21) is zero are given by

$$\theta \in \left\{\frac{n\pi \pm \frac{\pi}{2}}{\omega}, \frac{1}{\omega}\angle z_L, \frac{\angle z_L + 2n\pi \pm \frac{\pi}{2}}{\omega}\right\}, \qquad (4.22)$$

where $\omega \neq 0$ and $n \in \mathbb{Z}^+$. The value of θ that minimizes $c(\theta)$ is easily verified by substituting the values of θ from (4.22) into (4.20) and is given by

$$\hat{\theta} = \frac{1}{\omega}\angle z_L. \qquad (4.23)$$

Hence, in the presence of symmetric noise, the estimator in (4.10) that minimizes the asymptotic variance reduces to the simple expression in (4.23), which was first considered in [135]. However, in [135], neither the optimality (in terms of minimizing the asymptotic variance) nor the efficiency of the estimator in (4.23) was considered.

Quantifying Relative Efficiency

One way of interpreting Theorem 4.3.1 is to observe that when the sensing noise is Gaussian, no information is lost by analog phase modulation if ω is chosen sufficiently small. On the other hand, information is lost when the sensing noise follows other distributions. To see this more clearly, we define the relative efficiency between the asymptotic variance and the Fisher information as:

$$\mathscr{E}(\eta) = \left[I(\eta)\inf_{\omega\in(0,2\pi/\theta_R]}\text{AsV}(\omega)\right]^{-1}. \qquad (4.24)$$

It can easily be verified that $\mathscr{E}(\eta)$ is scale-invariant in the sense that $\mathscr{E}(\alpha\eta) = \mathscr{E}(\eta)$ for any $\alpha \in \mathbb{R}$. Moreover, based on Theorem 4.3.1 and (4.17), $0 \leq \mathscr{E}(\eta) \leq 1$, where the equality in the upper-bound is achieved only if η is Gaussian.

The relative efficiency in (4.24) depends only on the distribution of the sensing noise. The values of $\mathscr{E}(\eta)$ for several distributions are provided in Table 4.1. The result in Table 4.1 for the Gaussian case has been established in Theorem 4.3.1. For the

Distribution	Gaussian	Laplace	Cauchy	Uniform
$\mathscr{E}(\eta)$	1	$2/3$	$0.5c^2e^{-c}(1-e^{-c})^{-1} \approx 0.65$	0

Table 4.1: $\mathscr{E}(\eta)$ for different distributions.

Laplace sensing noise, $\varphi(\omega) = (1 + \omega^2\sigma_\eta^2/2)^{-1}$, $\text{AsV}(\omega) = \sigma_\eta^2(1 + \sigma_\eta^2\omega^2/2)/(1 + 2\sigma_\eta^2\omega^2)$, and $\inf_{\omega \in (0, 2\pi/\theta_R]} \text{AsV}(\omega) = 3\sigma_\eta^2/4$, by inspecting the third derivative of $\text{AsV}(\omega)$. Similarly for the case of Cauchy distribution, $\varphi(\omega) = e^{-\gamma\omega}$, $\text{AsV}(\omega) = e^{2\gamma\omega}(1 - e^{-2\gamma\omega})/2\omega^2$, and $\inf_{\omega \in (0, 2\pi/\theta_R]} \text{AsV}(\omega) = 4\gamma^2e^c(1 - e^{-2c})/c^2$, by examining the first derivative of $\text{AsV}(\omega)$ where γ is defined as the scale parameter of the Cauchy random variable, $c := 2 + W(-2e^{-2})$, and $W(\cdot)$ is the Lambert W-function [136]. For the uniform distribution, an extension of the definition in (4.1) can be used to argue that the Fisher information is infinite [134, pp. 119], and the relative efficiency of the estimator as defined in (4.24) is zero.

We have seen that the Gaussian sensing noise is the only distribution with the highest possible efficiency when the observations x_l are transmitted with phase modulation over Gaussian multiple-access channels and the estimator in (4.10) is used. However, it is possible that other sensing noise distributions, which yield less efficiency, have better asymptotic variances. This is because efficiency is defined relative to the Fisher information. For example, for Laplace sensing noise, the proposed estimator is not asymptotically efficient, but has better asymptotic variance than in the Gaussian case, since its inverse Fisher information, $[I(\eta)]^{-1}$, is lower. In conclusion, Gaussian sensing noise has the only distribution that does not suffer a loss in efficiency when the sensed data x_l is mapped to constant modulus transmissions over Gaussian multiple-access channels.

4.4 Numerical Results

In Figures 4.2 and 4.3, the asymptotic variance and the value of $[I(\eta)]^{-1}$ in dB are plotted versus ω, when the sensing noise is Gaussian, Laplace, uniform and Cauchy distributed.
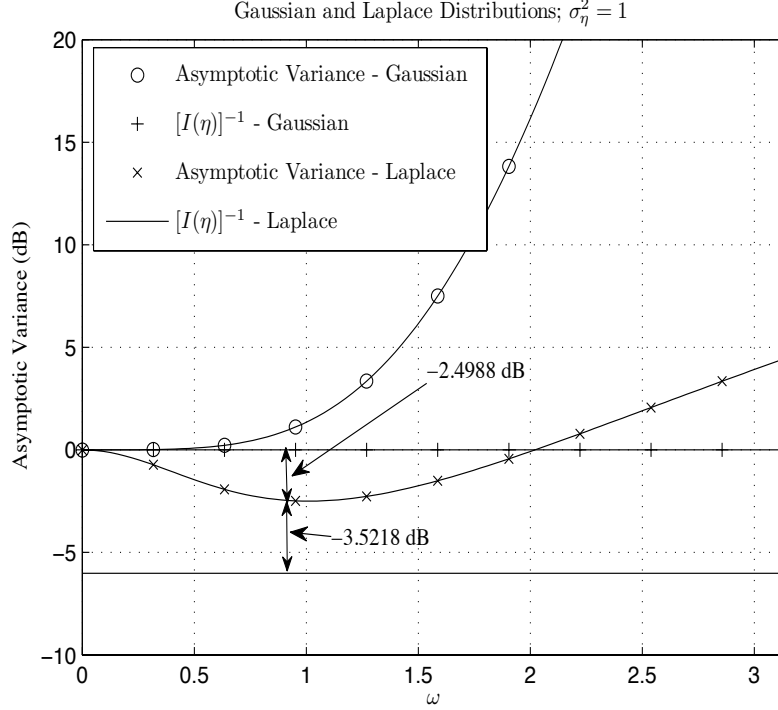
Figure 4.2: Plot of asymptotic variance vs. ω.

From Figure 4.2, the asymptotic variance approaches $[I(\eta)]^{-1}$ only as $\omega \to 0$ for Gaussian sensing noise, and is bounded away from $[I(\eta)]^{-1}$ for other values of ω. The estimator in (4.10) is not efficient when the sensing noise is non-Gaussian. Using the definition of relative efficiency in (4.24), it can seen from Figure 4.2 that $\mathscr{E}(\eta)$ in the case of Gaussian sensing noise is 0dB, and in the case of Laplace sensing noise is about -3.5dB. In Figure 4.2, it can be verified that $\inf_{\omega} \mathrm{AsV}(\omega) \approx 0.75$, which is about -2.5dB at $\omega = 1/\sqrt{2}$, which is lower than the Gaussian sensing noise case.

From Figure 4.3 the relative efficiency for Cauchy noise is about -3.8dB, verifying the value shown in Table 4.1. The inverse Fisher information for the uniform case is 0 ($-\infty$ dB) and is not shown in Figure 4.3. The relative efficiency as defined in (4.24), for uniform noise, is therefore zero. When the sensing noise follows the Cauchy, uniform or Laplace distributions, the estimator is not asymptotically efficient.
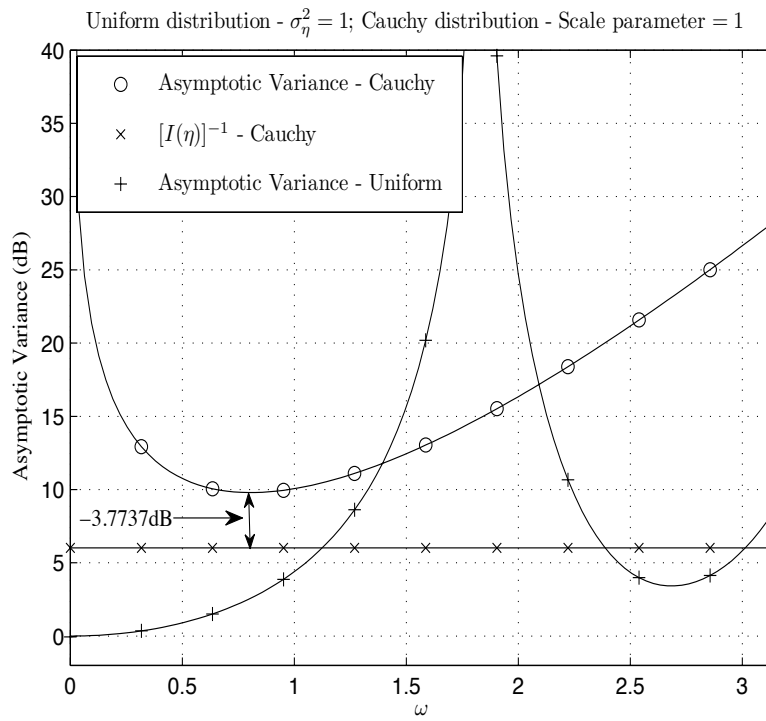
Figure 4.3: Plot of asymptotic variance vs. ω. Note that the value of $[I(\eta)]^{-1}$ is 0 $(-\infty$ dB) for the uniform sensing noise case and is not shown.

Chapter 5

DISTRIBUTED VARIANCE AND SNR ESTIMATION USING CONSTANT

MODULUS SIGNALING OVER GAUSSIAN MULTIPLE-ACCESS CHANNELS

5.1 Problem Summary

In this chapter, the location and scale parameter of a signal embedded in noise are estimated in a distributed fashion. Several sensors are exposed to a signal in (not necessarily Gaussian) noise as seen in Figure 1. These sensors phase modulate the observations using a constant-modulus scheme and transmit these signals to a fusion center (FC) over a Gaussian multiple-access channel [55, pp. 378]. Due to the additive nature of the multiple-access channel, the signals transmitted from the sensors add and approximate the characteristic function of the signal and noise, as the number of sensors increases. At the FC, a noisy version of this empirical characteristic function s received in Gaussian noise, and the location and scale parameter are estimated from this value. All sensors transmit using the same single value of ω, the parameter of the characteristic function. The value of ω is a design parameter in the phase-modulation scheme and is determined based on performance measures. A single transmission from each sensor to the FC is used for the estimation of the location parameter and the scale parameter. A single snapshot in time is sufficient for the estimation.

Once the signal is received at the FC, a minimum-variance estimator is used to jointly estimate the location and scale parameters. Additionally, from the structure of the characteristic functions, naive estimators are developed for each distribution. The performance of the estimates are measured using the asymptotic covariance matrix of the estimates. The location and scale parameter estimates are used to construct the estimate for SNR, and the asymptotic variance of the SNR estimator is also computed.

In contrast to the distributed estimation framework considered in this work, in centralized estimation, the observations of the signals embedded in noise are directly available to the estimator [109–114]. In [110–113], the location and scale parame-
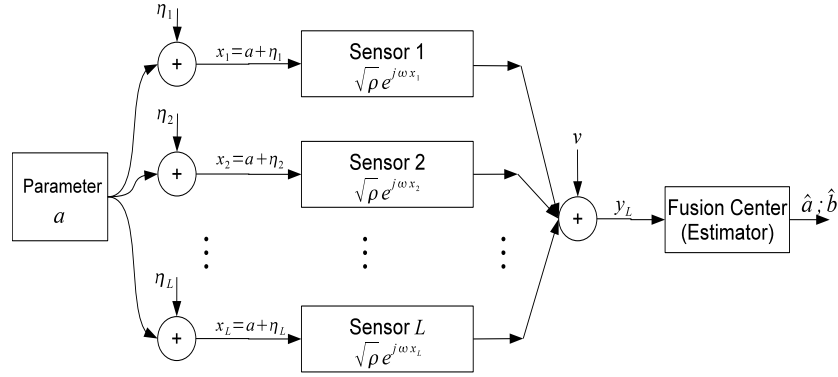
Figure 5.1: System model: Wireless sensor network with constant modulus transmissions from the sensors. The estimator is located at the fusion center.

ters are separately estimated from the characteristic function of the signal embedded in noise. It is also assumed in these works that the estimator has full access to a continuum of infinitely many values of the argument of the characteristic function, ω. In the current distributed estimation application, the evaluation of the characteristic function for many values of ω at the fusion center corresponds to many transmissions per sensor observation, requiring large bandwidth. In contrast, in this framework, all sensors transmit using a single value of ω, indicating limited bandwidth requirements.

In this chapter, distributed estimation of the location parameter and scale parameter of a random signal is performed. In contrast to [109–114], where centralized estimation is used to find the SNR, a distributed framework is used. In order to conserve bandwidth, a single value of ω is used for transmissions by all the sensors. In contrast to [61], where the estimation is performed independently at each sensor, due to the phase modulation used here, a single transmission from each sensor is enough for successful estimation at the FC. At the FC, the location parameter and the scale parameter are simultaneously estimated, using a minimum-variance estimator, and a naive estimator based on the structure of the characteristic function of each noise distribution. It is shown that the estimates of the location parameter and the scale parameter are independent of each other in all cases. The naive estimators have the same performance as the minimum-variance estimator, but with lower complexity. In each case, the values

of the ω that minimize the asymptotic variances of the location parameter and the scale parameter are also calculated. It is also shown with the help of simulations that the estimators are asymptotically efficient only if the noise distribution is Gaussian.

5.2   System Model

A sensor network, illustrated in Figure 5.1, consisting of L sensors is considered. The sensors observe a deterministic parameter, θ, in noise. The value, x_l, observed at the lth sensor is

$$x_l = \theta + \sigma\eta_l \tag{5.1}$$

for $l = 1,...,L$, where θ is a deterministic, real-valued, unknown parameter in a bounded interval of known length, $[0, \theta_R]$, where $\theta_R < \infty$, and η_l are iid real-valued random variables drawn from a distribution symmetric about its median, zero, and $\sigma > 0$ is a scale parameter. In what follows, the location parameter of x_l is defined as the median of the distribution of x_l. The sensing SNR is defined as $\gamma := \theta^2/\sigma^2$. Due to constraints in the transmit power, we consider a scheme where the lth sensor transmits its measurement, x_l, using a constant modulus base-band equivalent signal, $\sqrt{\rho}e^{j\omega x_l}$, over a Gaussian multiple access channel so that the received signal at the fusion center is given by

$$y_L = \sqrt{\rho}\sum_{l=1}^{L} e^{j\omega x_l} + \nu, \tag{5.2}$$

where $\nu \sim \mathscr{CN}(0, \sigma_\nu^2)$ is the noise on the channel. All sensors transmit using the same single value of ω, requiring a single transmission per sensor. The transmissions are assumed to be appropriately pulse-shaped and phase modulated to consume finite bandwidth. The transmission power at each sensor is the same and is given by ρ. Two cases of power constraint are considered in this paper. In the first case, a total power constraint scenario is considered, where $\rho = P/L$. Irrespective of the number of sensors in the system, the power in the system remains the same. Due to this power constraint, the channel noise plays an important role in performance as will be shown. The other transmission scheme is a per-sensor power constraint, where $\rho = P$. Adding sensors

adds power to the system, and as $L \to \infty$, the channel noise can be ignored. This can also be considered a special case of the per-sensor power constraint approach with $\sigma_v^2 \to 0$.

The parameter, $\omega \in (0, 2\pi/\theta_R]$, in the right hand side of (5.2), is a design parameter to be optimized, and $v \sim \mathscr{CN}(0, \sigma_v^2)$ is the channel noise independent of $\{\eta_l\}_{l=1}^L$. Note that the restriction $\omega \in (0, 2\pi/\theta_R]$ is necessary even in the absence of sensing and channel noise ($y_L = \sqrt{P}e^{j\omega\theta}$) to uniquely determine θ from y_L.

The objective of this work is to estimate the values of θ and σ. Using these estimates of θ and σ, the estimate of the SNR of the system is calculated. In order to generalize the problem to include distributions for which moments are not defined, such as the Cauchy distribution, the problem can be interpreted as estimating the location parameter and the scale parameter of x_l, represented by θ and σ, respectively.

5.3 Total Power Constraint

In the total power constraint regime, the total transmit power is held to a constant irrespective of the number of sensors in the system. In a system of L sensors, if the total power available is P, then each sensor transmits with a power of P/L. The signal at the FC, shown in (5.2) is given by

$$y_L = \sqrt{\frac{P}{L}} \sum_{l=1}^L e^{j\omega x_l} + v. \tag{5.3}$$

The Estimator

At the FC, the estimator acts on the received signal, y_L. Defining

$$z_L := \frac{y_L}{\sqrt{L}} = \sqrt{P}\frac{1}{L} \sum_{i=1}^L e^{j\omega x_i} + \frac{v}{\sqrt{L}}, \tag{5.4}$$

it can easily be seen that in the absence of channel noise ($\sigma_v^2 = 0$), $|z_L| \leq \sqrt{P}$. Due to noise in the system, however, this may not always be the case. The effects of having $z_L > \sqrt{P}$ on the estimator will be examined in detail later in this section. Asymptoti-

cally, as $L \to \infty$,

$$z := \lim_{L \to \infty} z_L = \sqrt{P} \lim_{L \to \infty} \frac{1}{L} \sum_{i=1}^{L} e^{j\omega x_i}$$

$$= \sqrt{P} e^{j\omega\theta} \varphi_\eta(\sigma\omega), \tag{5.5}$$

where

$$\varphi_\eta(\sigma\omega) = E\left[e^{j\omega\sigma\eta_i}\right] \tag{5.6}$$

is the characteristic function of η_i. The characteristic function, $\varphi_\eta(\sigma\omega) \in \mathbb{R}$, since the distribution of η_i is symmetric about the median, $\forall i$. Also define $\mathbf{z}_L := [z_L^R z_L^I]^T$ where z_L^R and z_L^I are the real and imaginary parts of z_L, respectively. The vector $\mathbf{z}_L$ converges for large L to $\bar{\mathbf{z}} = [\bar{z}^R \bar{z}^I]^T$, where $\bar{z}^R = \lim_{L \to \infty} z_L^R$ and $\bar{z}^I = \lim_{L \to \infty} z_L^I$ and $\bar{z}^R$ and $\bar{z}^I$ are also the real and imaginary parts of z, respectively. Due to the central limit theorem, this convergence takes place in such a way that

$$\tilde{\mathbf{z}} = \lim_{L \to \infty} \sqrt{L}(\mathbf{z}_L - \bar{\mathbf{z}}) \tag{5.7}$$

is a 2×1 Gaussian random vector with zero mean and a 2×2 covariance matrix $\Sigma(\theta)$ with elements

$$\Sigma_{11}(\theta) = P\left[v_c \cos^2(\omega\theta) + v_s \sin^2(\omega\theta)\right] + \frac{1}{2}\sigma_v^2$$

$$\Sigma_{22}(\theta) = P\left[v_s \cos^2(\omega\theta) + v_c \sin^2(\omega\theta)\right] + \frac{1}{2}\sigma_v^2$$

$$\Sigma_{12}(\theta) = \Sigma_{21}(\theta) = P(v_c - v_s)\sin(\omega\theta)\cos(\omega\theta), \tag{5.8}$$

where $v_c := \mathrm{var}[\cos(\omega\eta_l)] = 1/2 + \varphi_\eta(2\sigma\omega)/2 - \varphi_\eta^2(\sigma\omega)$ and $v_s := \mathrm{var}[\sin(\omega\eta_l)] = 1/2 - \varphi_\eta(2\sigma\omega)/2$.

From $\mathbf{z}_L$ obtained at the FC, the values of θ and σ need to be estimated. The estimator for $[\hat{\theta}\hat{\sigma}]^T$ which yields the minimum variance is given by [115, (3.6.2), pp. 82]

$$\begin{bmatrix} \hat{\theta} \\ \hat{\sigma} \end{bmatrix} = \underset{\theta,\sigma}{\mathrm{argmin}}[\mathbf{z}_L - \bar{\mathbf{z}}]\Sigma^{-1}(\theta)[\mathbf{z}_L - \bar{\mathbf{z}}]^T. \tag{5.9}$$

84

The asymptotic covariance of this estimator is given by [115, Lemma 3.1]

$$C\left(\hat{\theta}, \hat{\sigma}\right) = \left[\mathbf{J_z}^T \mathbf{\Sigma}^{-1}(\theta)\mathbf{J_z}\right]^{-1}, \tag{5.10}$$

where $\mathbf{J_z}$ is the Jacobian matrix of $\bar{z}$ with respect to θ and σ and is given by

$$\mathbf{J_z} = \omega\sqrt{P}e^{-\omega^2\sigma^2/2} \begin{bmatrix} -\sin(\omega\theta) & -\omega\sigma\cos(\omega\theta) \\ \cos(\omega\theta) & -\omega\sigma\sin(\omega\theta) \end{bmatrix}. \tag{5.11}$$

This yields the following asymptotic covariance matrix:

$$C\left(\hat{\theta}, \hat{\sigma}\right) = \begin{bmatrix} \frac{P+\sigma_v^2-P\varphi_\eta(2\sigma\omega)}{2P\omega^2\varphi_\eta^2(\sigma\omega)} & 0 \\ 0 & \frac{P+\sigma_v^2-2P\varphi_\eta^2(\sigma\omega)+P\varphi_\eta(2\sigma\omega)}{2P\left[\frac{\partial\varphi_\eta(\sigma\omega)}{\partial\sigma}\right]^2} \end{bmatrix}. \tag{5.12}$$

From the structure of z_L in (5.5), alternate estimators for θ and σ can be built. Separating the signal into its absolute and phase components,

$$|z_L| = \sqrt{P}\varphi_\eta(\sigma\omega), \tag{5.13}$$

$$\angle z_L = \omega\theta, \tag{5.14}$$

where (5.13) depends on σ and not θ, whereas (5.14) depends on θ and not σ, and can be used to construct low-complexity estimators. The estimator for $\hat{\theta}$ is the solution to (5.14) and the estimator of $\hat{\sigma}$ is the solution to (5.13). While these estimators are low-complexity, their performance needs to be studied. In what follows, the relationship between these estimators and the minimum-variance joint estimator in (5.9) is established.

Theorem 5.3.1 *The estimates $\hat{\theta}$ and $\hat{\sigma}$ that solve (5.14) and (5.13), respectively, are those that minimize (5.9) when $|z_L| \leq \sqrt{P}$.*

Proof The estimator in (5.9) can be simplified to

$$\begin{bmatrix} \hat{\theta} \\ \hat{\sigma} \end{bmatrix} = \underset{\sigma}{\arg\min}\,\underset{\theta}{\arg\min}\Big\{ -|z_L|^2\left[\varphi_\eta^2(\sigma\omega)+1\right] + P\varphi_\eta^2(\sigma\omega)\left[\varphi_\eta(2\sigma\omega)-1\right]$$

$$+ \left[z^R\cos(\omega\theta)+z^I\sin(\omega\theta)\right]^2$$

$$- 2\sqrt{P}s(\theta)\varphi_\eta(\sigma\omega)\left[z^R\cos(\omega\theta)+z^I\sin(\omega\theta)\right]\Big\}, \tag{5.15}$$

85

where the joint minimization is no longer required due to the separation of θ and σ. Defining $s(\theta) = z^R \cos(\omega\theta) + z^I \sin(\omega\theta)$, the problem is rewritten first as

$$
\begin{bmatrix} \hat{\theta} \\ \hat{\sigma} \end{bmatrix} = \underset{s(\theta)}{\mathrm{argmin}} \Big\{ -|z_L|^2 \left[\varphi_\eta^2(\sigma\omega) + 1 \right] + P\varphi_\eta^2(\sigma\omega) \left[\varphi_\eta(2\sigma\omega) - 1 \right]
$$
$$
+ \left[z^R \cos(\omega\theta) + z^I \sin(\omega\theta) \right]^2
$$
$$
- 2\sqrt{P} s(\theta) \varphi_\eta(\sigma\omega) \left[z^R \cos(\omega\theta) + z^I \sin(\omega\theta) \right] \Big\}, \tag{5.16}
$$

which yields

$$
s^{\mathrm{opt}}(\theta) = |z_L|. \tag{5.17}
$$

The minimization problem in (5.15) can now be rewritten as

$$
\hat{\sigma} = \underset{\sigma > 0}{\mathrm{argmin}} \left[1 - \varphi_\eta(2\sigma\omega) \right] \left[|z_L| - \sqrt{P}\varphi_\eta(\sigma\omega) \right]. \tag{5.18}
$$

The first term in (5.18) can be made arbitrarily small as $\sigma \to 0$. For large L, it can be seen from (5.4) that the effect of the channel noise is diminished, and with high probability, $|z_L| \leq \sqrt{P}$. When $|z_L| > \sqrt{P}$, the objective function is minimized when $\sigma \to 0$, the estimator returns $\hat{\sigma} \to 0$, which indicates that the estimator has failed. When $|z_L| \leq \sqrt{P}$, the objective function is minimized when $|z_L| - \sqrt{P}\varphi_\eta(\sigma\omega) = 0$, which is identical to (5.13). Substituting this value in (5.17), the equation in (5.14) is obtained, completing the proof.

∎

Since the two estimators are identical, their performance will also be the same, as given by (4.12), which is a diagonal matrix. This implies that the estimate of the location parameter and the estimate of the scale parameter are asymptotically independent. The asymptotic variances of the location and scale parameters can be denoted individually as $\mathrm{AsV}_{\hat{\theta}}(\omega)$ and $\mathrm{AsV}_{\hat{\sigma}}(\omega)$, respectively.

The estimator for θ presented as the solution to (5.14) is the same as the estimator used in [62, 135], where only the estimate of the location parameter was considered.

Therefore, in the rest if this paper we will focus on the estimation of σ and the performance of this estimator. For the estimation of γ and to study the performance of the estimator of γ, the results of [62, 135] are used.

From the scale parameter and the location parameter of x_l, the SNR of the transmission can be estimated as

$$\hat{\gamma} = \frac{\hat{\theta}^2}{\hat{\sigma}^2}. \tag{5.19}$$

From the asymptotic covariance matrix of $\hat{\theta}$ and $\hat{\sigma}$, the asymptotic variance of $\hat{\gamma}$ is given by [108]

$$\mathrm{AsV}_{\hat{\gamma}}(\omega) = \left[\begin{array}{cc} \frac{\partial \gamma}{\partial \theta} & \frac{\partial \gamma}{\partial \sigma} \end{array}\right] C\left(\hat{\theta}, \hat{\sigma}\right) \left[\begin{array}{c} \frac{\partial \gamma}{\partial \theta} \\ \frac{\partial \gamma}{\partial \sigma} \end{array}\right], \tag{5.20}$$

where $\gamma = \theta^2/\sigma^2$, and $\mathrm{AsV}_{\hat{\gamma}}(\omega)$ depends only on $\hat{\theta}$, $\hat{\sigma}$ and the covariance matrix of $\hat{\theta}$ and $\hat{\sigma}$. Therefore, in what follows, for SNR estimation, we will concentrate on the estimation of $\hat{\theta}$ and $\hat{\sigma}$, and evaluate the performance of these estimates.

Theorem 5.3.2 *Let ω_1, ω_2 and ω_3 minimize $\mathrm{AsV}_{\hat{\theta}}(\omega)$, $\mathrm{AsV}_{\hat{\sigma}}(\omega)$ and $\mathrm{AsV}_{\hat{\gamma}}(\omega)$, respectively. If $\mathrm{AsV}_{\hat{\theta}}(\omega)$ and $\mathrm{AsV}_{\hat{\sigma}}(\omega)$ are convex functions of ω, then*

$$\omega_1 \leq \omega_3 \leq \omega_2. \tag{5.21}$$

Proof If ω_1 minimizes $\mathrm{AsV}_{\hat{\theta}}(\omega)$, then

$$\left.\frac{\partial \mathrm{AsV}_{\hat{\theta}}(\omega)}{\partial \omega}\right|_{\omega=\omega_1} = 0, \quad \left.\frac{\partial^2 \mathrm{AsV}_{\hat{\theta}}(\omega)}{\partial \omega^2}\right|_{\omega=\omega_1} > 0. \tag{5.22}$$

Similarly for ω_2

$$\left.\frac{\partial \mathrm{AsV}_{\hat{\sigma}}(\omega)}{\partial \omega}\right|_{\omega=\omega_2} = 0, \quad \left.\frac{\partial^2 \mathrm{AsV}_{\hat{\sigma}}(\omega)}{\partial \omega^2}\right|_{\omega=\omega_2} > 0. \tag{5.23}$$

When the asymptotic variances expressions are convex in ω, then the minima are unique.

From (5.20), the expression for the asymptotic variance of $\hat{\gamma}$ is given by

$$\mathrm{AsV}_{\hat{\gamma}}(\omega) = \alpha \mathrm{AsV}_{\hat{\theta}}(\omega) + \beta \mathrm{AsV}_{\hat{\sigma}}(\omega), \tag{5.24}$$

87

where $\alpha = 4\theta^2/\sigma^4 > 0$ and $\beta = 16\theta^2/\sigma^6 > 0$. Since the coefficients of $\text{AsV}_{\hat{\theta}}(\omega)$ and $\text{AsV}_{\hat{\sigma}}(\omega)$ in (5.24) are positive, if $\text{AsV}_{\hat{\theta}}(\omega)$ and $\text{AsV}_{\hat{\sigma}}(\omega)$ are convex function of ω, then $\text{AsV}_{\hat{\gamma}}(\omega)$ is also a convex function of ω. If ω_3 is the minimizer of $\text{AsV}_{\hat{\gamma}}(\omega)$, it is required to verify that

$$\left.\frac{\partial \text{AsV}_{\hat{\gamma}}(\omega)}{\partial \omega}\right|_{\omega=\omega_3} = 0, \quad \left.\frac{\partial^2 \text{AsV}_{\hat{\gamma}}(\omega)}{\partial \omega^2}\right|_{\omega=\omega_3} > 0. \tag{5.25}$$

To verify the second-derivative condition of (5.25), rewriting the left-hand-side in terms of $\text{AsV}_{\hat{\theta}}(\omega)$ and $\text{AsV}_{\hat{\sigma}}(\omega)$ yields

$$\alpha \left.\frac{\partial^2 \text{AsV}_{\hat{\sigma}}(\omega)}{\partial \omega^2}\right|_{\omega=\omega_3} + \beta \left.\frac{\partial^2 \text{AsV}_{\hat{\sigma}}(\omega)}{\partial \omega^2}\right|_{\omega=\omega_3}, \tag{5.26}$$

which is greater than zero since $\alpha > 0$, $\beta > 0$ and due to the convexity of $\text{AsV}_{\hat{\theta}}(\omega)$ and $\text{AsV}_{\hat{\sigma}}(\omega)$.

The condition for the first-derivative of (5.25) can be rewritten similarly so that that the condition to be verified is given by

$$\frac{\left.\frac{\partial \text{AsV}_{\hat{\gamma}}(\omega)}{\partial \omega}\right|_{\omega=\omega_3}}{\left.\frac{\partial \text{AsV}_{\hat{\gamma}}(\omega)}{\partial \omega}\right|_{\omega=\omega_3}} = -\frac{\beta}{\alpha}. \tag{5.27}$$

The right hand side of (5.27) should always be negative since both $\alpha > 0$ and $\beta > 0$. This happens only when one of the slopes of $\text{AsV}_{\hat{\theta}}(\omega)$ and $\text{AsV}_{\hat{\sigma}}(\omega)$ is positive and the other is negative. By studying the functions, it can be seen that when the functions are convex, and when $\omega_1 < \omega_2$, the ω axis can be divided into three regions: (i) $\omega < \omega_1$, where both $\text{AsV}_{\hat{\theta}}(\omega)$ and $\text{AsV}_{\hat{\sigma}}(\omega)$ have negative slope; (ii) $\omega_1 < \omega < \omega_2$, where $\text{AsV}_{\hat{\theta}}(\omega)$ has a positive slope and $\text{AsV}_{\hat{\sigma}}(\omega)$ has a negative slope; and (iii) $\omega > \omega_2$, where $\text{AsV}_{\hat{\theta}}(\omega)$ and $\text{AsV}_{\hat{\sigma}}(\omega)$ both have positive slope. Therefore, the condition in (5.27) is satisfied only when $\omega_1 \leq \omega_3 \leq \omega_2$. A similar argument can be made when $\omega_1 > \omega_2$.

∎

88

In what follows, Theorem 5.3.1 is verified for three sensing noise distributions: Gaussian, Laplace and Cauchy. In addition, the optimum values of ω for estimating θ, σ and γ are also determined.

Gaussian Distribution

The case of Gaussian distributed sensing noise is considered first. The noise at the sensors is given by $\eta_i \sim \mathcal{N}(0, \sigma^2), \forall i$, where σ is the standard deviation of the Gaussian distribution. The characteristic function in this case is given by

$$\varphi_\eta(\sigma\omega) = e^{-\omega^2\sigma^2/2} \tag{5.28}$$

and the value of z is

$$z = \sqrt{P}e^{j\omega\theta}e^{-\omega^2\sigma^2/2}. \tag{5.29}$$

The low complexity estimators constructed from z using (5.13) and (5.14) are given by

$$\hat{\theta} = \frac{1}{\omega}\angle z, \tag{5.30}$$

$$\hat{\sigma} = \frac{1}{\omega}\sqrt{\log\left(\frac{P}{|z|^2}\right)}. \tag{5.31}$$

The asymptotic covariance matrix for these estimates can be calculated to be

$$C\left(\hat{\theta}, \hat{\sigma}\right) = \begin{bmatrix} \frac{P+\sigma_v^2-Pe^{-2\omega^2\sigma^2}}{2P\omega^2e^{-\omega^2\sigma^2}} & 0 \\ 0 & \frac{P+\sigma_v^2-2Pe^{-\omega^2\sigma^2}+Pe^{-2\omega^2\sigma^2}}{2P\omega^4\sigma^2e^{-\omega^2\sigma^2}} \end{bmatrix}. \tag{5.32}$$

By substituting the Gaussian characteristic function from (5.28) in (5.12), it can be easily verified that the covariance matrices in (5.12) and (5.32) are identical, as expected.

The value of ω that minimizes the asymptotic variance of $\hat{\sigma}$ needs to be computed. Making the substitution $\beta \leftarrow \omega^2\sigma^2$ and differentiating with respect to β, the following equation is required to be solved to find the stationary points of the asymptotic variance of $\hat{\sigma}$

$$\beta\left[e^{2\beta}\left(\frac{\sigma_v^2}{P}+1\right)-1\right]-e^{2\beta}\left(\frac{\sigma_v^2}{P}+1\right)+2e^\beta-1 = 0. \tag{5.33}$$

89

It is straightforward to show that in the Gaussian case, $\partial^2 \text{AsV}_{\hat{\sigma}}(\omega)/\partial \omega^2 > 0$. There-fore, the asymptotic variance is convex, and the solution to (5.33) leads to the unique minimum, $\omega_{\sigma}^{\text{opt}} = \sqrt{\beta_{\sigma}^{\text{opt}}}/\sigma$, where $\beta_{\sigma}^{\text{opt}}$ is the solution to (5.33). Similarly, it can be shown that the asymptotic variance of $\hat{\theta}$ is convex. The value of ω that minimizes the asymptotic variance is given by $\omega_{\theta}^{\text{opt}} = \sqrt{\beta_{\theta}^{\text{opt}}}/\sigma$, where $\beta_{\theta}^{\text{opt}}$ is the solution to

$$\left(\frac{\sigma_v^2}{P} + 1\right)(\beta - 1)e^{2\beta} + (\beta + 1) = 0. \tag{5.34}$$

Neither (5.33) nor (5.34) can be solved analytically, but the solutions can be obtained numerically.

The asymptotic variance of the SNR estimate is calculated using (5.20) and is given by

$$\text{AsV}_{\hat{\gamma}}(\omega) = \frac{\omega^2\sigma^4\left[P + \sigma_v^2 - Pe^{-2\omega^2\sigma^2}\right] + \theta^2\left[P + \sigma_v^2 - 2Pe^{-\omega^2\sigma^2} + Pe^{-2\omega^2\sigma^2}\right]}{2P\omega^4\sigma^4 e^{-\omega^2\sigma^2}}. \tag{5.35}$$

Let the value of ω that minimizes $\text{AsV}_{\hat{\gamma}}(\omega)$ be denoted by $\omega_{\gamma}^{\text{opt}}$. From Theorem 5.3.2, $\omega_{\gamma}^{\text{opt}}$ is the unique minimizer of $\text{AsV}_{\hat{\gamma}}(\omega)$, and $\omega_{\theta}^{\text{opt}} \leq \omega_{\gamma}^{\text{opt}} \leq \omega_{\sigma}^{\text{opt}}$. It is easy to verify that $\omega_{\gamma}^{\text{opt}} = \sqrt{\beta_{\gamma}^{\text{opt}}}/\sigma$, where $\beta_{\gamma}^{\text{opt}}$ is the solution to

$$\beta\left[\beta\left\{e^{2\beta}\left(\frac{\sigma_v^2}{P} + 1\right) + 1\right\} - e^{2\beta}\left(\frac{\sigma_v^2}{P} + 1\right) + 1\right]$$
$$+\gamma\left[\beta\left\{e^{2\beta}\left(\frac{\sigma_v^2}{P} + 1\right) - 1\right\} - 2\left\{e^{2\beta}\left(\frac{\sigma_v^2}{P} + 1\right) - 2e^{\beta} + 1\right\}\right] = 0. \tag{5.36}$$

Laplace Distribution

Let η_i be drawn from a Laplace distribution of mean zero and variance σ^2. The char-acteristic function is

$$\varphi_\eta(\sigma\omega) = \frac{1}{1 + \frac{\omega^2\sigma^2}{2}} \tag{5.37}$$

and the value of z is

$$z = \frac{\sqrt{P}e^{j\omega\theta}}{1 + \frac{\omega^2\sigma^2}{2}}. \tag{5.38}$$

90

The naive estimators in this case are

$$\hat{\theta} = \frac{1}{\omega}\angle z, \tag{5.39}$$

$$\hat{\sigma} = \frac{\sqrt{2}}{\omega}\sqrt{\frac{\sqrt{P}}{|z|} - 1}, \tag{5.40}$$

with the asymptotic covariance matrix given by

$$C(\hat{\theta},\hat{\sigma}) = \begin{bmatrix} \dfrac{\left(\frac{\sigma_v^2}{P}+1\right)(1+4\omega^2\sigma^2)-1}{2\omega^2(1+4\omega^2\sigma^2)^2(1+\omega^2\sigma^2)^{-2}} & 0 \\ 0 & \dfrac{\left(\frac{\sigma_v^2}{P}+1\right)(1+4\omega^2\sigma^2)(1+\omega^2\sigma^2)^2+(1+\omega^2\sigma^2)^2-2(1+4\omega^2\sigma^2)}{8\omega^4\sigma^2(1+4\omega^2\sigma^2)(1+\omega^2\sigma^2)^{-2}} \end{bmatrix}. \tag{5.41}$$

Using (5.20), the asymptotic variance of $\hat{\gamma}$ is given by

$$\text{AsV}_{\hat{\gamma}}(\omega) = \frac{\theta^2\left(1+\omega^2\sigma^2\right)^2}{2P\omega^4\sigma^8\left(1+4\omega^2\sigma^2\right)}\Big[4\sigma^4\omega^2\left\{4P\omega^2\sigma^2+\sigma_v^2\left(1+4\omega^2\sigma^2\right)\right\}$$
$$+\theta^2\left\{2\sigma^4\omega^4P\left(5+2\omega^2\sigma^2\right)+\sigma_v^2\left(1+4\omega^2\sigma^2\right)\left(1+\omega^2\sigma^2\right)^2\right\}\Big]. \tag{5.42}$$

To minimize the asymptotic variance of $\hat{\theta}$ it can be shown that $\omega_\theta^{\text{opt}}$ is given by $\omega_\theta^{\text{opt}} = \sqrt{\beta_\theta^{\text{opt}}}/\sigma$, where [62]

$$\sqrt{\beta_\theta^{\text{opt}}} = \frac{1}{12}\left(\frac{c}{\frac{\sigma_v^2}{P}+1} + \frac{25\frac{\sigma_v^2}{P}+4}{c} + 2\right), \tag{5.43}$$

and

$$c = \Bigg[125\left(\frac{\sigma_v^2}{P}\right)^3 + 258\left(\frac{\sigma_v^2}{P}\right)^2 + 141\left(\frac{\sigma_v^2}{P}\right)$$
$$+ 3\sqrt{3}\sqrt{\left(\frac{\sigma_v^2}{P}\right)\left(\frac{\sigma_v^2}{P}+1\right)^3\left(375\frac{\sigma_v^2}{P}+32\right)} + 8\Bigg]^{1/3}. \tag{5.44}$$

To minimize the asymptotic variance of $\hat{\sigma}$, one needs to calculate $\omega_\sigma^{\text{opt}} = \sqrt{\beta_\sigma^{\text{opt}}}/\sigma$, where $\beta_\sigma^{\text{opt}}$ is the solution to the quintic equation

$$16\left(\frac{\sigma_v^2}{P}+1\right)\beta^5 + 2\left(12\frac{\sigma_v^2}{P}+13\right)\beta^4 - \left(7\frac{\sigma_v^2}{P}+8\right)\beta^3 - 23\frac{\sigma_v^2}{P}\beta^2 - 9\frac{\sigma_v^2}{P}\beta - \frac{\sigma_v^2}{P} = 0. \tag{5.45}$$

Similarly, the asymptotic variance of $\hat{\gamma}$ is minimized at $\omega_\gamma^{\text{opt}} = \sqrt{\beta_\gamma^{\text{opt}}}/\sigma$, where $\beta_\gamma^{\text{opt}}$ is the solution to the quintic equation

$$16\left(\gamma+2+\gamma\frac{\sigma_v^2}{P}+2\frac{\sigma_v^2}{P}\right)\beta^5 + 2\left(13\gamma-8+12\gamma\frac{\sigma_v^2}{P}-8\frac{\sigma_v^2}{P}\right)\beta^4$$
$$+7\left(7\gamma\frac{\sigma_v^2}{P}-14\frac{\sigma_v^2}{P}-8\gamma\right)\beta^3 - (23\gamma+2)\frac{\sigma_v^2}{P}\beta^2 - 9\gamma\frac{\sigma_v^2}{P}\beta - \gamma\frac{\sigma_v^2}{P} = 0. \quad (5.46)$$

The quintic equations in (5.45) and (5.46) cannot be solved analytically. However, the solutions to these can be obtained numerically.

Cauchy Distribution

Since the Cauchy distribution does not have any moments defined, the scale parameter in this case is selected to be the Cauchy parameter. The characteristic function is given by

$$\varphi_\eta(\sigma\omega) = e^{-\sigma\omega} \quad (5.47)$$

to yield

$$z = \sqrt{P}e^{j\omega\theta}e^{-\omega\sigma} \quad (5.48)$$

and the naive estimates of θ and σ are given as

$$\hat{\theta} = \frac{1}{\omega}\angle z, \quad (5.49)$$

$$\hat{\sigma} = \frac{1}{\omega}\log\left(\frac{\sqrt{P}}{|z|}\right). \quad (5.50)$$

These naive estimators have the asymptotic covariance matrix given by

$$C\left(\hat{\theta},\hat{\sigma}\right) = \begin{bmatrix} \frac{P+\sigma_v^2-Pe^{2\omega\sigma}}{2P\omega^2 e^{-2\omega\sigma}} & 0 \\ 0 & \frac{P+\sigma_v^2-Pe^{2\omega\sigma}}{2P\omega^2 e^{-2\omega\sigma}} \end{bmatrix}, \quad (5.51)$$

which is a scaled 2×2 identity matrix. The asymptotic variance of $\hat{\gamma}$ can be calculated using (5.20) and is given by

$$\text{AsV}_{\hat{\gamma}}(\omega) = \frac{2\theta^2\left(\theta^2+\sigma^2\right)\left(P+\sigma_v^2-Pe^{2\omega\sigma}\right)}{P\omega^2\sigma^6 e^{-2\omega\sigma}}. \quad (5.52)$$

Since the asymptotic variances of both $\hat{\theta}$ and $\hat{\sigma}$ are identical, from Theorem 5.3.2, the same value of ω minimizes the asymptotic variances of all $\hat{\theta}$, $\hat{\sigma}$ and $\hat{\gamma}$. Taking the first derivative of the asymptotic variance with respect to ω and equating to zero, the value of ω that minimizes the asymptotic variances is given by

$$\omega^{\text{opt}} = \frac{2 + W\left[-\frac{2P}{e^2\left(P+\sigma_v^2\right)}\right]}{2\sigma},$$
(5.53)

where $W(\cdot)$ is the Lambert-W function [136].

5.4   Per-Sensor Power Constraint

In the case of per-sensor power constraint, the total transmit power increases as the number of sensors in the system increases, with the channel noise remaining the same. Each sensor transmits with a power of P and the signal at the FC, shown in (5.2) is given by

$$y_L = \sqrt{P}\sum_{l=1}^{L} e^{j\omega x_l} + v.$$
(5.54)

As the number of sensors increases, the effect of channel noise becomes negligible and can be ignored. In fact, the results in the case of per-sensor power constraint can be interpreted as a special case of the results in Section 5.3, with $\sigma_v^2 \to 0$. While this simply a special case of the results presented in the previous section, the development is included since closed form solutions can be obtained for ω^{opt} for all cases considered.

The Estimator

At the FC, the signal from (5.54) is modified to give

$$\zeta_L := \frac{y_L}{L} = \sqrt{P}\frac{1}{L}\sum_{i=1}^{L} e^{j\omega x_i} + \frac{v}{L},$$
(5.55)

which as $L \to \infty$, converges in probability to

$$\zeta = \lim_{L\to\infty} \zeta_L = \sqrt{P}\lim_{L\to\infty}\frac{1}{L}\sum_{i=1}^{L} e^{j\omega x_i} = \sqrt{P}e^{j\omega\theta}\varphi_\eta(\sigma\omega).$$
(5.56)

Defining $\zeta_L = [\zeta_L^R \zeta_L^I]$ and $\zeta = [\zeta^R \zeta^I]$, ζ_L converges to ζ in such a way that

$$\tilde{\zeta} = \lim_{L\to\infty} \sqrt{L}(\zeta_L - \zeta)$$
(5.57)

93

is a 2×1 Gaussian random vector with zero mean and a 2×2 covariance matrix $\tilde{\Sigma}(\theta)$ with elements

$$\tilde{\Sigma}_{11}(\theta) = P\left[\tilde{v}_c \cos^2(\omega\theta) + \tilde{v}_s \sin^2(\omega\theta)\right]$$

$$\tilde{\Sigma}_{22}(\theta) = P\left[\tilde{v}_s \cos^2(\omega\theta) + \tilde{v}_c \sin^2(\omega\theta)\right]$$

$$\tilde{\Sigma}_{12}(\theta) = \tilde{\Sigma}_{21}(\theta) = P(\tilde{v}_c - \tilde{v}_s)\sin(\omega\theta)\cos(\omega\theta), \tag{5.58}$$

where $\tilde{v}_c := \text{var}[\cos(\omega\eta_l)] = 1/2 + \varphi_\eta(2\sigma\omega)/2 - \varphi_\eta^2(\sigma\omega)$ and $\tilde{v}_s := \text{var}[\sin(\omega\eta_l)] = 1/2 - \varphi_\eta(2\sigma\omega)/2$. The minimum variance estimator for $[\hat{\theta}\,\hat{\sigma}]^T$ in this case is given by

$$\begin{bmatrix} \hat{\theta} \\ \hat{\sigma} \end{bmatrix} = \underset{\theta,\sigma}{\text{argmin}}[\zeta_L - \zeta]\tilde{\Sigma}^{-1}(\theta)[\zeta_L - \zeta]^T, \tag{5.59}$$

and the asymptotic covariance matrix of the estimates is given by

$$C\left(\hat{\theta},\hat{\sigma}\right) = \left[\mathbf{J}_{\mathbf{z}}^T \tilde{\Sigma}^{-1}(\theta)\mathbf{J}_{\mathbf{z}}\right]^{-1} = \begin{bmatrix} \frac{1-\varphi_\eta(2\sigma\omega)}{2\omega^2\varphi_\eta^2(\sigma\omega)} & 0 \\ 0 & \frac{1-2\varphi_\eta^2(\sigma\omega)+\varphi_\eta(2\sigma\omega)}{2\left[\frac{\partial\varphi_\eta(\sigma\omega)}{\partial\sigma}\right]^2} \end{bmatrix}, \tag{5.60}$$

which can be verified to be (5.12) with $\sigma_v \to 0$. The estimate of the computed as given in (5.19), with asymptotic variance as given in (5.20). Theorem 5.3.1 and Theorem 5.3.2 continue to hold.

The three sensing noise distributions considered previously, the Gaussian distribution, the Laplace distribution and the Cauchy distribution are considered again for the per-sensor power constraint case. Since the sensing noise stays the same, the low-complexity estimators stay the same, but their performance changes. In each case, the performance is evaluated and the values of ω that minimize the asymptotic variances of $\hat{\theta}$, $\hat{\sigma}$ and $\hat{\gamma}$ are calculated.

Gaussian Distribution

The performance in this case is given by substituting $\sigma_v = 0$ in (5.32) to give

$$C\left(\hat{\theta},\hat{\sigma}\right) = \begin{bmatrix} \frac{1-e^{-2\omega^2\sigma^2}}{2\omega^2 e^{-\omega^2\sigma^2}} & 0 \\ 0 & \frac{\left(1-e^{-\omega^2\sigma^2}\right)^2}{2\omega^4\sigma^2 e^{-\omega^2\sigma^2}} \end{bmatrix}. \tag{5.61}$$

94

The asymptotic variance of $\hat{\gamma}$ is given by

$$\text{AsV}_{\hat{\gamma}}(\omega) = \frac{\theta^2 \left(1 - 2e^{-\omega^2 \sigma^2} + e^{-2\omega^2 \sigma^2}\right) + \omega^2 \sigma^4 \left(1 - e^{-2\omega^2 \sigma^2}\right)}{2\omega^4 \sigma^4 e^{-\omega^2 \sigma^2}}. \tag{5.62}$$

The value of ω that minimizes the asymptotic variance of $\hat{\theta}$ is given by

$$\omega_{\theta}^{\text{opt}} = \underset{\omega}{\text{argmin}} \frac{1 - e^{-2\omega^2 \sigma^2}}{2\omega^2 e^{-\omega^2 \sigma^2}}. \tag{5.63}$$

It can easily be verified that the objective is minimized as $\omega_{\theta}^{\text{opt}} \to 0$. In a similar way, it can be shown that $\omega_{\sigma}^{\text{opt}} \to 0$ minimizes the asymptotic variance of $\hat{\sigma}$. From Theorem 5.3.2, $\text{AsV}_{\hat{\gamma}}(\omega)$ is also minimized when $\omega_{\gamma}^{\text{opt}} \to 0$.

Laplace Distribution

The asymptotic covariance matrix is given by (5.41) with $\sigma_\nu \to 0$:

$$C\left(\hat{\theta}, \hat{\sigma}\right) = \begin{bmatrix} \frac{2\sigma^2 \left(1 + \omega^2 \sigma^2\right)^2}{1 + 4\omega^2 \sigma^2} & 0 \\ 0 & \frac{\left(2\omega^2 \sigma^2 - 1\right)\left(1 + \omega^2 \sigma^2\right)^3}{4\omega^2 \left(1 + 4\omega^2 \sigma^2\right)} \end{bmatrix}. \tag{5.64}$$

The asymptotic variance of the estimate of γ is given by

$$\text{AsV}_{\hat{\gamma}}(\omega) = \frac{\gamma \left(1 + \omega^2 \sigma^2\right)^2 \left(8 + 5\gamma + 2\theta^2 \omega^2\right)}{\left(1 + 4\omega^2 \sigma^2\right)}. \tag{5.65}$$

To identify the value of ω that yields the best performance for estimating θ, the following problem needs to be solved:

$$\omega_{\theta}^{\text{opt}} = \underset{\omega}{\text{argmin}} \frac{2\sigma^2 \left(1 + \omega^2 \sigma^2\right)^2}{1 + 4\omega^2 \sigma^2}. \tag{5.66}$$

By inspecting the first derivative, it can be verified that $\omega_{\theta}^{\text{opt}} = 1/\sigma\sqrt{2}$. For the case of $\hat{\sigma}$

$$\omega_{\sigma}^{\text{opt}} = \underset{\omega}{\text{argmin}} \frac{\left(2\omega^2 \sigma^2 - 1\right)\left(1 + \omega^2 \sigma^2\right)^3}{4\omega^2 (1 + 4\omega^2 \sigma^2)}. \tag{5.67}$$

This is minimized at $\omega_{\sigma}^{\text{opt}} = \sqrt{3\sqrt{33} - 13}/4\sigma > \omega_{\theta}^{\text{opt}}$. The value of ω that minimizes $\text{AsV}_{\hat{\gamma}}(\omega)$ is similarly calculated to be

$$\omega_{\gamma}^{\text{opt}} = \frac{\sqrt{-13\theta^2 - 16\sigma^2 + \sqrt{(9\theta^2 + 16\sigma^2)(33\theta^2 + 16\sigma^2)}}}{4\theta\sigma}. \tag{5.68}$$

Cauchy Distribution

In the case of Cauchy distributed sensing noise, the asymptotic covariance matrix for the estimates, $\hat{\theta}$ and $\hat{\sigma}$, is given by

$$C\left(\hat{\theta},\hat{\sigma}\right) = \begin{bmatrix} \frac{1-e^{-2\omega\sigma}}{2\omega^2 e^{-2\omega\sigma}} & 0 \\ 0 & \frac{1-e^{-2\omega\sigma}}{2\omega^2 e^{-2\omega\sigma}} \end{bmatrix}, \tag{5.69}$$

which, similar to (5.51), is a scaled 2×2 identity matrix. The asymptotic variance of $\hat{\gamma}$ is given by

$$\mathrm{AsV}_{\hat{\gamma}}(\omega) = \frac{2\gamma(\gamma+1)\left(1-e^{2\omega\sigma}\right)}{\omega^2 \sigma^2 e^{-2\omega\sigma}}. \tag{5.70}$$

Since $\mathrm{AsV}_{\hat{\theta}}(\omega)$ and $\mathrm{AsV}_{\hat{\sigma}}(\omega)$ are identical, the value of ω that minimizes them is the same. Therefore, from Theorem 5.3.2, the same value of ω minimizes $\mathrm{AsV}_{\hat{\theta}}(\omega)$, $\mathrm{AsV}_{\hat{\sigma}}(\omega)$ and $\mathrm{AsV}_{\hat{\gamma}}(\omega)$, and is given by

$$\omega^{\mathrm{opt}} = (2 + W(-2e^{-2}))/2\sigma. \tag{5.71}$$

5.5   Simulation Results

Simulations are used to verify the numerical results obtained above. In each case of sensing distribution, the minimum variance estimator is simulated, and compared against the respective naive estimators for θ and σ. These estimators are then compared against the CRLB.

In Figure 5.2, the naive estimators and the minimum-variance estimator are compared when the sensing noise is Gaussian. The asymptotic variances of the two estimators are plotted and compared. It can be seen from the graph that for both the estimate of the location parameter, $\hat{\theta}$, and scale parameter, $\hat{\sigma}$, the performance is the same for both the naive estimator and the minimum variance estimator. When compared against the CRLB, in the case of Gaussian sensing noise, both the estimators are asymptotically efficient, since the asymptotic variances are the same as the respective values of the CRLB. This result was also seen previously in [63].

In Figure 5.3, the sensing noise is Laplace distributed. In this case, it can be verified that the performance of the naive estimator and the minimum-variance estima-
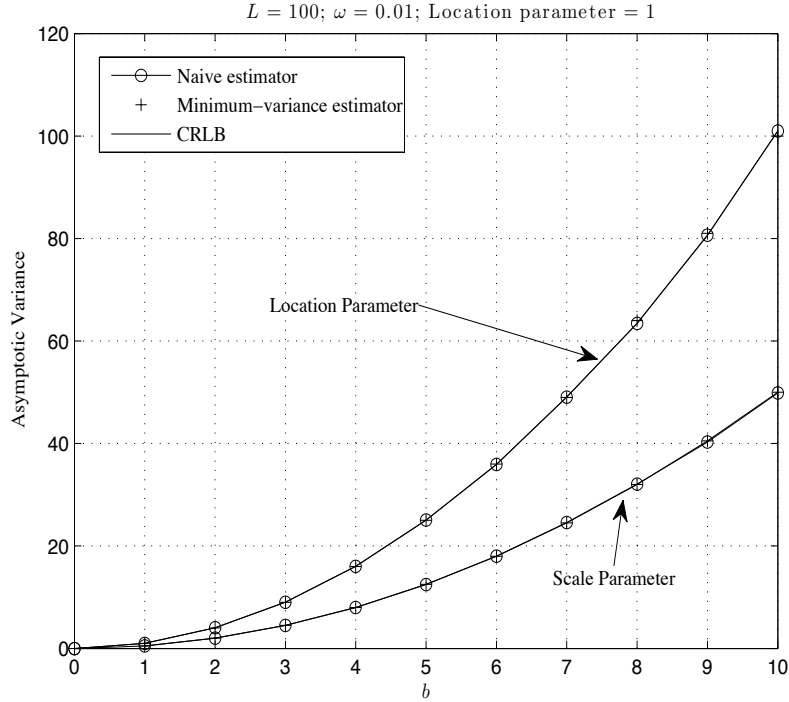
Figure 5.2: Asymptotic variance vs. scale parameter. Sensing noise is Gaussian distributed. The asymptotic variances match the CRLB.

tor is the same for each of the location parameter and scale parameter. However, the estimators are not asymptotically efficient as the the asymptotic variances are larger than the CRLB.

Cauchy distributed sensing noise was considered for the results shown in Figure 5.4. The estimators of both the location parameter and the scale parameter have the same performance. This is verified in the figure. Also, both parameters have the same CRLB, which are lower than the asymptotic variances of the location parameter and the scale parameter. Therefore, in the case when the sensing noise is Cauchy distributed, the estimators are also not asymptotically efficient.
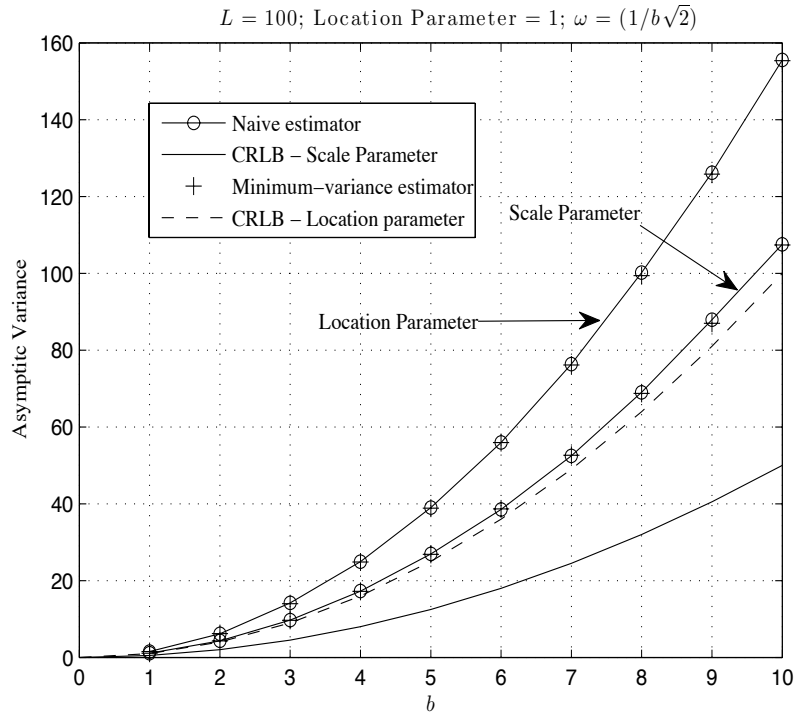
Figure 5.3: Asymptotic variance and CRLB vs. scale parameter. Sensing noise is Laplace distributed.
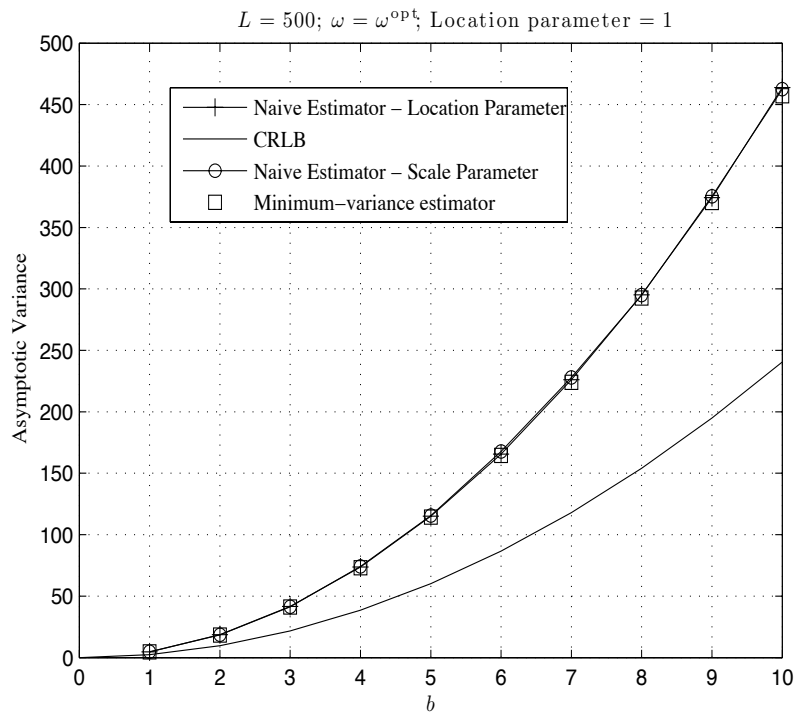
Figure 5.4: Performance vs. σ. Sensing noise is Cauchy distributed.

Chapter 6

CONCLUSIONS AND FUTURE WORK

6.1   Summary

In the preceding chapters, four distributed inference problems were presented. In the first case, distributed estimation was studied with a single antenna at the FC. In the second case, distributed detection with multiple antennas at the FC was considered. In both cases, the channels between the sensors and the FC were fading and the systems were studied with differing amounts of channel information at the sensors. In Chapter 4 and Chapter 5, constant-modulus phase modulated transmissions from the sensors were aggregated at the FC over Gaussian multiple-access channels. The scale parameter and the location parameters were estimated, then combined to form an SNR estimate. The performance of these estimators was studied. The asymptotic efficiency of the estimator of the location parameter was also studied.

In Chapter 2, the asymptotic variance of a linear estimator over fading multiple-access channels was evaluated for distributed estimation with different feedback scenarios and channel conditions. It was argued that the ratio of the asymptotic variances can be viewed as the factor by which the number of sensors for the system with the larger asymptotic variance would have to be increased so that the two systems have the same variance, for large number of sensors (about 50 or less as seen in the simulations). It was observed that for multiple access channels, performance with no CSI at the sensors was very poor. When the sensors have full channel information, the optimal sensor gains to obtain an achievable benchmark were derived, to give the smallest possible variance over fading channels. Furthermore, as the available power increased, this performance approached the AWGN performance. However, the drawback of this approach was the need for complete channel knowledge at the sensors and the required calculations to find the optimal sensor gains. When the channels were Rayleigh fading, the phase-only case had a performance loss of a factor of exactly $4/\pi$ when compared

99

to the AWGN channel case. This penalty was shown to decrease for line of sight scenarios.

The effects of inexact phase information at the sensors were also investigated. Continuous errors in phase feedback, phase quantization and errors on the feedback channel were also considered. Remarkably, in the asymptotic regime, when the number of sensors is large, it was possible to decouple the individual effects of phase-only feedback, quantization, and error in feedback, analytically. It was shown that using as few as three bits of channel phase information only caused deterioration of about 5% in the asymptotic variance, and that these systems were also robust to errors on the feedback channels. In the case of correlated channels, it was determined that a finitely correlated model guaranteed convergence to the asymptotic variance. In addition, a metric was derived to measure the speed of convergence and its dependence on the effect of noise, power and channel correlation on the speed of convergence was determined. With simulations, it was shown that only a few tens of sensors were needed for the asymptotic results to hold. Simulations were used to verify the analytical results for different fading models and feedback scenarios, and to show how the value of σ_A^2 was affected by correlation, M, observation noise and P.

A distributed detection system with sensors transmitting observations to a fusion center with multiple antennas was considered in Chapter 3. The error exponent was computed from the conditional probability of error. It was shown, that in certain cases, the error exponent converged to zero, indicating that that the error probability was not decaying exponentially, and the average asymptotic probability of error was used to evaluate such systems.

The performance with AWGN channels between the sensors and the FC was used as a benchmark. When the sensors had no channel information, Rayleigh fading channels and Ricean fading channels were considered between the sensors and the fusion center. When the channels were Ricean fading, the results were evaluated using

the error exponent, which is a function of the Ricean-K factor. As the number of antennas increased, or as the Ricean-K factor increased, performance improved. When, $K = 0$, i.e., when the channels were Rayleigh fading, the error exponent was zero, which indicated poor performance and the average asymptotic probability of error was computed. Finally, in all cases, adding antennas at the FC provided improvement in performance.

When the sensors had full channel information, the sensor gains were set to maximize the error exponent. When there were multiple antennas at the FC, the optimization problem was not tractable. Therefore, one lower bound and two upper bounds were computed and the minimum of the two upper bounds was used as the tight upper bound.

When the sensors only adjusted their phases for transmission, the performance was independent of the number of antennas at the FC. The performance was between $\mathscr{E}_{AWGN}(1)$ and $(\pi/4)\mathscr{E}_{AWGN}(1)$.

Having multiple antennas at the fusion center provided a gain of at most 2. However, if both the number of sensors and antennas scaled to infinity in such a way that the number of antennas at the FC scaled at least as fast as the number of sensors, larger gain was shown to be achieved. However, such a system is not practical for implementation.

Implementable, low-complexity, sub-optimal schemes were developed. In one approach, the system was configured to beamform to the antenna that provided the best performance, where the FC still used the data gathered at the other antennas. On an average, this was shown to perform better than in the single antenna case. Another approach was to assume there was no sensing noise, and the sensor gains were tuned for such a system even when sensing noise was present. In such a situation, the system performed optimally when the sensing noise in the system was low. A hybrid scheme was proposed which selected the better of these two methods depending on the sensing

SNR.

Depending on the number of sensors and antennas at the FC, and their rates of growth, the following system design recommendation can be made. If CSIS is available and the number of antennas at the FC is very much less than the number of sensors, then for better performance, it is recommended to increase the number of sensors, rather than the number of antennas at the FC. However, if the number of antennas at the FC can be increased at a much faster rate than the number of sensors, it is possible to achieve greater gains due to adding antennas at the FC.

In Chapter 4, the relationship between the Fisher information and the characteristic function was studied through two bounds. The condition for equality was also derived, for the first time in literature.

This result was used to prove the asymptotic efficiency of a distributed estimator that minimized the asymptotic variance in the presence of Gaussian sensing noise. Different sensing noise distributions were considered, and in all cases, the loss in efficiency was quantified through a scale-invariant relative efficiency metric that takes values between 0 and 1. This metric depends only on the distribution of the sensing noise used, and was computed for the Gaussian, Laplace, Cauchy and uniform cases. These relative efficiency values were interpreted as the amount of information lost due to constant modulus transmissions over Gaussian multiple-access channels relative to having perfect access to all sensor measurements. Numerical evaluations confirmed the result that the estimator of the location parameter derived in the chapter was asymptotically efficient only when the sensing noise is Gaussian.

A problem of simultaneous distributed estimation of the scale parameter and location parameter of a signal embedded in noise was considered in Chapter 5 for different sensing noise distributions. Sensors observed a parameter in sensing noise and modulated the observations using a constant-modulus exponential scheme. The sensors transmitted the observations over a Gaussian multiple-access channel to a fusion center.

Due to the additive nature of the channel, the signal received at the FC converged to the characteristic function of the sensing noise distribution as the number of sensors grew large. Two cases of sensor power were considered, one with a power constraint on each sensor, and one with a total power constraint across all the sensors.

At the fusion center, two types of estimators were used to estimate the location parameter and scale parameter. One of them, a minimum-variance estimator, was used to simultaneously estimate the parameters. Additionally, for each of the different sensing noise distributions, a low-complexity estimator was derived based on the structure of the characteristic function of the distribution. It was shown that these estimators are identical. For each case of sensing distribution, the optimum transmission parameter, ω, was calculated.

The asymptotic efficiency of the estimators was also evaluated. It was found that only in the case of Gaussian sensing noise, the estimators are asymptotically efficient.

6.2   Future Work
Variable PAPR transmissions

In the problems considered in Chapter 2 and Chapter 3, the peak to average power ratio (PAPR) is infinite. In the constant modulus problem considered in Chapter 4 and Chapter 5, the PAPR is one. The case with infinite PAPR indicates the need for power amplifiers with a large dynamic range. When the PAPR is one, all sensors transmit at the same power level all the time. This indicates that if a sensor has a power source with finite energy, the lifetime of the sensor is fixed. If some transmissions can occur with a lower energy, the overall life of the sensor can be increased. Therefore, if the transmission is redefined in such a way that

$$y = \sum_{i=1}^{L} \sqrt{\rho_i(x_i)} e^{j\omega f(x_i)} + v, \tag{6.1}$$

the task would be to choose $\rho_i(\cdot)$ and $f(\cdot)$ to satisfy a given PAPR, while maintaining good performance, which is similar in structure to the received signal in (2.1), with phase-only CSIS. It can be easily seen from Section 2.4 that the penalty incurred due

to such a transmission is given by

$$P_G := \frac{E[\|\rho_i\|]}{E^2[\|\sqrt{\rho}_i\|]}.$$ (6.2)

The minimum value of $P_G = 1$ when all the values of ρ_i are deterministic and equal. In order to find the best distribution of ρ, the following optimization problem may be posed:

$$\underset{\rho}{\operatorname{argmin}}\, P_G \quad \text{subject to } \sup(\rho) = P_P$$

$$\frac{\sup(\rho)}{E[\|\rho_i\|]} = P_A,$$ (6.3)

where P_P is the peak allowable power and P_A is the PAPR of the system. This problem can be solved to obtain different distributions of ρ under different conditions imposed on the nature of the distribution. In each case, the resulting estimator can be evaluated and studied.

Distributed Consensus

The problems in this dissertation all have a centralized sensor network architecture where the sensors observe a parameter embedded in noise and transmit their observations to the FC with minimal processing at the sensors. An alternate structure to this the paradigm of distributed consensus, where sensors communicate amongst themselves without a fusion center. Graph theory is used to determine the connectivity of the sensor network, consequently to establish a communication scheme. These computations are too demanding to be carried out at the sensors. While the consensus model assumes no centralized computer (such as a fusion center), this communication scheme is determined outside the network and fed to the network. Such a system does not account for changes in the network during operation. Future work in this are could be to develop distributed algorithms for routing and scheduling that can be broken down into fragments and processed at each sensor.

Additionally, estimators and communication schemes developed in Chapter 4 and Chapter 5 can be extended to the case of networks with no FC. The performance and efficiency of the algorithms under these conditions can be studied.

REFERENCES

[1]   G. J. Pottie and W. J. Kaiser, "Wireless integrated network sensors," *Communications of the ACM*, vol. 43, no. 5, pp. 51–58, May 2000.

[2]   N. Priyantha, A. Chakraborty, and H. Balakrishnan, "The cricket location-support system," *in the Proceedings of the ACM International Conference on Mobile Computing and Networking*, pp. 32–43, August 2000.

[3]   A. Cerpa, J. Elson, D. Estrin, L. Girod, M. Hamilton, and J. Zhao, "Habitat monitoring: Application driver for wireless communications technology," *in the Proceedings of the 2001 ACM SIGCOMM Workshop on Data Communications in Latin America and the Caribbean*, pp. 20–34, April 2001.

[4]   D. Li, K. D. Wong, Y. H. Hu, and A. M. Sayeed, "Detection, classification, and tracking of targets," *IEEE Signal Processing Magazine*, vol. 19, no. 17-29, 2002.

[5]   A. Mainwaring, J. Polastre, R. Szewczyk, D. Culler, and J. Anderson, "Wireless sensor networks for habitat monitoring," *in the Proceedings of the ACM international workshop on Wireless sensor networks and applications*, pp. 88–97, September 2002.

[6]   M. Maroti, G. Simon, A. Ledeczi, and J. Sztipanovits, "Shooter localization in urban terrain," *IEEE Computer Magazine*, pp. 60–61, August 2004.

[7]   J. Sallai, G. Balogh, M. Maroti, and A. Ledeczi, "Acoustic ranging in resource constrained sensor networks," *Technical Report ISIS-04-504, Institute for Software Integrated Systems*, 2004.

[8]   G. Simon, M. Maroti, A. Ledeczi, G. Balogh, B. Kusy, A. Nadas, G. Pap, J. Sallai, and K. Frampton, "Sensor network-based countersniper system," *in the Proceedings of the ACM Second International Conference on Embedded Networked Sensor Systems (SenSys 04)*, pp. 1–12, November 2004.

[9]   B. M. Sadler, "Fundamentals of energy-contrained sensor network systems," *IEEE A&E Systems Magazines*, vol. 20, no. 8, pp. 17–34, August 2005.

[10]   G. Pottie and W. Kaiser, *Principles of Embedded Networked Systems Design*. New York: Cambridge University Press, 2005.

[11]   V. Shnayder, B.-R. Chen, K. Lorincz, and T. Fulford-Jones, "Sensor networks for medical care," *Harvard University Technical Report TR-08-05*, April 2005.

[12] H. Kwon, H. Krishnamoorthi, V. Berisha, and A. Spanias, "A sensor network for real-time acoustic scene analysis," *IEEE International Symposium on Circuits and Systems*, pp. 169–172, May 2009.

[13] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Communication Magazine*, pp. 102–114, August 2002.

[14] B. Warneke, M. Last, B. Liebowitz, and K. S. J. Pister, "Smart dust: communicating with a cubic-millimeter computer," *Computer*, vol. 34, no. 1, pp. 44–51, January 2001.

[15] J. Rabaey, J. Ammer, T. Karalar, S. Li, B. Otis, M. Sheets, and T. Tuan, "Picoradios for wireless sensor networks: The next challenge in ultra-low-power design," *in the Proceedings of the IEEE International Solid-State Circuits Conference*, vol. 1, pp. 200–201, February 2002.

[16] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Transactions on Wireless Communications*, vol. 1, no. 4, pp. 660–670, 2002.

[17] E. Welsh, W. Fish, and P. Frantz, "Gnomes: A testbed for low-power heterogeneous wireless sensor networks," *in the Proceedings of the IEEE International Symposium on Circuits and Systems*, vol. 4, pp. 836–839, May 2003.

[18] J. Polastre, R. Szewczyk, C. Sharp, and D. Culler, "The mote revolution: Low power wireless sensor networks," *Symposium on High Performance Chips*, 2004.

[19] "The smart-Its project," available online at http://www.smart-its.org.

[20] "Crossbow Technology Inc." available online at http://www.xbow.com.

[21] L. Kleinrock and J. Silvester, "Optimum transmission radii for packet radio networks or why six is a magic number," in *NTC '78; National Telecommunications Conference, Birmingham, Ala., December 3-6, 1978, Conference Record. Volume 1. (A79-40501 17-32) Piscataway, N.J., Institute of Electrical and Electronics Engineers, Inc., 1978, p. 4.3.1-4.3.5.*, vol. 1, 1978, pp. 4.3.1–4.3.5.

[22] S. M. Hedetniemi, S. T. Hedetniemi, and A. L. Liestman, "A survey of gossiping and broadcasting in communication networks," *Networks*, vol. 18, no. 4, pp. 319–349, 1988.

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

[24] C. Rago, P. Willett, and Y. Bar-Shalom, "Censoring sensors: a low-communication-rate scheme for distributed detection," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 32, no. 2, pp. 554–568, April 1996.

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

[26] P. Gupta and P. R. Kumar, *Critical Power for Asymptotic Connectivity in Wireless Networks*. Birkhauser, 1998.

[27] ——, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 388–404, March 2000.

[28] M. Grossglauser and D. Tse, "Mobility increases the capacity of ad-hoc wireless networks," *in the Proceedings of IEEE Infocom*, pp. 1360–1369, 2001.

[29] O. Dousse, P. Thiran, and M. Hasler, "Connectivity in ad-hoc and hybrid networks," *in the Proceedings of IEEE Infocom*, pp. 1079–1088, 2002.

[30] H. E. Gamal, "On the scaling laws of dense wireless sensor networks," *Proceedings of the Annual Allerton Conference on Communication, Control and Coding*, vol. 41, no. 3, pp. 1393–1401, 2003.

[31] F. Xue and P. R. Kumar, *The Number of Neighbors Needed for Connectivity of Wireless Networks*. The Netherlands: Kluwer Academic Publishers, 2004, ch. 10, pp. 169–181.

[32] R. Olfati-Saber and R. M. Murray, "Consensus problems in networks of agents with switching topology and time-delays," *IEEE Transactions on Signal Processing*, vol. 49, no. 9, pp. 1520–1533, September 2004.

[33] L. Xiao, S. Boyd, and S. Lall, "A scheme for robust distributed sensor fusion based on average consensus," *in the Proceedings of IPSN*, 2005.

[34] O. Dousse, F. Baccelli, and P. Thiran, "Impact of interferences on connectivity in ad hoc networks," *IEEE/ACM Transactions on Networking*, vol. 13, no. 2, pp. 425–436, April 2005.

[35] F. Xue, L. Xie, and P. R. Kumar, "The transport capacity of wireless networks over fading channels," *IEEE Transactions on Information Theory*, vol. 51, no. 3, pp. 834–847, March 2005.

[36] R. Olfati-Saber and J. S. Shamma, "Consensus filters for sensor networks and distributed sensor fusion," *in the Proceedings of the IEEE Conference on Decision and Control*, 2005.

[37] ——, "Consensus filters for sensor networks and distributed sensor fusion," *in the Proceedings of the Conference on Decision and Control*, 2005.

[38] V. Saligrama, M. Alanyali, and O. Savas, "Distributed detection in sensor networks with packet losses and finite capacity links," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4118–4132, November 2006.

[39] R. Olfati-Saber, E. Franco, E. Frazzoli, and J. S. Shamma, *Belief Consensus and Distributed Hypothesis Testing in Sensor Networks*, ser. Lecture Notes in Control and Information Sciences.   Springer, 2006, vol. 331, pp. 169–182.

[40] L.-L. Xie and P. Kumar, "On the path-loss attenuation regime for positive cost and linear scaling of transport capacity in wireless networks," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2313–2328, June 2006.

[41] F. Xue and P. Kumar, *Scaling Laws for Ad Hoc Wireless Networks: An Information Theoretic Approach*.   Now Publishers, 2006, vol. 1, no. 2, pp. 145–270.

[42] I. D. Schizas, A. Ribeiro, and G. B. Giannakis, "Consensus in ad hoc wsns with noisy links Part I: Distributed estimation of deterministic signals," *IEEE Transactions on Signal Processing*, vol. 56, no. 1, pp. 350–364, January 2008.

[43] I. D. Schizas, G. B. Giannakis, S. I. Roumeliotis, and A. Ribeiro, "Consensus in ad hoc wsns with noisy links Part II: Distributed estimation and smoothing of random signals," *IEEE Transactions on Signal Processing*, vol. 56, no. 4, pp. 1650–1666, April 2008.

[44] S. Kirti and A. Scaglione, "Scalable distributed kalman filtering through consensus," *in the Proceedings of ICASSP*, 2008.

[45] H. Medeiros, J. Park, and A. C. Kak, "Distributed object tracking using a cluster-based kalman filter in wireless camera networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 4, pp. 448–463, August 2008.

[46] A. D. G. Dimakis, A. D. Sarwate, and M. J. Wainwright, "Geographic gossip: Efficient averaging for sensor networks," *IEEE Transactions on Signal Processing*, vol. 56, no. 3, pp. 1205–1216, March 2008.

[47] S. Kar, S. Aldosari, and J. M. F. Moura, "Topology for distributed inference on graphs," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2609–2613, June 2008.

[48] S. Kar and J. M. F. Moura, "Sensor networks with random links: Topology design for distributed consensus," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3316–3326, July 2008.

[49] R. Carli, A. C. L. Schenato, and S. Zampieri, "Distributed kalman filtering based on consensus strategies," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 4, pp. 622–633, May 2008.

[50] U. A. Khan and J. M. F. Moura, "Distributing the kalman filter for large-scale systems," *IEEE Transactions on Signal Processing*, vol. 56, no. 10, pp. 4919–4935, October 2008.

[51] E. Kokiopoulou and P. Frossard, "Polynomial filtering for fast convergence in distributed consensus," *IEEE Transactions on Signal Processing*, vol. 57, no. 1, pp. 342–354, January 2009.

[52] M. Bohge and W. Trappe, "An authentication framework for hierarchical ad hoc sensor networks," in *WiSe '03: Proceedings of the 2nd ACM workshop on Wireless security*.   New York: ACM, 2003, pp. 79–87.

[53] M. Tubaishat and S. Madria, "Sensor networks: an overview," *Potentials, IEEE*, vol. 22, no. 2, pp. 20–23, April 2003.

[54] L. Sankaranarayanan, G. Kramer, and N. B. Mandayam, "Hierarchical sensor networks: capacity bounds and cooperative strategies using the multiple-access relay channel model," October 2004, pp. 191–199.

[55] T. M. Cover and J. A. Thomas, *Elements of Information Theory*.   John Wiley and Sons, 1991.

[56] M. Gastpar and M. Vetterli, "Source-channel communication in sensor networks," *Proceedings of the 2nd International Workshop on Information Processing in Sensor Networks (IPSN'03)*, pp. 162–177, April 2003.

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

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

[59] A. Ribeiro and G. B. Giannakis, "Bandwidth-Constrained Distributed Estimation for Wireless Sensor Networks Part I: Gaussian Case," *IEEE Transactions on Signal Processing*, vol. 54, no. 3, pp. 1131–1143, March 2006.

[60] ——, "Bandwidth-Constrained Distributed Estimation for Wireless Sensor Networks Part II: Unknown Probability Density Function," *IEEE Transactions on Signal Processing*, vol. 54, no. 7, pp. 2784–2796, July 2006.

[61] M. Senel, V. Kapnadak, and E. J. Coyle, "Distributed estimation for cognitive radio networks - the binary symmetric channel case," *Proc. SenSIP Workshop*, 2008.

[62] C. Tepedelenlioglu and A. B. Narasimhamurthy, "Universal distributed estimation over multiple access channels with constant modulus signaling," *IEEE Transactions on Signal Processing*, vol. 58, no. 9, pp. 4783–4794, September 2010.

[63] C. Tepedelenlioglu, M. K. Banavar, and A. Spanias, "On inequalities relating the characteristic function and Fisher information," submitted to the IEEE Transactions on Information Theory. Preprint available online at http://arxiv.org/abs/1007.1483, 2010.

[64] C. Tepedelenlioglu and S. Dasarathan, "Distributed detection over gaussian multiple access channels with constant modulus signaling," submitted to the IEEE Transactions on Signal Processing., 2010.

[65] M. K. Banavar, C. Tepedelenlioğlu, and A. Spanias, "Distributed snr estimation using constant modulus signaling over gaussian multiple-access channels," *2011 IEEE Digital Signal Processing Workshop and IEEE Signal Processing Education Workshop (accepted)*, January 2011.

[66] G. Mergen and L. Tong, "Type based estimation over multiaccess channels," *IEEE Transactions on Signal Processing*, vol. 54, no. 2, pp. 613–626, February 2006.

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

[68] D. L. Hall and J. Llinas, "An introduction to multisensor data fusion," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 6–23, January 1997.

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

[70] J.-F. Chamberland and V. V. Veeravalli, "Decentralized detection in sensor networks," *IEEE Transactions on Signal Processing*, vol. 51, no. 2, pp. 407–416, February 2003.

[71] ——, "Asymptotic results for decentralized detection in power constrained wireless sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 6, pp. 1007–1015, August 2004.

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

[73] K. A. A. Tarzai, S. K. Jayaweera, and V. Aravinthan, "Performance of decentralized detection in a resource-constrained sensor network with non-orthogonal communications," *In Proc. 39th Annual Asilomar Conference on Signals, Systems and Computers*, pp. 437–441, October 2005.

[74] S. K. Jayaweera, "Bayesian fusion performance and system optimization for distributed stochastic Gaussian signal detection under communication constraints," *IEEE Transactions on Signal Processing*, vol. 55, no. 4, pp. 1238–1250, April 2007.

[75] K. Liu, H. E. Gamal, and A. Sayeed, "Decentralized inference over multiple-access channels," *IEEE Transactions on Signal Processing*, vol. 55, no. 7, pp. 3445–3455, July 2007.

[76] T. Wimalajeewa and S. K. Jayaweera, "Optimal power scheduling for correlated data fusion in wireless sensor networks via constrained PSO," *IEEE Transactions on Wireless Communications*, vol. 7, no. 9, pp. 3608–3618, September 2008.

[77] M. K. Banavar, A. D. Smith, C. Tepedelenlioglu, and A. Spanias, "Distributed detection over fading MACs with multiple antennas at the fusion center," Available online at http://arxiv.org/abs/1001.3173, 2010.

[78] Z.-Q. Luo, "Universal decentralized estimation in a bandwidth constrained sensor network," *IEEE Transactions on Information Theory*, vol. 51, no. 6, p. 22102219, June 2005.

[79] ——, "Universal decentralized estimation in a bandwidth constrained sensor network," *IEEE Transactions on Information Theory*, vol. 51, no. 6, pp. 2210–2219, June 2005.

[80] S. Cui, J.-J. Xiao, A. J. Goldsmith, Z.-Q. Luo, and H. V. Poor, "Energy-efficient joint estimation in sensor networks - analog vs. digital," *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 4, pp. 745–748, March 2005.

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

[82] J.-J. Xiao, A. Ribeiro, Z.-Q. Luo, and G. B. Giannakis, "Distributed compression-estimation using wireless sensor networks," *IEEE Signal Processing Magazine*, vol. 23, no. 4, pp. 27–41, July 2006.

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

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

[85] S. Y. Chueng, S. C. Ergen, and P. Varaiya, "Traffic surveillance with wireless magnetic sensors," *Proceedings of the 12th ITS World Congress*, November 2005.

[86] A. Haoui, R. Kavaler, and P. Varaiya, "Wireless magnetic sensors for traffic surveillance, transportation research part C: Emerging technologies," *Emerging Commercial Technologies*, vol. 16, no. 3, pp. 294–306, June 2008.

113

[87]   S. M. Kay, *Fundamentals of Statistical Signal Processing. Vol II: Detection Theory*, A. V. Openheim, Ed.   Prentice Hall Signal Processing Series: Prentice Hall, 1998.

[88]   J. D. Papastavrou and M. Athans, "Distributed detection by a large team of sensors in tandem," *Proceedings of the 29th IEEE Conference on Decision and Control*, pp. 246–251, December 1990.

[89]   D. Cochran, H. Gish, and D. Sinno, "A geometric approach to multi-channel signal detection," *IEEE Transactions on Signal Processing*, vol. 43, no. 9, pp. 2049–2057, September 1995.

[90]   M. Kam, W. Chang, and Q. Zhu, "Hardware complexity of binary distributed detection systems with isolated local bayesian detectors," *IEEE Transaction on Systems, Man and Cybernetics*, vol. 21, no. 3, pp. 565–571, May/June 1991.

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

[92]   X. Zhang, H. V. Poor, and M. Chiang, "Optimal power allocation for distributed detection over MIMO channels in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 56, no. 9, pp. 4124–4140, September 2008.

[93]   W. Li and H. Dai, "Distributed detection in wireless sensor networks using a multiple access channel," *IEEE Transactions on Signal Processing*, vol. 55, no. 3, pp. 822–833, March 2007.

[94]   K. Bai and C. Tepedelenlioğlu, "Distributed detection in UWB wireless sensor networks," *Proceedings of the IEEE ICASSP 2008*, pp. 2261–2264, April 2008.

[95]   J.-G. Chen, N. Ansari, and Z. Siveski, "Distributed detection for cellular CDMA," *Electronics Letters*, vol. 32, no. 3, pp. 169–171, February 1996.

[96]   K. Liu and A. M. Sayeed, "Type-based decentralized detection in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 55, no. 5, pp. 1899–1910, May 2007.

[97]   A. Anandkumar and L. Tong, "Type-based random access for distributed detection over multiaccess fading channels," *IEEE Transactions on Signal Processing*, vol. 55, no. 10, pp. 5032–5043, October 2007.

[98] C. R. Berger, M. Guerriero, S. Zhou, and P. Willett, "PAC vs. MAC for decentralized detection using noncoherent modulation," *IEEE Transactions in Signal Processing*, vol. 57, no. 9, pp. 3562–3575, September 2009.

[99] S. Yiu and R. Schober, "Nonorthogonal transmission and noncoherent fusion of censored decisions," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 1, pp. 263–273, January 2009.

[100] A. D. Smith, M. K. Banavar, C. Tepedelenlioğlu, and A. Spanias, "Distributed estimation over fading macs with multiple antennas at the fusion center," in *Proceedings of the Asilomar Conference on Signals, Systems and Computers*, November 2009, pp. 424–428.

[101] M. K. Banavar, A. D. Smith, C. Tepedelenlioğlu, and A. Spanias, "Distributed detection over fading MACs with multiple antennas at the fusion center," *Proceedings of the IEEE ICASSP 2010*, pp. 2894–2897, March 2010.

[102] T. C. Aysal and K. E. Barner, "Sensor data cryptography in wireless sensor networks," *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 2, pp. 273–289, June 2008.

[103] V. Kapnadak, M. Senel, and E. J. Coyle, "Distributed incumbent estimation for cognitive wireless networks," in *Information Sciences and Systems, 2008. CISS 2008. Proceedings of the 42nd Annual Conference on*, March 2008, pp. 588–593.

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

[105] C. Tepedelenlioğlu, M. K. Banavar, and A. Spanias, "Asymptotic analysis of distributed estimation over fading multiple access channels," in *Conference Record of the Forty-First Asilomar Conference on Signals, Systems and Computers, 2007. ACSSC 2007.*, November 2007, pp. 2140–2144.

[106] M. K. Banavar, C. Tepedelenlioğlu, and A. Spanias, "Performance of distributed estimation over multiple access fading channels with partial feedback," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 2008. ICASSP 2008.*, April 2008, pp. 2253 – 2256.

[107] A. D. Smith, "Distributed parameter estimated using sensor networks within a MIMO scenario," Master's thesis, Arizona State University, Tempe, AZ, USA, 2008.

[108] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. New Jersey: Prentice Hall, 1993.

[109] D. R. Pauluzzi and N. C. Beaulieu, "A comparison of SNR estimation techniques for the AWGN channel," *IEEE Transactions on Communications*, vol. 48, no. 10, pp. 1681 – 1691, October 2000.

[110] I. A. Koutrouvelis, "Regression-type estimation of the parameters of stable laws," *Journal of the American Statistical Association*, vol. 75, no. 372, pp. 918–928, December 1980.

[111] ——, "An iterative procedure for the estimation of the parameters of stable laws," *Communications in Statistics - Simulation and Computation*, vol. 10, no. 1, pp. 17–28, 1981.

[112] A. Feuerverger and P. McDunnough, "On the efficiency of empirical characteristic function procedures," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 43, no. 1, pp. 20–27, 1981.

[113] I. A. Koutrouvelis, "Estimation of location and scale in Cauchy distributions using the empirical characteristic function," *Biometrika*, vol. 69, no. 1, pp. 205–213, April 1982.

[114] R. L. Eubank and V. N. LaRiccia, "Location and scale parameter estimation from randomly censored data," Department of Statistics, Southern Methodist University, Tech. Rep., August 1982.

[115] B. Porat, *Digital processing of random signals: theory and methods*. New Jersey: Prentice-Hall, 1993.

[116] C. Tepedelenlioğlu, A. Abdi, and G. B. Giannakis, "The Ricean *K* factor: estimation and performance analysis," *IEEE Transactions on Wireless Communications*, vol. 2, no. 4, pp. 799–810, July 2003.

[117] A. Goldsmith, *Wireless Communications*, 1st ed. New York: Cambridge University Press, 2005.

[118] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York: Cambridge University Press, 2004.

[119] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*. Courier Dover Publications, 1965.

[120] Y. S. Shmaliy, "Von Mises/Tikhonov-based distributions for systems with differential phase measurement," *Signal Processing*, vol. 85, no. 4, pp. 693–703, April 2004.

[121] I. Gradshteyn and I. Ryzhik, *Table of Integrals, Series and Products*, 7th ed. Academic Press, 2007.

[122] P. Billingsley, *Probabilty and Measure*, 3rd ed. Wiley, 1995.

[123] W. Hoeffding and H. Robbins, "The central limit theorem for dependent random variables," *Duke Math J.*, vol. 15, no. 3, pp. 773–780, 1948.

[124] G. H. Golub and C. F. V. Loan, *Matrix computations*, 3rd ed. Baltimore, MD: John Hopkins University Press, 1996.

[125] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity - Part I: System description," *IEEE Transactions on Communications*, vol. 51, pp. 1927–1938, November 2003.

[126] A. M. Tulino and S. Verdu, *Random Matrix Theory and Wireless Communications*. USA: now Publishers, 2004.

[127] Y. Q. Yin, Z. D. Bai, and P. R. Krishnaiah, "On the limit of the largest eigenvalue of the large dimensional sample covariance matrix," *Probability Theory and Related Fields*, vol. 78, no. 4, pp. 509–521, August 1988.

[128] Z. D. Bai and Y. Q. Yin, "Limit of the smallest eigenvalue of a large dimensional sample covariance matrix," *The Annals of Probability*, vol. 21, no. 3, pp. 1275–1294, 1993.

[129] Z.-Q. Luo, W.-K. Ma, A. M.-C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 20–34, May 2010.

[130] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 1.21," http://cvxr.com/cvx, Jul. 2010.

[131] Z. Zhang, "Inequalities for characteristic functions involving Fisher information," *C. R. Acad. Sci. Paris*, vol. 344, no. 5, pp. 327–330, March 2007.

[132] R. Zamir, "A proof of the Fisher information inequality via a data processing argument," *Information Theory, IEEE Transactions on*, vol. 44, no. 3, pp. 1246–1250, May 1998.

[133] O. Johnson, *Information Theory and the Central Limit Theorem*. London: Imperial College Press, 2004.

[134] E. L. Lehmann and G. Casella, *Theory of Point Estimation*. New York: Springer-Verlag, 1998.

[135] C. Tepedelenlioğlu and A. B. Narasimhamurthy, "Distributed estimation with constant modulus signals over multiple access channels," *Proceedings of ICASSP 2010*, pp. 2290–2293, March 2010.

[136] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, "On the Lambert W function," *Advances in Computational Mathematics*, vol. 5, pp. 329–359, 1996.