1-31-2013

On the spatial predictability of wireless channels and robust networked cooperation in mobile sensor networks

Mehrzad Malmirchegini

Mehrzad Malmirchegini

Candidate

Electrical and Computer Engineering

Department

This dissertation is approved, and it is acceptable in quality and form for publication:

Approved by the Dissertation Committee:

Yasamin Mostofi , Chairperson

Maria Cristina Pereyra

Pradeep Sen

Balu Santhanam

On the spatial predictability of wireless channels and robust networked cooperation in mobile sensor networks

by

Mehrzad Malmirchegini

B.S., Iran University of Science and Technology, 2004
M.S., Sharif University of Technology, 2007

DISSERTATION

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Doctor of Philosophy
Engineering

The University of New Mexico

Albuquerque, New Mexico

December, 2012

Dedication

To my dear parents for their love

Acknowledgments

First of all, I would like to thank my advisor Professor Yasamin Mostofi. It has been an honor for me to be one of her first Ph.D. student. She is a great instructor and an exceptional researcher and most importantly a caring person and a trustable friend. I appreciate all her contributions of time and ideas to make my Ph.D. experience productive and stimulating. The joy and enthusiasm she has for her research was contagious and motivational for me, even during tough times in the Ph.D. pursuit. I am also thankful for the excellent example she has provided as an innovative and successful researcher and a dedicated professor.

I would like to thank Professor Pereyra, Professor Sen and Professor Santhanam for being on my dissertation committee. They are among the most knowledgeable people I have had the chance to meet at UNM. Professor Pereyra has been my instructor for three advanced and graduate-level math classes. She is the one who broadened my perspectives in mathematical analysis and inspired me to apply them in solving engineering problems. I want to thank Professor Sen for agreeing to be on my dissertation committee despite his time constraints. His organized and creative thinking techniques have been always inspiring to me. I also want to thank Professor Santhanam, whose expertise in Stochastic Processing and Signal Processing proved to be useful in my research.

These acknowledgments would be incomplete without mentioning the immense support of all my colleagues at the Cooperative Network Lab at UNM, specially Alireza Ghaffarkhah, Alejandro Gonzalez-Ruiz and Yuan Yan.

Finally, I want to thank my family, specially my father and mother, Moslem and Mahin who raised me with a love of science and supported me in all my pursuits. My love for them is eternal.

On the spatial predictability of wireless channels and robust networked cooperation in mobile sensor networks

by

Mehrzad Malmirchegini

B.S., Iran University of Science and Technology, 2004
M.S., Sharif University of Technology, 2007

Ph.D., University of New Mexico, 2012

Abstract

The wireless sensor network revolution has created the possibility of exploring and controlling the environment in ways not possible before. The vision of a multi-agent network cooperatively learning and adapting in harsh unknown environments to achieve a common goal is closer than ever. In such networks, communication plays a key role in the overall performance of the network as each mobile agent improves its knowledge by processing the information received from others. Therefore, proper prediction of the communication signal strength and *fundamentally understanding the spatial predictability of a wireless channel*, based on only a few measurements, become considerably important. *The first contribution of this thesis* is to propose a framework for predicting the spatial variations of wireless channels and to fundamentally understand wireless channel predictability. This framework can have a significant impact on intelligent connectivity maintenance in mobile sensor networks.

More specifically, in Chapter 2, we develop a probabilistic framework for predicting the channel spatial variations, based on a small number of measurements. By using this framework, we then propose a mathematical foundation for understanding the impact of different environments, in terms of their underlying parameters, on wireless channel predictability. Furthermore, we show how sampling positions can be optimized to improve the prediction quality. Inspired by the recent results in non-uniform sampling theory, we then pursue a different path in Chapter 3 and show how the sparsity of the wireless channel in the frequency domain can be exploited in order to estimate channel variations based on a small number of measurements. The sparsity-based estimator is model-free and independent of the underlying channel parameters. Along this line, we then demonstrate the underlying tradeoffs between these two frameworks and propose an integrated approach which takes advantage of both channel compressibility in the frequency domain and probabilistic characterization in the spatial domain. All the theoretical results are validated with experimental measurements using our robotic testbed.

The second contribution of the thesis is to consider different cooperative network operations with imperfect local sensing and under realistic modeling of communication links. More specifically, we consider the group agreement problem, where the cooperative network is trying to reach consensus on the occurrence of an event, by communicating over fading channels. This problem has received little attention in the literature as compared to the estimation consensus problem. However, it can find several applications such as networked fire detection and cooperative spectrum sensing in cognitive radio networks. Thus, another contribution of this dissertation is to fundamentally understand the behavior of such a cooperative network operation under imperfect communication links. To do so, we propose a novel consensus-seeking protocol that utilizes information of link qualities and noise variances to improve the performance and increase the robustness of the network to local sensing limitations. We mathematically characterize the impact of fading, noise, network connectivity and time-varying topology on consensus performance, which becomes challenging due to all the introduced uncertainties. We consider two different strategies,

in terms of using the available transmissions: fusion and diversity, and shed light on the underlying tradeoffs in terms of speed of convergence and memoryless asymptotic behavior. Motivated by our analysis, we then propose an integrated framework, which keeps the benefits of both diversity and fusion approaches. We mathematically analyze the proposed framework and show how it achieves accurate consensus asymptotically.

Contents

List of Figures

Chapter 1

Introduction

In the past few years, the sensor network revolution has created the possibility of exploring and controlling the environment in ways not possible before [7–9]. The vision of a multi-agent robotic network cooperatively learning and adapting in a harsh unknown environment to achieve a common goal is closer than ever. These cooperative networks arise in many different applications such as target tracking, environmental monitoring, surveillance and security, and military systems. Since each agent has a limited sensing capability, the group relies on networked sensing and decision-making to accomplish the task. Therefore, communication plays a key role in the overall performance of such networks, as each agent improves its sensing qualities by communicating over the network. Thus, maintaining connectivity becomes considerably important in such networks.

In the robotics and control community, considerable progress has been made in the area of networked robotic and control systems [10]. However, ideal or over-simplified models have typically been used to model the communication links among agents. In order to realize the full potentials of these networks, an integrative approach to communication and motion planning issues is essential, i.e., each robot should have an awareness of the impact of its motion decisions on link qualities, when planning its trajectory [6]. This requires that

each robot assesses the quality of the communication link in the locations that it has not yet visited. As a result, proper prediction of the communication signal strength and *fundamentally understanding the spatial predictability of a wireless channel*, based on only a few measurements, become considerably important. In the communications community, rich literature was developed, over the past decades, for the characterization and modeling of wireless channels [11–14]. If all the information about object positions, geometry and dielectric properties is available, ray tracing methods could be used to model the spatial variations of the received signal strength in a given area [15]. However, such approaches require knowing the environment, in terms of locations of the objects and their dielectric properties, which is prohibitive for real-time networked robotic applications. Furthermore, such approaches can not provide the needed fundamental understanding of wireless channel predictability. In this dissertation we tackle this problem. We utilize the probabilistic modeling of the wireless channels as well as their sparsity in the frequency domain and propose two frameworks for the spatial prediction of wireless channels based on sparse channel measurements. The proposed frameworks do not require the knowledge of the environmental features and provide a fundamental understanding of wireless channel predictability in different environments, which is currently missing in the literature. More specifically, we bring a foundational theoretical understanding of the impact of different environments on wireless channel predictability and verify our results with experimental measurements. We furthermore characterize the underlying tradeoffs between these two frameworks, which motivates developing our integrated probabilistic and sparsity-based approach of Chapter 3.

To address cooperative operation with limited local sensing and imperfect communications, we then consider the following cooperative network operation: *binary consensus over fading communication channels*. We propose nonlinear binary consensus-seeking protocols that can account for sensing and communication limitations. We then mathematically characterize the impact of fading, noise, network connectivity and time-varying topology on the consensus performance, which becomes challenging due to all the intro-

duced uncertainties. To improve the performance and robustness of network cooperation, we furthermore consider two different decision-making strategies, in term of using the available transmissions: *fusion* and *diversity*. We characterize the underlying tradeoffs between the two approaches and propose an integrated framework that is aimed at keeping the strengths of both.

1.1 Related work and contributions

In this section, we discuss the related work and contributions of each chapter of the dissertation individually.

Chapter 2: Understanding the spatial predictability of wireless channels

As we mentioned earlier, a mobile cooperative network needs to maintain its connectivity in order to accomplish its task. In order to achieve this, each robot should consider the impact of its motion decisions on its link qualities, when planning its trajectory. This requires each robot to assess the quality of the communication links in the locations that it has not yet visited. As a result, proper prediction of the communication signal strength in a given area, based on only a few measurements, becomes considerably important. As a robot moves around, it can learn the signal strength at positions along its motion trajectory. However, there is simply not enough time to directly measure the channel at every location in the area. Therefore, the spatial variations of a channel should be estimated based on a considerably incomplete data set, which is a challenging task.

If all the information about object positions, geometry and dielectric properties is available, ray tracing methods could be used to find spatial variations of the received signal

strength in the area of interest [15, 16]. Another alternative approach, which is the most accurate and rigorous, is to solve Maxwell's equations for a specific environment numerically [17, 18]. However, such approaches require knowing the environment, in terms of locations of the objects and their dielectric properties, which is prohibitive for real-time networked robotic applications. Furthermore, such approaches can not provide a fundamental understanding of wireless channel predictability.

In general, the spatial variations of a phenomena can be described as a field. We next describe some of the concurrent recent work for general estimation of a general field. Among different statistical models for the spatial variations of a physical phenomenon, Gaussian process has received considerable attention recently. A Gaussian process (or Kriging) has been widely used as a nonlinear regression technique to predict the spatio-temporal variations of physical quantities that obey the Gaussian random field model [19]. In [20], authors introduced a statistical framework to estimate the distribution of a random field and its gradient. Under the assumptions that the field is uncorrelated temporally with a limited-range correlation in space, they developed a distributed kriged Kalman that enables the network to compute the predictive mean functions of the random field and its gradient. In [21], the authors developed a novel class of self-organizing multi-agent systems that perform a given task by exploiting predictive posterior statistics from the recursive estimation of a spatial Gaussian process. They assume that the mean of the Gaussian field can be spatially modeled as a linear combination of Gaussian kernels. In [22], the authors extended that work to spatio-temporal Gaussian processes. There have also been some works on the estimation of other types of spatial fields. For instance, in [23], authors developed a distributed interpolation scheme for deterministic field estimation. The field is assumed to be locally Lipschitz and an iterative implementation of nearest neighbor and inverse distance weighting interpolations has been proposed. In [24], the authors utilized a Gaussian process model to improve cooperation in a cognitive radio network. In [25], authors studied distributed estimation of an unknown acoustic field, which is modeled deterministically by using a number of sparse acoustic sources and distance-dependent de-

cays. They showed that by using an average consensus algorithm and compressive sensing theory, each sensor node can estimate the field in a decentralized manner.

In the wireless communication literature, it is well established that a communication channel between two nodes can be probabilistically modeled as a multi-scale dynamical system with three major dynamics: small-scale fading (multipath), large-scale fading (shadowing) and path loss [11–13]. In Chapter 2 of this dissertation, we utilize this well-established probabilistic modeling and propose a framework to spatially predict and also understand the spatial predictability of wireless channels in different environments. More specifically, we develop a probabilistic framework in which each robot can spatially predict the channel, based on a small number of measurements. We then mathematically characterize the impact of different environments, in terms of their underlying parameters, on channel spatial predictability. We furthermore show how sampling positions can be optimized to improve the prediction quality. We emphasize that we are not suggesting that a wireless channel is fully predictable, as it is not. Rather, our goal is to develop a mathematical characterization of how predictable a wireless channel can be and understand the impact of different environments, in terms of the underlying parameters, on its predictability. As a result, our prediction of channel spatial variations is not going to be perfect, unless several measurements are gathered, but will be informative for applications such as communication-aware motion planning in robotic networks, where a prediction of the link qualities is needed. To the best of our knowledge, this is the first time that a framework has been developed to mathematically characterize and understand the spatial predictability of wireless channels based on a small number of a priori measurements.

Fig. 1.1 (left) shows the simple disk model, which is commonly used in the networked robotic and control literatures. In this model the link quality is assumed above an acceptable threshold in a disk around the transmitter, with no connectivity outside of the disk, as shown in Fig. 1.1 (left). Fig. 1.1 (middle) and Fig. 1.1 (right), for instance, illustrate how the proposed probabilistic framework of Chapter 2 enables a more realistic characteriza-

tion of wireless channels and their connectivity, as compared to the commonly-used disk model of Fig. 1.1 (left). In summary, predicting the spatial variations of a random field,

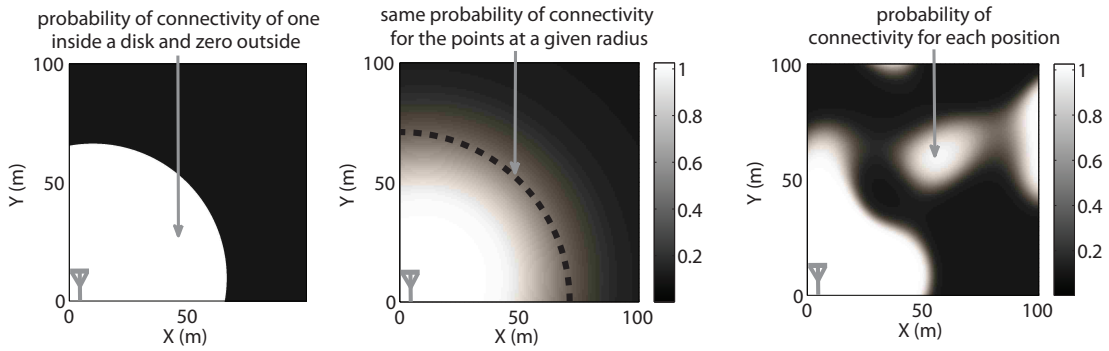

Figure 1.1: Different connectivity models for the communication channel to a fixed transmitter at (0,0) coordinate: (left) simplified disk model that is commonly used in the robotic-network literature (middle) our probabilistic path loss model, and (right) our general probabilistic model.

based on sparse sampling, has recently been of interest in other areas such as meteorology, ecological systems, acoustic field estimation, robotic field sensing, and cognitive radios just to name a few [24, 26–28]. However, to the best of authors' knowledge, no framework has been developed to mathematically understand the spatial predictability of a general random field, or wireless channels in particular, in different environments. As such, the contribution of this chapter is beyond only understanding the spatial predictability of wireless channels and can possibly benefit other areas that require estimation of a random field, based on sparse measurements. For the implication of this framework for other applications not considered in this thesis, readers are referred to other work in our group such as [29] (for work on robotic routers) and [6, 30] (for work on networked surveillance and target tracking) as examples. The results of this chapter are partially published in [31, 32].

Chapter 3: An integrated sparsity and model-based framework for channel prediction

Recently, there has been significant breakthroughs in the area of non-uniform sampling theory [1–3, 33]. The famous Nyquist-Shannon sampling theorem [34] revolutionized several different fields by showing that, under certain conditions, it is indeed possible to reconstruct a uniformly-sampled signal perfectly. The new theory of compressive sampling (also known by other terms such as compressed sensing, compressive sensing or sparse sensing), on the other hand, shows that under certain conditions, it is indeed possible to reconstruct a signal from a considerably incomplete set of observations, i.e. with a number of measurements much less than predicted by the Nyquist-Shannon theorem [1, 3]. This has opened new and fundamentally different possibilities, in terms of estimation and processing, in several different fields such as communications [35–37], signal processing [38–42], and sensor/mobile networks [43–46].

We next review some of the sparsity-related work in the general area of communications that happened concurrent to our work but for a different problem. In recent years, researchers have worked on the sparse modeling and estimation of the impulse response of wibeband channels by utilizing the sparsity of the channel taps (delay spread) in the time domain. They have considered different signaling waveforms used for sensing (e.g., single- or multi-carrier signaling waveforms) and the class to which the underlying channel belongs (e.g., frequency- or doubly-selective channel). For instance, in [37, 47], authors presented a new approach, called compressed channel sensing, to estimate the channel taps of a time-invariant wideband channel. They considered four specific classes of channels, namely, frequency- and doubly-selective single-antenna channels, and nonselective and frequency-selective multiple-antenna channels. Similarly, in [48], authors considered an orthogonal frequency division multiplexed (OFDM) transmission scheme with pilot symbol assisted modulation (PSAM). For a frequency selective channel with large delay spread (but relatively few nonzero delay coefficients), they proposed a compressive-based

approach to estimate the channel. Along a similar line, authors in [49] applied the theory of sparse signal recovery to the problem of estimating the doubly selective channels for multi-carrier communication systems. They showed that this technique reduces the number of required pilots tones. All these work exploit the sparsity of the delay taps in order to estimate time-invariant wideband channels, a problem different from the one we have considered in this thesis.

In Chapter 3, we consider the spatial variations of a narrowband channel to a fixed station. We are interested in mapping the channel spatial variations over a given field of operation, based on a small number of measurements, as is relevant to the robust operation of robotic networks. Our analysis of several real channel measurements show that narrowband wireless channels can be compressible in the frequency domain. Therefore, in Chapter 3, we show how to utilize the sparsity of the communication channels in the frequency domain in order to map the channel variations with a small number of measurements. The proposed sparsity-based estimation is model-free and independent of the underlying channel parameters. We then characterize the underlying tradeoffs between this sparsity-based framework and the probabilistic framework of Chapter 2, using real channel measurements. We show that the probabilistic framework performs well when the channel underlying parameters are estimated with a reasonable quality. However, if these parameters can not be estimated with a reasonable quality, the sparsity-based approach outperforms the probabilistic framework. Motivated by this analysis, we then propose an integrated channel prediction framework. The integrated approach properly takes advantage of both channel compressibility in the frequency domain and channel probabilistic characterization in the spatial domain. The results of this chapter are partially published in [32, 50].

Chapter 4: Binary consensus over fading channels

In recent years, there has been significant interest in cooperative sensing, estimation/detection and control. Consider the case where a group of nodes with sensing, processing, communication and actuation capabilities are given a task to perform jointly. Each node has limited local capabilities and can therefore only accomplish the task in a cooperative manner, in order to get around local sensing errors and limitations. One of the problems that arise in such cooperative networks is that of group agreements. Consensus problems arise when the agents need to reach an agreement on the value of a parameter of interest and can be categorized into two main groups: *estimation consensus* and *detection consensus*.

Estimation consensus refers to the problems where the parameter of interest can take values over an infinite set or an unknown finite set. For instance, it may be of interest that all the mobile agents that started in different directions move in the same direction in a cooperative multi-agent network [10], which requires that they all eventually agree on their headings. These problems received considerable attention in the past few years. In the biological sciences, for instance, there exists a rich literature on generating coordinated behaviors in a group of autonomous agents [51–55]. In the past decade, the control and systems community, in particular, has become interested in the mathematical analysis of consensus behaviors. This interest is sparked, in part, by several emerging applications of networked systems such as vehicle formation [56, 57], flocking [58, 59], rendezvous [60], autonomous vehicles [61], robot position synchronization [62], coupled oscillators [63], and several other applications in distributed sensor networks or distributed control of unmanned air/ground/underwater vehicles. In particular, the control community has applied tools from algebraic graph theory and advanced matrix analysis to characterize estimation consensus problems over graphs that are not fully connected [10, 64–71]. Convergence and equilibrium state of continuous-time and discrete-time consensus protocols have been studied for both time-invariant and time-varying topologies [64–67]. Furthermore, consensus protocols have been applied to formation problems [10, 68–70] as well as distributed

filtering [71]. [72] provides a comprehensive survey of the literature on such consensus problems. The impact of uncertainties introduced by communication links, however, has received less attention in the control literature (except for the work of [73]). More recently, there has been considerable interest in estimation consensus problems from the signal processing and communication community, with more emphasis on link uncertainties [74–78].

By *detection consensus*, on the other hand, we refer to the problems in which the parameter of interest takes values from a finite known set. Then the update protocol that each agent will utilize becomes nonlinear. We refer to a subset of detection consensus problems where the network is trying to reach an agreement over a parameter that can only have two values as *binary consensus* [79]. For instance, networked detection of fire falls into this category. While there exists a rich literature on estimation consensus, detection consensus problems only recently started to receive attention. In [80], the authors consider convergence in a detection consensus setup over perfect channels, with repeated sensing and known probabilistic sensing models. In [81], the authors consider a distributed hypothesis testing problem over perfect communication channels, to which they refer to as belief consensus. They consider the case where each node transmits its belief (conditional probability) to other nodes. As a result, their problem immediately takes the form of the traditional average estimation consensus, for which a rich literature exists. In [82–86] and references therein, quantized consensus problems are considered over perfect channels, in which every node can only send from a set of quantized values.

In [79], [87] and [88], *binary consensus* problems are considered, where the nodes start with an initial decision regarding the occurrence of an event. Through repeated communications, the goal for every node is to reach the majority of the initial votes, without knowing anything about the sensing qualities. [88] considered and characterized phase transition of such a binary consensus problem in the presence of a uniformly-distributed communication noise. Since the support of the probability density function of this noise

is bounded, there exists a transition point beyond which consensus will be guaranteed in this case. In [88], this transition point is characterized. In most applications, however, the agents will communicate their values wirelessly and will experience Gaussian receiver noise as opposed to a uniformly-distributed noise. In [79, 87, 89], authors considered reaching binary consensus over time-invariant network topologies with Additive White Gaussian Noise (AWGN) channels and regular graphs (all nodes have the same number of neighbors). Since the noise is not bounded in this case, there is no transition point beyond which consensus is guaranteed. Instead, they proposed a probabilistic approach to characterize the asymptotic and transient behavior of the network.

In Chapter 4, we consider binary consensus over *not fully-connected and rapidly-changing network topologies* with fading channels, where the goal of every node is to reach the majority of the initial votes. We mathematically characterize the impact of fading, noise, network connectivity and time-varying topology on consensus performance, which becomes challenging due to all the introduced uncertainties. To improve the performance and robustness of network cooperation, we propose novel consensus-seeking protocols that utilize information of link qualities and noise variances. Furthermore, we consider two different decision-making strategies, in term of using the available transmissions: *fusion* and *diversity*. In the first approach, the given resources are used to increase the flow of information in the network whereas the second strategy aims to increase robustness to link error by channel coding. There exist interesting tradeoffs between these two approaches in terms of speed of convergence and asymptotic behavior, as we shall explore. Our proposed framework builds a foundation for understanding both the asymptotic and transient behaviors of binary consensus in fading environments. While channel uncertainty can result in an undesirable asymptotic behavior, depending on the utilized decision-making strategy, we show that the network can still be in consensus for a long period of time (enough for practical purposes) with high probability. In order to characterize the transient behavior, we derive a tight approximation for the second largest eigenvalue of the average of the underlying linear dynamical system. The derived expressions show

how channel uncertainty and network topology affect binary consensus and shed light on the underlying tradeoffs. The results of this chapter are partially published in [90, 91].

Chapter 5: An integrated framework for binary consensus over time-invariant network topologies

In Chapter 5, we extend our analysis of Chapter 4 and consider the binary consensus problem over a general time-invariant network topology (not necessarily fully connected) with fading channels. More specifically, we consider the underlying tradeoffs of the fusion and diversity strategies that we analyzed in Chapter 4. The *main contribution* of this chapter is then to propose a framework that keeps the benefits of both fusion and diversity strategies, in terms of network information flow and link error robustness, for binary consensus over time-invariant network topologies with fading channels. We mathematically analyze the proposed framework and show that it achieves accurate consensus asymptotically. As an example, we then utilize the proposed framework over regular ring lattice networks. Our theoretical and simulation results indicate that the proposed technique improves the consensus performance considerably and overcomes the memoryless asymptotic behavior of the original problem. The results of this chapter are partially presented in [92].

Chapter 2

Understanding the spatial predictability of wireless channels

In the wireless communication literature, it is well established that a communication channel between two nodes can be probabilistically modeled as a multi-scale dynamical system with three major dynamics: small-scale fading (multipath), large-scale fading (shadowing) and path loss [11–13]. Fig. 2.1, for instance, shows the received signal power across a route in the basement of the ECE building. The three main dynamics are marked on the figure. The measured received signal is the small-scale fading. In order to extract the large-scale component, the received signal should be averaged locally over a distance of 5λ to 40λ (depending on the scenario), where λ is the transmission wavelength [12, 93]. In the example of Fig. 2.1, for instance, we averaged the channel locally over the length of $5\lambda = 62.5$cm, by using a moving average (frequency of operation is 2.4GHz). Once we have the large-scale component, the distance-dependent path loss is calculated by finding the best line fit to the log of the received measurements [11, 12, 94].

In this chapter we utilize such a probabilistic link model and fundamentally characterize the spatial predictability of wireless channels at unvisited locations, based on a few a

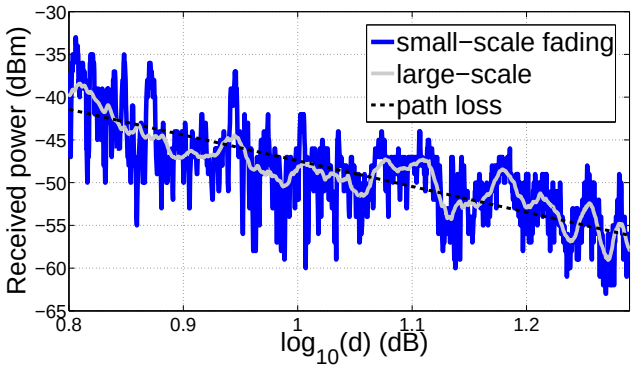

Figure 2.1: Underlying dynamics of the received signal power across a route in the basement of ECE building.

priori channel measurements. We then mathematically characterize the impact of different environments, in terms of their underlying parameters, on channel spatial predictability. Furthermore, we show the optimum distribution of the sparse sampling positions in order to maximize channel predictability.

This chapter is organized as follows. In Section 2.1, we describe our proposed probabilistic channel prediction framework. In Section 2.2, we mathematically characterize the impact of different underlying channel parameters on the prediction performance, assuming perfectly-estimated path loss parameters. In Section 2.3, we mathematically characterize the impact of different environments and sampling positions on the estimation of path loss parameters and show how to optimize the positions of the sparse samples. In Section 2.4, we analyze the impact of different environments on the estimation of shadowing power. Then, in Section 2.5, we extend the analysis of Section 2.2 to characterize wireless channel predictability in the presence of path loss estimation error. In Section 2.6, we show the performance of the proposed framework in predicting (and understanding the predictability of) the spatial variations of real channels, using several measurements in our building. A summary of the results of the chapter is provided in Section 2.7.

2.1 Probabilistic prediction of channel spatial variations

As mentioned in the previous section, a communication channel between two nodes can be modeled as a multi-scale dynamical system with three major dynamics: small-scale fading (multipath), large-scale fading (shadowing) and path loss. Let $\Upsilon_{\text{RX}}(q)$ denote the received signal strength (power), in the transmission from a fixed transmitter at $q_b \in \mathcal{K}$ to a mobile node at $q \in \mathcal{K}$, where $\mathcal{K} \subset \mathbb{R}^2$ denotes the workspace. Consider the case where the channel to the fixed transmitter is narrowband. Furthermore, assume that the workspace is not changing with time, i.e. the environmental features that impact the wireless transmission in the workspace are time-invariant. Our proposed framework can be extended to time-varying environments, as we briefly discuss later in this section. Then, we have the following at the output of the power detector: $\Upsilon_{\text{RX}}(q) = g(q)P_T + \varrho$, where P_T and $g(q)$ denote the transmitted power and channel gain (square of the amplitude of the baseband equivalent channel), at position q, respectively and ϱ represents the power of the receiver thermal noise [11]. Define $\Upsilon(q) \triangleq \Upsilon_{\text{RX}}(q) - \varrho$. We assume that the receiver can estimate and remove the noise power to obtain $\Upsilon(q)$.[1] $\Upsilon(q)$ is proportional to $g(q)$ and can be modeled as a multi-scale dynamical system with three major dynamics: multipath fading, shadowing and path loss. We can then characterize $\Upsilon(q)$ by a 2D non-stationary random field with the following form [11]: $\Upsilon(q) = \Upsilon_{\text{PL}}(q)\Upsilon_{\text{SH}}(q)\Upsilon_{\text{MP}}(q)$, where $\Upsilon_{\text{MP}}(q)$ and $\Upsilon_{\text{SH}}(q)$ are random variables representing the impact of multipath fading and shadowing components respectively and $\Upsilon_{\text{PL}}(q) = \frac{K_{\text{PL}}}{\|q-q_b\|^{n_{\text{PL}}}}$ is the distance-dependent path loss.[2] In this model, the multipath fading coefficient, $\Upsilon_{\text{MP}}(q)$, has a unit average. Let

[1] Most related device drivers provide an estimate on the noise power. MadWiFi, for instance, estimates the noise power by using the often-used formula of $K_{\text{Bol}} \times T_{\text{env}} \times BW$ [95], where K_{Bol} is the Boltzmann's constant, T_{env} is the environment temperature and BW is the utilized bandwidth. Its newer versions can even provide a better online assessment, by using the measurements from the silent mode (when no transmission) [96].

[2] We follow the convention of [12] and use the term "shadowing" to refer to the large-scale fading after its mean (path loss) is removed in the dB domain. More specifically, Υ_{SH} is the large-scale fading after its average (path loss) is removed in the dB domain. Furthermore, we use the term "multipath fading" to refer to the normalized small-scale fading, i.e. with unit average. Then,

$\Upsilon_{\text{dB}}(q) = 10\log_{10}\big(\Upsilon(q)\big)$ represent the received signal strength in dB. We have

$$\Upsilon_{\text{dB}}(q) = \underbrace{10\log_{10}\big(K_{\text{PL}}\big) + \overline{\Upsilon}_{\text{MP, dB}}}_{K_{\text{dB}}} - 10n_{\text{PL}}\log_{10}\big(\|q - q_b\|\big) + \nu(q) + \omega(q), \qquad (2.1)$$

where $\overline{\Upsilon}_{\text{MP, dB}} = 10\,\mathbb{E}\Big\{\log_{10}\big(\Upsilon_{\text{MP}}(q)\big)\Big\}$ is the average of the multipath fading in dB, $\nu(q) = 10\log_{10}\big(\Upsilon_{\text{SH}}(q)\big)$ is a zero-mean random variable representing the shadowing effect in dB and $\omega(q) = 10\log_{10}\big(\Upsilon_{\text{MP}}(q)\big) - \overline{\Upsilon}_{\text{MP, dB}}$ is a zero-mean random variable, independent of $\nu(q)$, which denotes the impact of multipath fading in dB, after removing its average. In the communication literature, the distributions of $\Upsilon_{\text{MP}}(q)$ and $\Upsilon_{\text{SH}}(q)$ (or equivalently the distributions of $\omega(q)$ and $\nu(q)$) are well established based on empirical data [13]. For instance, Nakagami distribution is shown to be a good match for the distribution of $\Upsilon_{\text{MP}}(q)$ in several environments [11]. In this case, we have the following Nakagami distribution, with parameter m and unit average, for the distribution of $\Upsilon_{\text{MP}}(q)$: $f_{\Upsilon_{\text{MP}}}(x) = \frac{m^m x^{m-1}}{\text{Gamma}(m)}\, e^{-mx}$, where Gamma$(.)$ represents the Gamma function. This then results in the following distribution for $\omega(q)$: $f_\omega(x) = \frac{\ln(10)}{10} 10^{(x+\overline{\Upsilon}_{\text{MP, dB}})/10} f_{\Upsilon_{\text{MP}}}\big(10^{(x+\overline{\Upsilon}_{\text{MP, dB}})/10}\big)$. Some experimental measurements have also suggested Gaussian to be a good enough yet simple fit for the distribution of $\omega(q)$ [97]. We will take advantage of this Gaussian simplification later in our framework. As for the shadowing variable, log-normal is shown to be a good match for the distribution of $\Upsilon_{\text{SH}}(q)$. Then, we have the following zero-mean Gaussian pdf for the distribution of $\nu(q)$: $f_\nu(x) = \frac{1}{\sqrt{2\pi\alpha}}\, e^{-x^2/2\alpha}$, where α is the variance of the shadowing variations around path loss.

Characterizing the spatial correlation of $\omega(q)$ and $\nu(q)$ is also considerably important for our model-based channel prediction framework. However, we do not attempt to predict the multipath component, $\omega(q)$, due to the fact that it typically decorrelates fast and that the form of its correlation function can change considerably, depending on the angle of arrival and position of the scatterers. Therefore, in our proposed framework we only predict the path loss and shadowing components of the channel. The impact of multipath will then

Υ_{MP} is the normalized small-scale fading.

appear in the characterization of the prediction error variance, as we shall see. As for the spatial correlation of shadowing, [98] characterizes an exponentially-decaying spatial correlation function, which is widely used: $\mathbb{E}\Big\{\nu(q_1)\nu(q_2)\Big\} = \alpha e^{-\|q_1-q_2\|/\beta}$, for $q_1, q_2 \in \mathcal{K}$ where α denotes the *shadowing power* and the *correlation distance*, β, controls the spatial correlation of the channel [98].

Next, we describe our proposed model-based channel prediction framework. Consider the case where a wireless channel to a fixed transmitter is sparsely sampled at positions $\mathcal{Q} = \{q_1, q_2, \cdots, q_k\} \subset \mathcal{K}$, in a given environment. These channel measurements can be gathered by one or a number of cooperative homogenous robots, equipped with identical receivers, making measurements along their trajectories. Let a region or an environment refer to an area over which the underlying channel parameters, such as α and β, can be considered constant. The four marked areas of Fig. 2.11 are examples of such regions. First, consider the case that all the k measurements belong to one region and that we are predicting the channel in the same region. We show how to relax this assumption later in this section. Let $D_{\mathcal{Q}}$ and $Y_{\mathcal{Q}} = [y_1, \cdots, y_k]^T \in \mathbb{R}^k$ denote the corresponding distance vector to the transmitter in dB and the vector of all the available channel measurements (in dB) respectively: $D_{\mathcal{Q}} = \Big[10\log_{10}(\|q_1 - q_b\|), \cdots, 10\log_{10}(\|q_k - q_b\|)\Big]^T$ and $Y_{\mathcal{Q}} = [y_1, \cdots, y_k]^T \in \mathbb{R}^k$. We have,

$$Y_{\mathcal{Q}} = \underbrace{\Big[1_k \quad -D_{\mathcal{Q}}\Big]}_{H_{\mathcal{Q}}} \theta + \vartheta_{\mathcal{Q}} + \Omega_{\mathcal{Q}}, \tag{2.2}$$

where 1_k denotes the vector of ones with the length of k, $\theta = [K_{\text{dB}} \;\; n_{\text{PL}}]^T$ is the vector of the path loss parameters, $\vartheta_{\mathcal{Q}} = \Big[\nu_1, \cdots, \nu_k\Big]^T$ with $\nu_i = \nu(q_i)$ and $\Omega_{\mathcal{Q}} = \Big[\omega_1, \cdots, \omega_k\Big]^T$ with $\omega_i = \omega(q_i)$, for $i = 1, \cdots, k$. Based on the log-normal model for shadowing, $\vartheta_{\mathcal{Q}}$ is a zero-mean Gaussian random vector with the covariance matrix $R_{\mathcal{Q}} \in \mathbb{R}^{k\times k}$, where $\Big[R_{\mathcal{Q}}\Big]_{i,j} = \alpha\, e^{-\|q_i-q_j\|/\beta}$, for $q_i, q_j \in \mathcal{Q}$. The term $\Omega_{\mathcal{Q}}$ denotes the impact of multipath fading in dB domain. As mentioned earlier, some empirical data have shown Gaussian to be a good match for the distribution of w_i [97]. For instance, Fig. 2.2 compares the match of

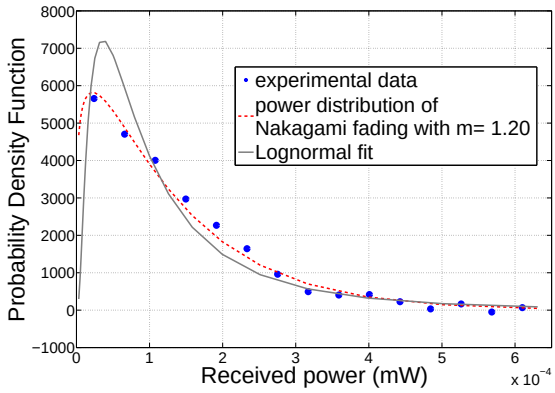

Figure 2.2: Comparison of Nakagami and lognormal for the distribution of small-scale fading.

both Nakagami and lognormal to the distribution of multipath fading (Υ_{MP}) for a stationary section of our collected data of Fig. 2.1. As can be seen, Nakagami provides a considerably good match while lognormal can be acceptable, depending on the required accuracy. Thus, in order to facilitate the mathematical derivations in our prediction framework, we take w_i to have a Gaussian distribution. In addition, multipath fading typically decorrelates considerably fast, making learning of its correlation function, based on sparse possibly non-localized samples, considerably challenging if not infeasible. There is also no one general function that can properly model its correlation in all the environments as its form depends heavily on the angle of arrival and position of the scatterers. While approaches based on the estimation of the power spectrum and linear prediction have been utilized to predict the immediate values of multipath, based on past observations, such approaches require dense sampling in order to capture correlated multipath samples. Finally, even if its correlation function is learned, it typically can not be taken advantage of, in the prediction framework, unless the location of the channel to be predicted is very close to the position of one of the available measurements. Thus, we take $\Omega_{\mathcal{Q}}$ to be an uncorrelated zero-mean Gaussian vector with the covariance of $\mathbb{E}\Big\{\Omega_{\mathcal{Q}}\Omega_{\mathcal{Q}}^T\Big\} = \sigma^2 I_{k\times k}$, where $I_{k\times k}$ is a $k \times k$ identity matrix and $\sigma^2 = \mathbb{E}\big\{\omega^2(q)\big\} = 100\int_0^\infty \log_{10}^2(x) f_{\Upsilon_{\text{MP}}}(x)dx - 100\Big(\int_0^\infty \log_{10}(x) f_{\Upsilon_{\text{MP}}}(x)dx\Big)^2$ is the power of multipath fading (in dB domain). In other words, our framework does not

attempt to predict the multipath component and assumes the worst case of uncorrelated multipath (worst from a prediction standpoint). The estimated variance of multipath then appears in our assessment of channel prediction error variance, as we shall see. Note, however, that this is only for the purpose of our modeling. When we show the performance of this framework, we use real measurements where the multipath component will have its natural distribution and correlation function.

We then define $\Xi_{\mathcal{Q}} \triangleq \vartheta_{\mathcal{Q}} + \Omega_{\mathcal{Q}}$, which is a zero-mean Gaussian vector with the covariance matrix of $R_{\text{tot},\mathcal{Q}} \triangleq R_{\mathcal{Q}} + \sigma^2 I_{k\times k}$. In our model-based probabilistic framework, we first need to estimate the parameters of the model (θ,α, β and σ^2) and then use these parameters to estimate the channel. Let $f_{Y_{\mathcal{Q}}}(Y_{\mathcal{Q}}|\theta,\alpha,\beta,\sigma^2)$ denote the conditional pdf of $Y_{\mathcal{Q}}$, given the parameters θ, α, β and σ^2. Under the assumption of independent multipath fading variables, Eq. 2.2 will result in the following:

$$f_{Y_{\mathcal{Q}}}(Y_{\mathcal{Q}}|\theta,\alpha,\beta,\sigma^2) = \frac{1}{(2\pi)^{k/2}\Big(\det\big[\alpha R_{\text{norm},\mathcal{Q}}(\beta)+\sigma^2 I_{k\times k}\big]\Big)^{1/2}} e^{-\frac{1}{2}\left(Y_{\mathcal{Q}}-H_{\mathcal{Q}}\theta\right)^T\left(\alpha R_{\text{norm},\mathcal{Q}}(\beta)+\sigma^2 I_{k\times k}\right)^{-1}\left(Y_{\mathcal{Q}}-H_{\mathcal{Q}}\theta\right)}, \tag{2.3}$$

where $R_{\text{norm},\mathcal{Q}} = \frac{1}{\alpha}R_{\mathcal{Q}}$ denotes the normalized version of $R_{\mathcal{Q}}$. Next, we characterize the Maximum Likelihood (ML) estimation of the underlying channel parameters.

$$\begin{aligned}
&[\hat{\theta}_{\text{ML}},\hat{\alpha}_{\text{ML}},\hat{\beta}_{\text{ML}},\hat{\sigma}^2_{\text{ML}}] = \text{argmax}_{\theta,\alpha,\beta,\sigma^2}\ \ln\left(f_{Y_{\mathcal{Q}}}(Y_{\mathcal{Q}}|\theta,\alpha,\beta,\sigma^2)\right)\\
&= \text{argmin}_{\theta,\alpha,\beta,\sigma^2}\ \left(Y_{\mathcal{Q}}-H_{\mathcal{Q}}\theta\right)^T\left(\alpha R_{\text{norm},\mathcal{Q}}(\beta)+\sigma^2 I_{k\times k}\right)^{-1}\left(Y_{\mathcal{Q}}-H_{\mathcal{Q}}\theta\right)\\
&+\ln\Big(\det\big[\alpha R_{\text{norm},\mathcal{Q}}(\beta)+\sigma^2 I_{k\times k}\big]\Big),
\end{aligned}$$

which results in:

$$\hat{\theta}_{\text{ML}} = \Big(H^T_{\mathcal{Q}}\big(\hat{\alpha}_{\text{ML}}R_{\text{norm},\mathcal{Q}}(\hat{\beta}_{\text{ML}})+\hat{\sigma}^2_{\text{ML}}\big)^{-1}H_{\mathcal{Q}}\Big)^{-1}H^T_{\mathcal{Q}}\big(\hat{\alpha}_{\text{ML}}R_{\text{norm},\mathcal{Q}}(\hat{\beta}_{\text{ML}})+\hat{\sigma}^2_{\text{ML}}\big)^{-1}Y_{\mathcal{Q}}. \tag{2.4}$$

Finding a closed-form expression for $\hat{\alpha}_{\text{ML}}, \hat{\beta}_{\text{ML}}$ and $\hat{\sigma}^2_{\text{ML}}$, however, is challenging. For the special case where Ω_k is negligible, the ML estimation of channel parameters can be simplified to:

$$\hat{\theta}_{\text{ML},\sigma^2=0} = \Big(H_{\mathcal{Q}}^T R^{-1}_{\text{norm},\mathcal{Q}}(\hat{\beta}_{\text{ML},\sigma^2=0}) H_{\mathcal{Q}}\Big)^{-1} H_{\mathcal{Q}}^T R^{-1}_{\text{norm},\mathcal{Q}}(\hat{\beta}_{\text{ML},\sigma^2=0}) Y_{\mathcal{Q}},$$

$$\begin{aligned}
\hat{\alpha}_{\text{ML},\sigma^2=0} &= \frac{1}{k}\big(Y_{\mathcal{Q}} - H_{\mathcal{Q}}\hat{\theta}_{\text{ML},\sigma^2=0}\big)^T R^{-1}_{\text{norm},\mathcal{Q}}(\hat{\beta}_{\text{ML},\sigma^2=0})\big(Y_{\mathcal{Q}} - H_{\mathcal{Q}}\hat{\theta}_{\text{ML},\sigma^2=0}\big), \\
\hat{\beta}_{\text{ML},\sigma^2=0} &= \arg\min_{\beta} \quad \Big[Y_{\mathcal{Q}}^T P^T_{\mathcal{Q},\text{ML}}(\beta) R^{-1}_{\text{norm},\mathcal{Q}}(\beta) P_{\mathcal{Q},\text{ML}}(\beta) Y_{\mathcal{Q}}\Big]^k \det\big[R_{\text{norm},\mathcal{Q}}(\beta)\big], \qquad (2.5)
\end{aligned}$$

where $P_{\mathcal{Q},\text{ML}}(\beta) = I_{k\times k} - H_{\mathcal{Q}}\Big(H_{\mathcal{Q}}^T R^{-1}_{\text{norm},\mathcal{Q}}(\beta) H_{\mathcal{Q}}\Big)^{-1} H_{\mathcal{Q}}^T R^{-1}_{\text{norm},\mathcal{Q}}(\beta)$. Under the assumption that β is known, it can be shown that $\hat{\theta}_{\text{ML},\sigma^2=0}$ is an unbiased estimator and achieves the Cramer-Rao bound. Furthermore, for large number of sampling points k, we can show that $\hat{\alpha}_{\text{ML},\sigma^2=0}$ is unbiased and achieves the Cramer-Rao bound as well. We will provide the detailed proof in Section 2.4. The ML estimator will therefore be our benchmark in the estimation of the channel parameters.

As can be seen, in order to estimate θ and α, we first need to estimate β, which is challenging. Furthermore, finding the ML estimation of the channel parameters for the general case, where $\sigma^2 \neq 0$, is computationally complex. Therefore, we next devise a suboptimum but simpler estimation strategy. Let $\chi = \alpha + \sigma^2$ denote the sum of the shadowing and multipath powers. A Least Square (LS) estimation of θ and χ then results in:

$$\hat{\theta}_{\text{LS}} = \big(H_{\mathcal{Q}}^T H_{\mathcal{Q}}\big)^{-1} H_{\mathcal{Q}}^T Y_{\mathcal{Q}}, \qquad (2.6)$$

$$\begin{aligned}
\hat{\chi}_{\text{LS}|\theta=\hat{\theta}_{\text{LS}}} &= \frac{1}{k} Y_{\mathcal{Q}}^T \Big(I_{k\times k} - H_{\mathcal{Q}}\big(H_{\mathcal{Q}}^T H_{\mathcal{Q}}\big)^{-1} H_{\mathcal{Q}}^T\Big)^2 Y_{\mathcal{Q}} \\
&= \frac{1}{k} Y_{\mathcal{Q}}^T \Big(I_{k\times k} - H_{\mathcal{Q}}\big(H_{\mathcal{Q}}^T H_{\mathcal{Q}}\big)^{-1} H_{\mathcal{Q}}^T\Big) Y_{\mathcal{Q}}, \qquad (2.7)
\end{aligned}$$

where $H_{\mathcal{Q}}$ is full rank, except for the case where the samples are equally-distanced from the transmitter. Since such a special case is very low probable, we assume that $H_{\mathcal{Q}}$ is

full rank unless otherwise is stated. We refer to this suboptimal approach as LS approach. We next discuss a more practical but suboptimum strategy to estimate β. Let $\mathcal{I}_l = \Big\{(i,j)\big|q_i, q_j \in \mathcal{Q} \text{ such that } \|q_i - q_j\| = l\Big\}$ denote the pairs of points in $\mathcal{Q}$ which are located at distance l from each other. Let $Y_{\mathcal{Q},\text{cent,LS}} = \Big(I_{k\times k} - H_{\mathcal{Q}}\big(H_{\mathcal{Q}}^T H_{\mathcal{Q}}\big)^{-1} H_{\mathcal{Q}}^T\Big) Y_{\mathcal{Q}}$ represent the centered version of the measurement vector, when path loss parameters are estimated using the LS estimator of Eq. 2.6. Define $\hat{r}_{\mathcal{Q}}(l) \triangleq \frac{1}{|\mathcal{I}_l|}\sum_{(i,j)\in\mathcal{I}_l}[Y_{\mathcal{Q},\text{cent,LS}}]_i[Y_{\mathcal{Q},\text{cent,LS}}]_j$ to be the numerical estimate of the spatial correlation function at distance l, where $|.|$ represents the cardinality of the argument set and $[.]_i$ denotes the ith element of the argument vector. We have $[\hat{\alpha}_{\text{LS}}, \hat{\beta}_{\text{LS}}] = \arg\min_{\alpha,\beta}\sum_{l\in\mathcal{L}_{\mathcal{Q}}} w(l)\Big[\ln\big(\alpha e^{-l/\beta}\big) - \ln\big(\hat{r}_{\mathcal{Q}}(l)\big)\Big]^2$, where $\mathcal{L}_{\mathcal{Q}} = \{l | 0 < \hat{r}_{\mathcal{Q}}(l) < \hat{\chi}_{\text{LS}|\theta=\hat{\theta}_{\text{LS}}}\}$ and $w(l)$ can be chosen based on our assessment of the accuracy of the estimation of $\hat{r}_{\mathcal{Q}}(l)$. For instance, if we have very few pairs of measurements at a specific distance, then the weight should be smaller. Let $\mathcal{L}_{\mathcal{Q}} = \{l_1, l_2, \cdots, l_{|\mathcal{L}_{\mathcal{Q}}|}\}$ denote an ordered set of all the possible distances among the measurement points. We have the following Least Square estimator of α and β: $\begin{bmatrix}\ln(\hat{\alpha}_{\text{LS}})\\ \frac{1}{\hat{\beta}_{\text{LS}}}\end{bmatrix} = (M_{\mathcal{L}_{\mathcal{Q}}}^T W_{\mathcal{L}_{\mathcal{Q}}} M_{\mathcal{L}_{\mathcal{Q}}})^{-1} M_{\mathcal{L}_{\mathcal{Q}}}^T W_{\mathcal{L}_{\mathcal{Q}}} b$ where $M_{\mathcal{L}_{\mathcal{Q}}} = \begin{bmatrix}1 & -l_1\\ \vdots & \vdots\\ 1 & -l_{|\mathcal{L}_{\mathcal{Q}}|}\end{bmatrix}, b = \begin{bmatrix}\ln\big(\hat{r}_{\mathcal{Q}}(l_1)\big)\\ \vdots\\ \ln\big(\hat{r}_{\mathcal{Q}}(l_{|\mathcal{L}_{\mathcal{Q}}|})\big)\end{bmatrix}$ and $W_{\mathcal{L}_{\mathcal{Q}}} = \text{diag}[w(l_1), \cdots, w(l_{|\mathcal{L}_{\mathcal{Q}}|})]$. We then have, $\hat{\sigma}^2_{\text{LS}} = \hat{\chi}_{\text{LS}|\theta=\hat{\theta}_{\text{LS}}} - \hat{\alpha}_{\text{LS}}$ for the estimation of the multipath power (in dB domain). Note that the estimated values of the shadowing parameters should satisfy: $0 < \hat{\alpha}_{\text{LS}} \leq \hat{\chi}_{\text{LS}|\theta=\hat{\theta}_{\text{LS}}}$ and $\hat{\beta}_{\text{LS}} > 0$. If due to the lack of enough measurements, any of these are violated, we take $\hat{\alpha}_{\text{LS}}$ and $\hat{\beta}_{\text{LS}}$ to be zero. This means that, in this case, we can not estimate the correlated part of the channel.

Once the underlying parameters of our model are estimated, channel at position $q \in \mathcal{K}$ can be estimated as follows. We have the following for the probability distribution of $\Upsilon_{\text{dB}}(q)$, conditioned on all the gathered measurements and the underlying parameters:

$f(\Upsilon_{\text{dB}}(q)|Y_{\mathcal{Q}},\theta,\alpha,\beta,\sigma^2) \sim \mathcal{N}\big(\tilde{\Upsilon}_{\text{dB},\mathcal{Q}}(q), \sigma^2_{\text{dB},\mathcal{Q}}(q)\big)$ with

$$\tilde{\Upsilon}_{\text{dB},\mathcal{Q}}(q) \triangleq \mathbb{E}\Big\{\Upsilon_{\text{dB}}(q) \;\Big|\; Y_{\mathcal{Q}},\theta,\alpha,\beta,\sigma^2\Big\} = h^T(q)\theta + \phi^T_{\mathcal{Q}}(q)R^{-1}_{\text{tot},\mathcal{Q}}\big(Y_{\mathcal{Q}} - H_{\mathcal{Q}}\theta\big) \text{ and}$$
$$\sigma^2_{\text{dB},\mathcal{Q}}(q) \triangleq \mathbb{E}\Big\{\big(\Upsilon_{\text{dB}}(q) - \tilde{\Upsilon}_{\text{dB},\mathcal{Q}}(q)\big)^2 \;\Big|\; \theta,\alpha,\beta,\sigma^2\Big\} = \alpha + \sigma^2 - \phi^T_{\mathcal{Q}}(q)R^{-1}_{\text{tot},\mathcal{Q}}\phi_{\mathcal{Q}}(q), \tag{2.8}$$

where $h(q) = \begin{bmatrix}1 & -D_{\{q\}}\end{bmatrix}^T$, $D_{\{q\}} = 10\log_{10}\big(\|q-q_b\|\big)$ and $\phi_{\mathcal{Q}}(q)$ denotes the cross covariance between $\mathcal{Q}$ and q, i.e. $\phi_{\mathcal{Q}}(q) = \alpha\big[e^{-\frac{\|q_1-q\|}{\beta}},\cdots,e^{-\frac{\|q_k-q\|}{\beta}}\big]^T$. Therefore, the Minimum Mean Square Error (MMSE) estimation of $\Upsilon_{\text{dB}}(q)$, assuming perfect estimation of the underlying parameters, is given by $\tilde{\Upsilon}_{\text{dB},\mathcal{Q}}(q)$. We then have the following by considering the true estimated parameters: $\hat{\Upsilon}_{\text{dB},\mathcal{Q}}(q) = h^T(q)\hat{\theta} + \hat{\phi}^T_{\mathcal{Q}}(q)\hat{R}^{-1}_{\text{tot},\mathcal{Q}}\big(Y_{\mathcal{Q}} - H_{\mathcal{Q}}\hat{\theta}\big)$, where $\hat{\phi}_{\mathcal{Q}}(q) = \Big[\hat{\alpha}\,e^{-\|q-q_1\|/\hat{\beta}},\cdots,\hat{\alpha}\,e^{-\|q-q_k\|/\hat{\beta}}\Big]^T$ and $\hat{R}_{\text{tot},\mathcal{Q}} = \hat{\alpha}R_{\text{norm},\mathcal{Q}}(\hat{\beta}) + \hat{\sigma}^2 I_{k\times k}$.

The prediction quality at position q improves, the more correlated the available channel measurements become with the value of the channel at position q. In order to mathematically assess this, the next lemma characterizes the average number of the available measurements at the β neighborhood of the point to be predicted, for the case of randomly-distributed available channel measurements in 1D. The β-neighborhood of a point q, in the workspace $\mathcal{K}$, is defined as $\{z \in \mathcal{K} | d(z,q) < \beta\}$, where $d(z,q)$ denotes the Euclidian distance between points z and q.

Lemma 1. *Consider the case that k channel measurements, at positions $\{q_1, q_2, \cdots, q_k\}$ are available, for predicting the channel at point q. Let $N_\beta(Q,q)$ represent the number of points in $Q = \{q_i\}_{i=1}^k$, which are located in the β neighborhood of q, where q and $\{q_i\}_{i=1}^k$ are i.i.d. random positions, uniformly distributed over the workspace $\mathcal{K} = [0, L]$. We then have, $\overline{N_\beta(Q,q)} = k\Big(2\frac{\beta}{L} - \frac{\beta^2}{L^2}\Big)$, where $\overline{N_\beta(Q,q)} = \mathbb{E}_{Q,q}\Big\{N_\beta(Q,q)\Big\}$ and $\mathbb{E}_{Q,q}\{.\}$ represents the expected value w.r.t. Q and q.*

Proof. Define $d_q(x) \triangleq |x-q|$, for two positions x and q. For a fixed $x \in \mathcal{K}$ and a uniform distribution of $\mathbf{q}$ over $\mathcal{K}$, $d_{\mathbf{q}}(x)$ is a random variable with the following Commutative

Density Function (CDF):

$$\text{prob}(d_{\mathbf{q}}(x) \leq \beta) = \begin{cases} 2\frac{\beta}{L} & 0 \leq \beta \leq \min(x, L-x) \\ \frac{\beta}{L} + \frac{\min(x,L-x)}{L} & \min(x, L-x) \leq \beta \leq \max(x, L-x) \\ 1 & \beta \geq \max(x, L-x) \end{cases} \tag{2.9}$$

We then have,

$$\text{prob}(d_{\mathbf{q}}(x) \leq \beta)\Big|_{\beta \leq \frac{L}{2}} = \begin{cases} \frac{\beta+x}{L} & 0 \leq x \leq \beta \\ 2\frac{\beta}{L} & \beta \leq x \leq L-\beta \\ 1 + \frac{\beta-x}{L} & L-\beta \leq x \leq L \end{cases} \tag{2.10}$$

and

$$\text{prob}(d_{\mathbf{q}}(x) \leq \beta)\Big|_{\beta > \frac{L}{2}} = \begin{cases} \frac{\beta+x}{L} & 0 \leq x \leq L-\beta \\ 1 & L-\beta \leq x \leq \beta \\ 1 + \frac{\beta-x}{L} & \beta \leq x \leq L \end{cases} \tag{2.11}$$

Let $\mathbf{X}$ and $\mathbf{Q} = \{\mathbf{q_i}\}_{i=1}^{k}$ represent i.i.d. random variables, uniformly distributed over $\mathcal{K}$, and $B_\beta(\mathbf{X})$ denote an open ball with the radius of β, centered at $\mathbf{X}$. For a fixed x, we have the following PDF:

$$\begin{aligned} &\text{prob}\big(N_\beta(\mathbf{Q}, \mathbf{X}) = i \big| \mathbf{X} = x\big) \\ &= \begin{pmatrix} k \\ i \end{pmatrix} \Big[\text{prob}\big(d_{\mathbf{q_1}}(\mathbf{X} = x) \leq \beta\big)\Big]^i \Big[1 - \text{prob}\big(d_{\mathbf{q_1}}(\mathbf{X} = x) \leq \beta\big)\Big]^{k-i}. \end{aligned} \tag{2.12}$$

This results in the following average:

$$\begin{aligned} \overline{N_\beta(\mathbf{Q}, \mathbf{X})} &= E\Big\{E\Big\{N_\beta(\mathbf{Q}, \mathbf{X}) | \mathbf{X} = x\Big\}\Big\} \\ &= E\Big\{k \times \text{prob}\big(d_{\mathbf{q}}(\mathbf{X} = x) \leq \beta\big)\Big\} \overset{\text{Eq. 2.10,2.11}}{=} k\Big(2\frac{\beta}{L} - \frac{\beta^2}{L^2}\Big), \end{aligned} \tag{2.13}$$

where $\overline{N_\beta(\mathbf{Q}, \mathbf{X})}$ indicates the average number of points in $\mathcal{Q}$, which are located in the β neighborhood of x. □

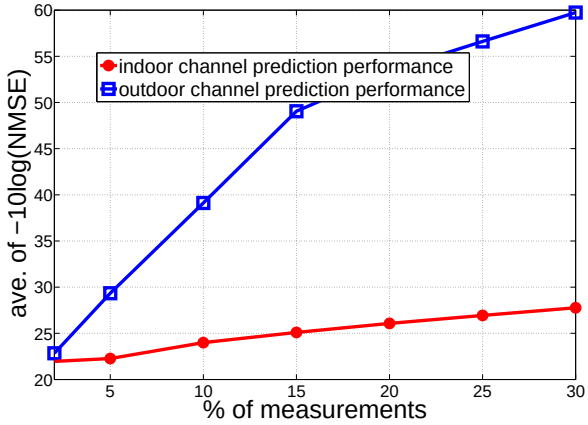

Figure 2.3: Channel prediction quality for both indoor and outdoor channel measurements, as a function of the percentage of the measurements gathered.

Special case - probabilistic path loss: If the knowledge of beta is not available or is not used in the prediction (thus beta is assumed zero), then Eq. 2.8 results in the same probability distribution for all the points that are equally-spaced from the transmitter. An example of this case can be seen in Fig. 1.1(top-middle), where we have the same predicted probability of connectivity (probability that the $\Upsilon_{\text{dB}}(q)$ is above a given threshold) for all the points at a given radius from the transmitter. Our more general case of Eq. 2.8 is then shown in Fig. 1.1(top-right), where a probability distribution (and a resulting probability of connectivity) is assigned to each point in the workspace. Both these cases result in a more comprehensive channel prediction than the commonly-used disk model of Fig. 1.1(top-left).

Next we show the reconstruction of two real channels, using our proposed method. The performance metric is the Average Normalized Mean Square Error (ANMSE) of the estimated channel, where the following Normalized Mean Square Error, $\text{NMSE} = \left[\frac{\int_{\mathcal{K}}(\Upsilon_{\text{dB}}(q)-\hat{\Upsilon}_{\text{dB},\mathcal{Q}}(q))^2 dA}{\int_{\mathcal{K}}\Upsilon^2_{\text{dB}}(q)dA}\right]$, is averaged over several different randomly-selected sampling positions, for a given percentage of collected samples. Fig. 2.3 shows the reconstruction performance for an outdoor channel across a street in downtown San Francisco [5]

as well as for an indoor channel measurement along a route in the basement of the ECE building. The indoor experiment uses an 802.11g WLAN card while the outdoor measurement is based on measuring receptions from an AT&T cell tower [5]. For both cases, all the underlying parameters are estimated using the LS approach of this section. Consider the outdoor case, for instance. We have the measurements of the received signal power, every 2mm along a street of length 16m in San Francisco, mounting to 8000 samples. Fig. 2.3 then shows the prediction performance where only a percentage of the total samples were available to a node. The available measurements are randomly chosen over the street. 5% measurements, for instance, means that a robot has collected 400 samples, randomly over that street, based on which it will predict the channel over the whole street. The prediction error variance is -29dB for the case of 5% measurements. It can be seen that both channels can be reconstructed with a good quality. The outdoor channel, however, can be reconstructed with a considerably better quality. This is expected as the indoor channel suffers from a more severe multipath fading, which makes it less spatially predictable.

2.1.1 Space-varying underlying parameters and adaptive channel prediction

So far, we considered channel prediction over a small enough space such that the underlying channel parameters can be considered constant over the workspace. However, if the available channel samples belong to a large enough space (such as the entire floor), the underlying parameters can be space-varying. In this part, we show how the previous framework can be extended to an adaptive approach, in order to address the case where the operation, and the corresponding available channel measurements, are over a large space. Basically, a robot can use its localization and mapping information (which it will have for navigation and collision avoidance) to detect when something changes in the structure of its environment. For instance, it can detect when it moves out of a room to a hallway or when it reaches an intersection. Thus, we assume that the underlying parameters can

possibly change when some environmental factors change. From analyzing several real measurements, this is a reasonable assumption. While there could possibly be cases that are not captured by this assumption, i.e. having a drastic change in an underlying parameter without any environmental change, such cases are rare and the robot can not know about it to adapt its strategy anyways.

Let a region denote a place of operation where there is no environmental changes and the underlying parameters can be considered constant (such as a room or a hallway with no intersection that leads to the transmitter). In order to allow the node to give less weight to the available measurements that are collected in different regions and/or are far from the position where the channel needs to be estimated, we introduce a forgetting factor and a distance-dependent weight. This allows the node to adapt the impact of a sample measurement on its prediction framework. The forgetting factor is used to let the node give less impact to a measurement if it belongs to a different region, as compared to the place where the robot needs to predict the channel. On the other hand, the distance-dependent weight allows the robot to give less weight to the farther measurements. Consider the case where the workspace consists of p different regions, i.e. $\mathcal{K} = \bigcup_{i=1}^{p} \mathcal{R}_i$. Let τ_i represent the region, where the ith measurement belongs to, i.e. $q_i \in \mathcal{R}_{\tau_i}$. Define the forgetting matrix F, with the following characteristics: 1) F is symmetric, 2) F is stochastic and 3) $[F]_{i,j} = f_{i,j}$ is proportional to the similarity between regions i and j. The third property implies that, $\max_j f_{j,i} = f_{i,i}$ and $f_{i,j} \geq f_{i,k}$ iff regions i and j have more environmental features in common, as compared to regions i and k. Furthermore, let $\mathcal{G}$ denote the functional space of all non-increasing functions on $\mathbf{R}^+$. For $q \in \mathcal{R}_m \subset \mathcal{K}$, we define the corresponding weight matrix as: $[\Psi_{\mathcal{Q}}(q)]_{i,i} = f_{\tau_i,m} \times g_{\tau_i,m}(\|q - q_i\|)$ and $[\Psi_{\mathcal{Q}}(q)]_{i,j} = 0$ for $i \neq j$, where $g_{\tau_i,m} \in \mathcal{G}$. One candidate for g is an exponential function: $g_{\tau_i,m}(\|q - q_i\|) = e^{-\frac{\|q-q_i\|}{b_{\tau_i,m}}}$. $f_{\tau_i,m}$ and $b_{\tau_i,m}$ are design parameters, which the robot can choose. They impact how conservative the robot will be in taking the measurements of different regions into account. Let $\hat{\theta}_{\text{WLS}}(q) = \min_\theta \left\| \Psi_{\mathcal{Q}}^{\frac{1}{2}}(q)\big(Y_{\mathcal{Q}} - H_{\mathcal{Q}}\theta\big) \right\|^2$ denote the weighted LS estimation of the path loss parameters, for prediction at position $q \in \mathcal{R}_m$. We

then have, $\hat{\theta}_{\text{WLS}}(q) = \left(H_{\mathcal{Q}}^T \Psi_{\mathcal{Q}}(q) H_{\mathcal{Q}}\right)^{-1} H_{\mathcal{Q}}^T \Psi_{\mathcal{Q}}(q) Y_{\mathcal{Q}}$. The channel and other underlying parameters can be similarly estimated.

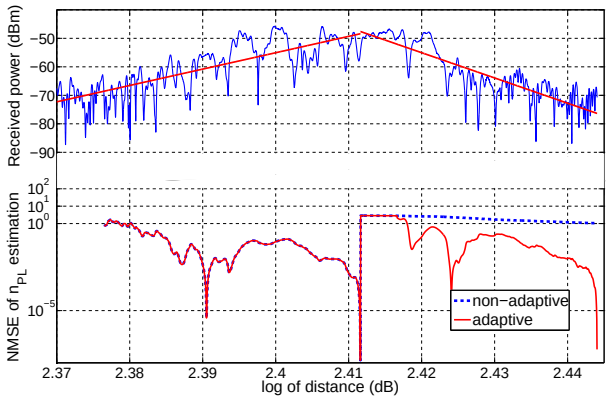

Figure 2.4: Performance of our adaptive approach, in estimating the path loss slope, when a robot moves along a street in San Francisco and samples the channel along its trajectory [5]–(top) channel received power across the street along with its best slope fit and (bottom) prediction error variance of the robot, as it moves along the street and measures the channel.

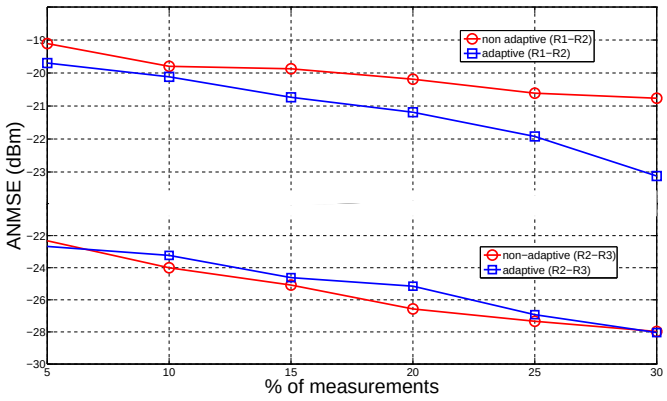

Figure 2.5: Impact of different environments on channel prediction performance, using real channel measurements. (top) indoor and outdoor, (middle) main room (R1) and hallway (R2) of Fig. 2.11 and (bottom) hallways R2 and R3 of Fig. 2.11.

Fig. 2.4 shows the performance of our adaptive approach when a robot moves along a street. The channel measurement is in reception from an AT&T cell tower, in a street

in San Francisco [5], which experiences very different path loss exponents due to the presence of an intersection that leads to the transmitter. The robot samples the channel as it moves along the street and estimates the path loss slope, without any a priori information in this environment. The figure compares the performance of the non-adaptive case with that of the adaptive one and shows that we can benefit considerably from the adaptation. Next, Fig. 2.5 (top) shows the prediction quality when a number of robots operate in our basement, over a large area and cooperate for channel prediction. The regions of operation are R1 and R2, as indicated in Fig. 2.11. Note that the performance is simulated, in this case, using real channel measurements in this environment. It can be seen that the adaptive approach can improve the performance as compared to the non-adaptive case. In the non-adaptive case, all the gathered and communicated measurements are utilized by each robot for channel prediction, without taking into account that these measurements may belong to different regions. It can be seen that we can benefit a couple of dBs, by using the adaptive approach. In other tests in different environments, we also observed that the adaptation may make a negligible difference if different regions are not that much different, in terms of their underlying parameters, as expected. Fig. 2.5 (bottom) shows an example of such a case for operation over a different area in our basement. It can be seen that the performance curves are very close.

Here, it is our goal to fundamentally understand the impact of different environments (in terms of their underlying parameters) on the proposed channel prediction framework. Consider the four marked regions of Fig. 2.11 for instance. We want to understand how the channel prediction quality changes (and justify the observed behaviors) when we move from one region to another. Therefore, in the rest of this chapter, we consider the non-adaptive channel prediction framework, to predict the channel over a region where the underlying parameters can be considered constant. In the next section, we characterize the impact of different environments (in terms of the underlying channel parameters) on the prediction framework.

2.2 Impact of channel parameters on the prediction error variance

In this section, we characterize the impact of the underlying channel parameters on the spatial predictability of a wireless channel. We assume that the underlying parameters are estimated perfectly in this section to avoid error propagation from parameter estimation to channel prediction. In the subsequent sections, we then extend our analysis to take the impact of the estimation error of key underlying parameters into account.

Let $\Upsilon_{\text{dB}}(q) = 10\log_{10}\big(\Upsilon(q)\big)$ represent the received signal strength at position $q \in \mathcal{K}$ in dB. Based on the gathered measurements at $\mathcal{Q} \subset \mathcal{K}$, the goal is to estimate the channel at $q \in \mathcal{K}\backslash\mathcal{Q}$, using the channel predictor, $\tilde{\Upsilon}_{\text{dB},\mathcal{Q}}(q)$ of Eq. 2.8, with the corresponding error covariance of $\sigma^2_{\text{dB},\mathcal{Q}}(q)$. We next characterize the impact of different channel parameters on this prediction. We first introduce the following lemmas.

Lemma 2. *Let $\Psi(t)$ be an invertible matrix for $t \in \mathbb{R}$. We have $\frac{d\Psi^{-n}}{dt} = -\Psi^{-n}\frac{d\Psi^{n}}{dt}\Psi^{-n}$, where n is a positive integer.*

Proof. Taking the derivative from both sides of equation $\Psi^n(t)\Psi^{-n}(t) = I_{k\times k}$, with respect to t, proves the lemma. □

Lemma 3. *Let J be an n-by-m matrix with the rank of m and Ψ be an n-by-n full rank matrix. If matrix Ψ is positive definite ($\Psi \succ 0$), then $J^T\Psi J$ is positive definite.*

Proof. See [99] for a proof. □

Theorem 1. *The estimation error variance, $\sigma^2_{dB,\mathcal{Q}}$, is an increasing function of α and σ^2 for $\alpha, \sigma^2 \in [0,\infty)$ and an invertible $R_{norm,\mathcal{Q}}$.*

Proof. We first show that the estimation error variance is an increasing function of σ^2. Let $\phi_{\text{norm},\mathcal{Q}}(q) = \frac{1}{\alpha}\phi_{\mathcal{Q}}(q)$ denote the normalized cross covariance between $\mathcal{Q}$ and q. We

have $\sigma^2_{\text{dB},\mathcal{Q}}(q) = \alpha + \sigma^2 - \alpha\phi^T_{\text{norm},\mathcal{Q}}(q)\big(R_{\text{norm},\mathcal{Q}} + \frac{\sigma^2}{\alpha}I_{k\times k}\big)^{-1}\phi_{\text{norm},\mathcal{Q}}(q)$. For $\alpha = 0$, we have $\frac{d}{d\sigma^2}\sigma^2_{\text{dB},\mathcal{Q}}(q)\Big|_{\alpha=0} = 1 > 0,\ \forall\sigma^2 \in [0,\infty)$. For $\alpha \neq 0$, taking the derivative with respect to σ^2 (using Lemma 2) and then applying Lemma 3 result in: $\frac{d}{d\sigma^2}\sigma^2_{\text{dB},\mathcal{Q}}(q) = 1 + \phi^T_{\text{norm},\mathcal{Q}}(q)\big(R_{\text{norm},\mathcal{Q}}+\frac{\sigma^2}{\alpha}I_{k\times k}\big)^{-2}\phi_{\text{norm},\mathcal{Q}}(q) > 0,\ \forall\sigma^2 \in [0,\infty)$ and for an invertible $R_{\text{norm},\mathcal{Q}}$, which completes the proof. We next prove that $\sigma^2_{\text{dB},\mathcal{Q}}(q)$ is an increasing function of α. First assume that $\sigma^2 \neq 0$. Taking the derivative with respect to α results in: $\frac{d}{d\alpha}\sigma^2_{\text{dB},\mathcal{Q}}(q) = 1-\phi^T_{\text{norm},\mathcal{Q}}(q)\Big[\big(R_{\text{norm},\mathcal{Q}}+\frac{\sigma^2}{\alpha}I_{k\times k}\big)^{-1}+\frac{\sigma^2}{\alpha}\big(R_{\text{norm},\mathcal{Q}}+\frac{\sigma^2}{\alpha}I_{k\times k}\big)^{-2}\Big]\phi_{\text{norm},\mathcal{Q}}(q)$. Define $f(\alpha) \triangleq \frac{d}{d\alpha}\sigma^2_{\text{dB},\mathcal{Q}}(q)$. f is of class C^∞ on $\mathbb{R}^+$ with the following properties: 1) $f(0) = 1$, 2) $f(\infty) = 1 - \phi^T_{\text{norm},\mathcal{Q}}(q)R^{-1}_{\text{norm},\mathcal{Q}}\phi_{\text{norm},\mathcal{Q}}(q) > 0$ and 3) $\frac{d}{d\alpha}f(\alpha) < 0$.

First property can be easily confirmed. We next prove the second property. Let $R_{\text{norm},\mathcal{Q}\bigcup\{q\}}$ represent the correlation matrix corresponding to $\mathcal{Q}\bigcup\{q\}$. We have $R_{\text{norm},\mathcal{Q}\bigcup\{q\}} = \begin{bmatrix} R_{\text{norm},\mathcal{Q}} & \phi_{\text{norm},\mathcal{Q}}(q) \\ \phi^T_{\text{norm},\mathcal{Q}}(q) & 1 \end{bmatrix}$, where $R_{\text{norm},\mathcal{Q}\bigcup\{q\}}$ is assumed invertible. Thus, under the assumption that $R_{\text{norm},\mathcal{Q}}$ is invertible, the second property can be easily confirmed, using the Schur complement of $R_{\text{norm},\mathcal{Q}}$ block [100]. Next we prove the third property. We have

$$\begin{aligned}\frac{d}{d\alpha}f(\alpha) &= -\phi^T_{\text{norm},\mathcal{Q}}(q)\frac{d}{d\alpha}\Big[\big(R_{\text{norm},\mathcal{Q}} + \frac{\sigma^2}{\alpha}I_{k\times k}\big)^{-1} + \frac{\sigma^2}{\alpha}\big(R_{\text{norm},\mathcal{Q}} + \frac{\sigma^2}{\alpha}I_{k\times k}\big)^{-2}\Big]\phi_{\text{norm},\mathcal{Q}}(q) \\ &= -\phi^T_{\text{norm},\mathcal{Q}}(q)\frac{\sigma^2}{\alpha}\frac{d}{d\alpha}\Big[\big(R_{\text{norm},\mathcal{Q}} + \frac{\sigma^2}{\alpha}I_{k\times k}\big)^{-2}\Big]\phi_{\text{norm},\mathcal{Q}}(q) \\ &= -2\sigma^4\phi^T_{\text{norm},\mathcal{Q}}(q)\big(\alpha R_{\text{norm},\mathcal{Q}} + \sigma^2 I_{k\times k}\big)^{-3}\phi_{\text{norm},\mathcal{Q}}(q). \qquad (2.14)\end{aligned}$$

Since $\big(\alpha R_{\text{norm},\mathcal{Q}} + \sigma^2 I_{k\times k}\big) \succ 0$, we can then easily show that $\frac{d}{d\alpha}f(\alpha) < 0$ using Lemma 3. By using these three properties, we have $f(\alpha) > 0$, which means that the estimation error variance is an increasing function of $\alpha \in [0,\infty)$. Furthermore, if $\sigma^2 = 0$ and $R_{\text{norm},\mathcal{Q}}$ is invertible, then $\frac{d}{d\alpha}\sigma^2_{\text{dB},\mathcal{Q}}(q)\big|_{\sigma^2=0} = f(\infty)$, which is positive as shown for property 2. Therefore, estimation variance is an increasing function of α in this case too. □

We next characterize the impact of β on the prediction quality, using properties of the Euclidean Distance Matrix (EDM) [101]. Given the position set $\mathcal{Q} = \{q_1, q_2, \cdots, q_k\} \subset \mathcal{K}$, the EDM $\Pi = \big[\pi_{i,j}\big] \in \mathbb{R}^{k\times k}$ is defined entry-wise as $\big[\Pi\big]_{i,j} = \pi_{i,j} = \|q_i - q_j\|^2$ for $i, j =$

$1, 2, \cdots, k$. We have the following properties for the EDM:
1) $\sqrt{\pi_{i,j}} \geq 0$ for $i \neq j$ and $\sqrt{\pi_{i,j}} = 0$ for $i = j$,
2) $\sqrt{\pi_{i,j}} = \sqrt{\pi_{j,i}}$,
3) $\sqrt{\pi_{i,l}} + \sqrt{\pi_{l,j}} \geq \sqrt{\pi_{i,j}}$ for $i \neq j \neq l$.

Theorem 2. *Matrix* $\Pi = \left[\pi_{i,j}\right] \in \mathbb{R}^{k\times k}$ *is EDM if and only if* $-V_k^T \Pi V_k \succeq 0, \Pi^T = \Pi$ *and* $\pi_{i,i} = 0$ *for* $1 \leq i \leq k$, *where* V_k *is the full-rank skinny Schoenberg auxiliary matrix:* $V_k \triangleq \frac{1}{\sqrt{2}} \begin{bmatrix} -1_{k-1}^T \\ I_{(k-1)\times(k-1)} \end{bmatrix} \in \mathbb{R}^{k\times k-1}$.

Proof. Readers are referred to [101] for the details of the proof. □

Theorem 3. *Let* $T = \left[t_{i,j}\right] \in \mathbb{R}^{k\times k}$ *represent the entry-wise square root of* $\Pi = \left[\pi_{i,j}\right] \in \mathbb{R}^{k\times k}$ *where* $t_{i,j} = \pi_{i,j}^{\frac{1}{2}}$. *If* Π *is EDM, then* T *is EDM. This case is of interest because it corresponds to the absolute distance matrix.*

Proof. Readers are referred to [101–103] for the details of the proof. □

Lemma 4. *The Hadamard product (Schur product) of two positive-definite matrices is positive-definite and the Hadamard product of two positive-semidefinite matrices is positive-semidefinite.*

Proof. Readers are referred to Theorem 7.5.3 of [99] for more details. □

Theorem 4. *The estimation error variance is a decreasing function of* $\beta \in (0, \infty)$ *for* $\sigma^2 \neq 0$ *and a non-increasing function of* $\beta \in (0, \infty)$ *for* $\sigma^2 = 0$ *and an invertible* $R_{norm,\mathcal{Q}}$.

Proof. Case of $\alpha = 0$ is not of interest in this theorem since we are interested in the impact of shadowing. Therefore, in this proof we assume that $\alpha \neq 0$. Let $\delta_{\mathcal{Q}}(q) = \left[\|q_1 - q\|, \|q_2 - q\|, \cdots, \|q_k - q\|\right]^T$ represent the distance vector between the set $\mathcal{Q}$ and position $q \notin \mathcal{Q}$ and $\Delta_{\mathcal{Q}}(q) \triangleq \text{diag}\left[\delta_{\mathcal{Q}}(q)\right]$. Let $\left[T_{\mathcal{Q}}\right]_{i,j} = \| q_i - q_j \|, \forall q_i, q_j \in \mathcal{Q}$, denote

the absolute distance matrix corresponding to the set $\mathcal{Q}$. First assume that $\sigma^2 \neq 0$. We have,

$$\frac{d}{d\beta}\sigma^2_{\text{dB},\mathcal{Q}}(q) = -\frac{1}{\beta^2}\phi^T_{\mathcal{Q}}(q)\Big[\Delta_{\mathcal{Q}}(q)R^{-1}_{\text{tot},\mathcal{Q}} - \alpha R^{-1}_{\text{tot},\mathcal{Q}}\big(T_{\mathcal{Q}} \bullet R_{\text{norm},\mathcal{Q}}\big)R^{-1}_{\text{tot},\mathcal{Q}} + R^{-1}_{\text{tot},\mathcal{Q}}\Delta_{\mathcal{Q}}(q)\Big]\phi_{\mathcal{Q}}(q)$$
$$= -\frac{1}{\beta^2}\phi^T_{\mathcal{Q}}(q)R^{-1}_{\text{tot},\mathcal{Q}}\Big[R_{\text{tot},\mathcal{Q}}\Delta_{\mathcal{Q}}(q) - \alpha T_{\mathcal{Q}} \bullet R_{\text{norm},\mathcal{Q}} + \Delta_{\mathcal{Q}}(q)R_{\text{tot},\mathcal{Q}}\Big]R^{-1}_{\text{tot},\mathcal{Q}}\phi_{\mathcal{Q}}(q)$$
$$= -\frac{1}{\beta^2}\phi^T_{\mathcal{Q}}(q)R^{-1}_{\text{tot},\mathcal{Q}}\Big[\alpha\Big(R_{\text{norm},\mathcal{Q}}\Delta_{\mathcal{Q}}(q) + \Delta_{\mathcal{Q}}(q)R_{\text{norm},\mathcal{Q}} - T_{\mathcal{Q}} \bullet R_{\text{norm},\mathcal{Q}}\Big) + 2\sigma^2\Delta_{\mathcal{Q}}(q)\Big]R^{-1}_{\text{tot},\mathcal{Q}}\phi_{\mathcal{Q}}(q),$$

where $(\bullet)$ denotes the Hadamard product. Moreover, it can be confirmed that $R_{\text{norm},\mathcal{Q}}\Delta_{\mathcal{Q}}(q) = R_{\text{norm},\mathcal{Q}} \bullet \big(1_k\delta^T_{\mathcal{Q}}(q)\big)$. Therefore, we have: $\frac{d}{d\beta}\sigma^2_{\text{dB},\mathcal{Q}}(q) = -\frac{1}{\beta^2}\phi^T_{\mathcal{Q}}(q)R^{-1}_{\text{tot},\mathcal{Q}}\Big[\alpha\Big(1_k\delta^T_{\mathcal{Q}}(q) + \delta_{\mathcal{Q}}(q)1^T_k - T_{\mathcal{Q}}\Big) \bullet R_{\text{norm},\mathcal{Q}} + 2\sigma^2\Delta_{\mathcal{Q}}(q)\Big]R^{-1}_{\text{tot},\mathcal{Q}}\phi_{\mathcal{Q}}(q)$. From Lemma 4, we know that the Hadamard product of two positive-semidefinite matrices is positive-semidefinite. Therefore, to prove that $\frac{d}{d\beta}\sigma^2_{\text{dB},\mathcal{Q}}(q)\big|_{\sigma^2\neq 0} < 0$, it suffices to show that $1_k\delta^T_{\mathcal{Q}}(q) + \delta_{\mathcal{Q}}(q)1^T_k - T_{\mathcal{Q}}$ is positive-semidefinite $\big($we know that $\Delta_{\mathcal{Q}}(q) \succ 0\big)$. Let $T_{\{q\}\bigcup\mathcal{Q}} = \begin{bmatrix} 0 & \delta^T_{\mathcal{Q}}(q) \\ \delta_{\mathcal{Q}}(q) & T_{\mathcal{Q}} \end{bmatrix} \in \mathbb{R}^{(k+1)\times(k+1)}$ represent the distance matrix corresponding to $\{q\}\bigcup\mathcal{Q}$. Let e_i denote a unit vector in $\mathbb{R}^{k+1}$, where all the entries are zero except for the ith one. Therefore, the Schoenberg auxiliary matrix can be represented as $V_{k+1} = \frac{1}{\sqrt{2}}\big[e_2 - e_1, \cdots, e_{k+1} - e_1\big]$. We have:

$$\begin{aligned}
&-\Big[V^T_{k+1}T_{\{q\}\bigcup\mathcal{Q}}V_{k+1}\Big]_{i,j} \\
&= -\frac{1}{2}\big(e_{i+1} - e_1\big)^T T_{\{q\}\bigcup\mathcal{Q}}\big(e_{j+1} - e_1\big) \\
&= -\frac{1}{2}\Big(e_{i+1}T_{\{q\}\bigcup\mathcal{Q}}e_{j+1} - e_1T_{\{q\}\bigcup\mathcal{Q}}e_{j+1} - e_{i+1}T_{\{q\}\bigcup\mathcal{Q}}e_1\Big) \\
&= \frac{1}{2}\Big(\|q_j - q\| + \|q_i - q\| - \|q_i - q_j\|\Big) \\
&= \frac{1}{2}\Big[1_k\delta^T_{\mathcal{Q}}(q) + \delta_{\mathcal{Q}}(q)1^T_k - T_{\mathcal{Q}}\Big]_{i,j}.
\end{aligned}$$

Then, matrix $T_{\{q\}\bigcup\mathcal{Q}}$ is EDM using Theorem 3. Therefore, applying Theorem 2 for EDM $T_{\{q\}\bigcup\mathcal{Q}}$ results in: $1_k\delta^T_{\mathcal{Q}}(q) + \delta_{\mathcal{Q}}(q)1^T_k - T_{\mathcal{Q}} = -2V^T_{k+1}T_{\{q\}\bigcup\mathcal{Q}}V_{k+1} \succeq 0$, which completes the proof. Next consider the case where $\sigma^2 = 0$. A similar derivation will result in $\frac{d}{d\beta}\sigma^2_{\text{dB},\mathcal{Q}}(q)\Big|_{\sigma^2=0} \leq 0$, under the assumption that $R_{\text{norm},\mathcal{Q}}$ is invertible. Therefore, the estimation error variance is a non-increasing function of β in this case. □

Note that path loss parameters, K_{dB} and n_{PL}, do not affect the estimation error variance in this case. In Section 2.6, we show the impact of different environments (with different underlying parameters) on channel predictability, using several measurements in our building. In the next section, we characterize the impact of the underlying parameters on the estimation of path loss parameters.

2.3 Impact of channel parameters on path loss estimation

In this section, we explore the effect of the underlying channel parameters on the estimation of path loss parameters. To provide a benchmark, we first consider the ML estimator of Eq. 2.4, where we assume that α, β and σ^2 are perfectly known. We then consider the Least Square estimator of Eq. 2.6 for a more realistic case, where α, β and σ^2 are not known at the time of estimating path loss parameters. Let $\hat{\theta}_{\text{ML}} = [\hat{K}_{\text{dB,ML}} \;\; \hat{n}_{\text{PL,ML}}]^T$ denote the ML estimation of path loss parameters as denoted by Eq. 2.4. We have the following error covariance matrix: $C_{\theta,\text{ML}} = \mathbb{E}\left\{\left(\theta - \hat{\theta}_{\text{ML}}\right)\left(\theta - \hat{\theta}_{\text{ML}}\right)^T\right\} = \left(H_{\mathcal{Q}}^T R_{\text{tot},\mathcal{Q}}^{-1} H_{\mathcal{Q}}\right)^{-1}$, where $\sigma^2_{\hat{K}_{\text{dB,ML}}} = [C_{\theta,\text{ML}}]_{1,1}$ and $\sigma^2_{\hat{n}_{\text{PL,ML}}} = [C_{\theta,\text{ML}}]_{2,2}$ denote the ML estimation error variance of $\hat{K}_{\text{dB,ML}}$ and $\hat{n}_{\text{PL,ML}}$ respectively. We have the following Theorem.

Theorem 5. *Both $\sigma^2_{\hat{K}_{dB,ML}}$ and $\sigma^2_{\hat{n}_{PL,ML}}$ are increasing functions of α and σ^2 for $\alpha, \sigma^2 \in [0, \infty)$ and an invertible $R_{norm,\mathcal{Q}}$.*

Proof. We have $C_{\theta,\text{ML}} = \left(H_{\mathcal{Q}}^T\left(\alpha R_{\text{norm},\mathcal{Q}} + \sigma^2 I_{k\times k}\right)^{-1} H_{\mathcal{Q}}\right)^{-1}$, where $R_{\text{norm},\mathcal{Q}} = \frac{1}{\alpha} R_{\mathcal{Q}}$. Taking the derivative with respect to α results in:

$$\begin{aligned}\frac{dC_{\theta,\text{ML}}}{d\alpha} &= -C_{\theta,\text{ML}} \frac{d}{d\alpha}\Big(H_{\mathcal{Q}}^T\left(\alpha R_{\text{norm},\mathcal{Q}} + \sigma^2 I_{k\times k}\right)^{-1} H_{\mathcal{Q}}\Big) C_{\theta,\text{ML}} \\ &= C_{\theta,\text{ML}} H_{\mathcal{Q}}^T\left(\alpha R_{\text{norm},\mathcal{Q}} + \sigma^2 I_{k\times k}\right)^{-1} R_{\text{norm},\mathcal{Q}}\left(\alpha R_{\text{norm},\mathcal{Q}} + \sigma^2 I_{k\times k}\right)^{-1} H_{\mathcal{Q}} C_{\theta,\text{ML}}.\end{aligned} \tag{2.15}$$

By using Lemma 3 and the assumption that $R_{\text{norm},\mathcal{Q}} \succ 0$, we can easily see that $\frac{dC_{\theta,\text{ML}}}{d\alpha} \succ 0$. Let $e_1 = [1 \quad 0]^T$ and $e_2 = [0 \quad 1]^T$ denote unit vectors in $\mathbb{R}^2$. We have: $\frac{d}{d\alpha}\sigma^2_{\hat{K}_{\text{dB,ML}}} =$

$\frac{d}{d\alpha}\big(e_1^T C_{\theta,\text{ML}} e_1\big) = e_1^T \frac{dC_{\theta,\text{ML}}}{d\alpha} e_1 > 0$ and $\frac{d}{d\alpha}\sigma^2_{\hat{n}_{\text{PL,ML}}} = \frac{d}{d\alpha}\big(e_2^T C_{\theta,\text{ML}} e_2\big) = e_2^T \frac{dC_{\theta,\text{ML}}}{d\alpha} e_2 > 0$. To show that the estimation error of path loss parameters is an increasing function of σ^2, it suffices to show that $\frac{dC_{\theta,\text{ML}}}{d\sigma^2} \succ 0$. We have, $\frac{dC_{\theta,\text{ML}}}{d\sigma^2} = -C_{\theta,\text{ML}} H_{\mathcal{Q}}^T \frac{d}{d\sigma^2}\big(\alpha R_{\text{norm},\mathcal{Q}} + \sigma^2 I_{k\times k}\big)^{-1} H_{\mathcal{Q}} C_{\theta,\text{ML}} = C_{\theta,\text{ML}} H_{\mathcal{Q}}^T \big(\alpha R_{\text{norm},\mathcal{Q}} + \sigma^2 I_{k\times k}\big)^{-2} H_{\mathcal{Q}} C_{\theta,\text{ML}} \succ 0$, for $\alpha, \sigma^2 \in [0,\infty)$ and an invertible $R_{\text{norm},\mathcal{Q}}$. □

In general, the estimation error variance of path loss parameters does not have monotonic behavior as a function of β. To get a better understanding of the impact of correlation distance on the estimation of path loss parameters, we consider two extreme cases of $\beta = 0$ and $\beta = \infty$. More specifically, we characterize the optimum positions of the measurement points at both extremes and find the minimum achievable estimation error variance.

A. Case of $\beta = 0$: In this case, $R_{\text{tot},\mathcal{Q}}\big(\beta = 0\big) = (\alpha + \sigma^2) I_{k\times k}$ and the error covariance matrix of path loss parameters can be characterized as:

$$\lim_{\beta\to 0} C_{\theta,\text{ML}} = (\alpha+\sigma^2)(H_{\mathcal{Q}}^T H_{\mathcal{Q}})^{-1} = (\alpha+\sigma^2)\begin{bmatrix} k & -1_k^T D_{\mathcal{Q}} \\ -1_k^T D_{\mathcal{Q}} & D_{\mathcal{Q}}^T D_{\mathcal{Q}} \end{bmatrix}^{-1}$$
$$= \frac{\alpha+\sigma^2}{D_{\mathcal{Q}}^T A_k D_{\mathcal{Q}}} \begin{bmatrix} D_{\mathcal{Q}}^T D_{\mathcal{Q}} & 1_k^T D_{\mathcal{Q}} \\ 1_k^T D_{\mathcal{Q}} & k \end{bmatrix}, \tag{2.16}$$

where $A_k = kI_{k\times k} - 1_k 1_k^T$. As can be seen, the estimation error variances of both K_{dB} and n_{PL} are functions of sampling positions ($\mathcal{Q}$).

Lemma 5. *Matrix $A_k = kI_{k\times k} - 1_k 1_k^T$ has 0 and k as eigenvalues with the multiplicity of 1 and $k-1$ respectively. Let $v_1 \in$ span$\{1_k\}$ and $v_2 \in 1_k^{\perp}$, where $1_k^{\perp} = \Big\{v \Big| v^T 1_k = 0\Big\}$. We have $A_k v_1 = 0$ and $A_k v_2 = k v_2$.*

Proof. The proof is straightforward and is omitted. □

Theorem 6. *Let $D_{\mathcal{Q}}^{1_k}$ and $D_{\mathcal{Q}}^{1_k^{\perp}}$ denote the projection of $D_{\mathcal{Q}}$ to span$\{1_k\}$ and $1_k^{\perp}$ subspaces respectively. The optimum positioning, which minimizes both $\sigma^2_{\hat{K}_{dB,ML}}$ and $\sigma^2_{\hat{n}_{PL,ML}}$ for the*

case of $\beta = 0$*, is*

$$\mathcal{Q}^{opt}_{PL,\beta=0} = \arg\max_{\mathcal{Q}} ||D_{\mathcal{Q}}||_2^2, \text{ s.t. } \mathcal{Q} \subset \mathcal{K} \text{ and } D_{\mathcal{Q}}^{1_k} = 0. \tag{2.17}$$

Proof. We have the following optimum positioning in order to minimize the estimation error variance of K_{dB}, using Rayleigh-Ritz theorem [99]:

$$\begin{aligned}\mathcal{Q}^{\text{opt}}_{\sigma^2_{\hat{K}_{\text{dB,ML}},\beta=0}} &= \arg \min_{\text{s.t. } \mathcal{Q}\subset\mathcal{K}} \sigma^2_{\hat{K}_{\text{dB,ML}},\beta=0} = \arg \max_{\text{s.t. } \mathcal{Q}\subset\mathcal{K}} \frac{D_{\mathcal{Q}}^T A_k D_{\mathcal{Q}}}{D_{\mathcal{Q}}^T D_{\mathcal{Q}}} \\ &= \Big\{Q | Q \subset \mathcal{K} \text{ and } D_{\mathcal{Q}}^{1_k} = 0\Big\}.\end{aligned}$$

This optimization problem can have multiple solutions, depending on the structure of the space, all of which achieve the minimum error variance of $\frac{\alpha+\sigma^2}{k}$. Similarly, we have the following to minimize the estimation error variance of n_{PL}:

$$\begin{aligned}\mathcal{Q}^{\text{opt}}_{\sigma^2_{\hat{n}_{\text{PL,ML}},\beta=0}} &= \arg \min_{\text{s.t. } \mathcal{Q}\subset\mathcal{K}} \sigma^2_{\hat{n}_{\text{PL,ML}},\beta=0} = \arg \max_{\text{s.t. } \mathcal{Q}\subset\mathcal{K}} D_{\mathcal{Q}}^T A_k D_{\mathcal{Q}} \\ &= \arg \max_{\text{s.t. } \mathcal{Q}\subset\mathcal{K} \text{ and } D_{\mathcal{Q}}^{1_k}=0} ||D_{\mathcal{Q}}||_2^2. \end{aligned} \tag{2.18}$$

Therefore, Eq. 2.17 represents the optimum positioning which satisfies both objectives. □

Next, we provide an intuitive interpretation. Similar to Eq. 2.2, the measurement vector can be represented by $Y_{\mathcal{Q}} = (K_{\text{dB}} \times sqrtk)u_1 + (-n_{\text{PL}}\|D_{\mathcal{Q}}\|_2)u_2 + \Xi_{\mathcal{Q}}$, where $u_1 = \frac{1_k}{\sqrt{k}}$ and $u_2 = \frac{D_{\mathcal{Q}}}{\|D_{\mathcal{Q}}\|_2}$ are normalized vectors. Then, the problem becomes similar to the decoding problem in CDMA (Code Division Multiple Access) systems. Thus, we have $D_{\mathcal{Q}} \in 1_k^{\perp}$. Moreover, maximizing k and $\|D_{\mathcal{Q}}\|_2$, which can be interpreted as maximizing the SNR of each term, results in a better estimation of K_{dB} and n_{PL} respectively.

B. Case of $\beta = \infty$*:* Next we characterize the impact of correlation on the estimation quality of path loss parameters, when β goes to ∞. To simplify the derivations, we define two variables: $\rho = \frac{\alpha}{\sigma^2}$ for $\sigma^2 \neq 0$, which denotes the ratio of the power of shadowing to multipath power (in dB) and $\chi = \alpha + \sigma^2$, which represents the sum of the two powers.

The following can be easily confirmed for $\sigma^2 \neq 0$:

1) $\lim_{\beta\to\infty} R_{\text{tot},\mathcal{Q}} = \big(\rho 1_k 1_k^T + I_{k\times k}\big)\frac{\chi}{1+\rho}$,

2) $\lim_{\beta\to\infty} R_{\text{tot},\mathcal{Q}}^{-1} = \big(I_{k\times k} - \frac{\rho}{1+\rho k} 1_k 1_k^T\big)\frac{1+\rho}{\chi}$ (using Matrix Inversion Lemma),

3) $\lim_{\beta\to\infty} 1_k^T R_{\text{tot},\mathcal{Q}}^{-1} 1_k = k\frac{1+\rho}{1+\rho k}\frac{1}{\chi}$,

4) $\lim_{\beta\to\infty} 1_k^T R_{\text{tot},\mathcal{Q}}^{-1} D_{\mathcal{Q}} = \big(1_k^T D_{\mathcal{Q}}\big)\frac{1+\rho}{1+\rho k}\frac{1}{\chi}$,

5) $\lim_{\beta\to\infty} D_{\mathcal{Q}}^T R_{\text{tot},\mathcal{Q}}^{-1} D_{\mathcal{Q}} = \Big(D_{\mathcal{Q}}^T D_{\mathcal{Q}} - \frac{\rho}{1+\rho k}\big(1_k^T D_{\mathcal{Q}}\big)^2\Big)\frac{1+\rho}{\chi} = D_{\mathcal{Q}}^T\Big(\rho A_k + I_{k\times k}\Big) D_{\mathcal{Q}}\frac{1+\rho}{1+\rho k}\frac{1}{\chi}$.

Using the above equations, we have:

$$\begin{aligned}
\lim_{\beta\to\infty} C_{\theta,\text{ML}} &= \lim_{\beta\to\infty}\Big(H_{\mathcal{Q}}^T R_{\text{tot},\mathcal{Q}}^{-1} H_{\mathcal{Q}}\Big)^{-1} = \lim_{\beta\to\infty}\begin{bmatrix} 1_k^T R_{\text{tot},\mathcal{Q}}^{-1} 1_k & -1_k^T R_{\text{tot},\mathcal{Q}}^{-1} D_{\mathcal{Q}} \\ -1_k^T R_{\text{tot},\mathcal{Q}}^{-1} D_{\mathcal{Q}} & D_{\mathcal{Q}}^T R_{\text{tot},\mathcal{Q}}^{-1} D_{\mathcal{Q}}\end{bmatrix}^{-1} \\
&= \frac{1+\rho k}{1+\rho}\chi\frac{1}{\big(1+\rho k\big) D_{\mathcal{Q}}^T A_k D_{\mathcal{Q}}}\begin{bmatrix} D_{\mathcal{Q}}^T\Big(\rho A_k + I_{k\times k}\Big) D_{\mathcal{Q}} & 1_k^T D_{\mathcal{Q}} \\ 1_k^T D_{\mathcal{Q}} & k\end{bmatrix} \\
&= \frac{\chi}{1+\rho}\begin{bmatrix} \rho + \frac{D_{\mathcal{Q}}^T D_{\mathcal{Q}}}{D_{\mathcal{Q}}^T A_k D_{\mathcal{Q}}} & \frac{1_k^T D_{\mathcal{Q}}}{D_{\mathcal{Q}}^T A_k D_{\mathcal{Q}}} \\ \frac{1_k^T D_{\mathcal{Q}}}{D_{\mathcal{Q}}^T A_k D_{\mathcal{Q}}} & \frac{k}{D_{\mathcal{Q}}^T A_k D_{\mathcal{Q}}}\end{bmatrix} = \begin{bmatrix} \alpha + \frac{D_{\mathcal{Q}}^T D_{\mathcal{Q}}}{D_{\mathcal{Q}}^T A_k D_{\mathcal{Q}}}\sigma^2 & \frac{1_k^T D_{\mathcal{Q}}}{D_{\mathcal{Q}}^T A_k D_{\mathcal{Q}}}\sigma^2 \\ \frac{1_k^T D_{\mathcal{Q}}}{D_{\mathcal{Q}}^T A_k D_{\mathcal{Q}}}\sigma^2 & \frac{k}{D_{\mathcal{Q}}^T A_k D_{\mathcal{Q}}}\sigma^2\end{bmatrix}.
\end{aligned} \tag{2.19}$$

Remark 1. *It can be seen from Eq. 2.19 that Theorem 6 also characterizes the optimum positioning for this case. Moreover, if $\mathcal{Q}_{PL}^{opt}$ denotes the solution of Eq. 2.17, then we have,*

$$\begin{aligned}
&\sigma^2_{\hat{K}_{dB,ML}}\Big|_{\beta=0} = \frac{\alpha+\sigma^2}{k},\ \sigma^2_{\hat{K}_{dB,ML}}\Big|_{\beta=\infty} = \alpha + \frac{\sigma^2}{k}, \\
&\sigma^2_{\hat{n}_{PL,ML}}\Big|_{\beta=0} = \frac{\alpha+\sigma^2}{\|D_{\mathcal{Q}_{PL}^{opt}}\|^2} \text{ and } \sigma^2_{\hat{n}_{PL,ML}}\Big|_{\beta=\infty} = \frac{\sigma^2}{\|D_{\mathcal{Q}_{PL}^{opt}}\|^2}.
\end{aligned} \tag{2.20}$$

As can be seen, the fully correlated case provides a smaller estimation error variance for n_{PL} and larger for K_{dB}. In Section 2.5, we will show that the slope of path loss, n_{PL}, has the most impact on the overall channel estimation error variance. Thus, case of $\beta = \infty$ would be more desirable than $\beta = 0$.

Remark 2. *Consider the case where multipath effect is negligible, i.e., $\sigma^2 = 0$. We have*

$$\lim_{\beta\to\infty} C_{\theta,ML,\sigma^2=0} = \begin{bmatrix}\alpha & 0 \\ 0 & 0\end{bmatrix}. \tag{2.21}$$

For this case, the measurement vector becomes $Y_{\mathcal{Q}} = H_{\mathcal{Q}}\theta + \varrho 1_k$, where $\varrho \sim \mathcal{N}(0, \alpha)$ with $\mathcal{N}$ denoting a Gaussian distribution. Thus, for $k \geq 2$, the slope of path loss, $-n_{PL}$, can be perfectly estimated. However, the uncertainty of ϱ results in a bias in the estimation of K_{dB}, as can be seen from Eq. 2.21. It can also be seen that the estimation error covariance is not a function of the sampling positions anymore.

We next characterize the LS estimation of path loss parameters. Let $\hat{\theta}_{\text{LS}}$ denote the LS estimation of path loss parameters as denoted by Eq. 2.6. We have the following error covariance matrix: $C_{\theta,\text{LS}} = \left(H_{\mathcal{Q}}^T H_{\mathcal{Q}}\right)^{-1} H_{\mathcal{Q}}^T R_{\text{tot},\mathcal{Q}} H_{\mathcal{Q}} \left(H_{\mathcal{Q}}^T H_{\mathcal{Q}}\right)^{-1}$. The following Theorem characterizes some properties of this estimator.

Theorem 7. *Let $\hat{\theta}_{LS}$ and $C_{\theta,LS}$ represent the Least Square estimator of path loss parameters and the corresponding estimation error covariance matrix respectively. Let $\sigma^2_{\hat{K}_{dB,LS}}$ and $\sigma^2_{\hat{n}_{PL,LS}}$ denote the LS error variances of $\hat{K}_{dB,LS}$ and $\hat{n}_{PL,LS}$ respectively. We have the following properties:*
1) $C_{\theta,LS} \succeq C_{\theta,ML}$,
2) $\sigma^2_{\hat{K}_{dB,LS}}$ and $\sigma^2_{\hat{n}_{PL,LS}}$ are increasing functions of σ^2 for α, $\sigma^2 \in [0, \infty)$. Moreover, $\sigma^2_{\hat{K}_{dB,LS}}$ and $\sigma^2_{\hat{n}_{PL,LS}}$ are increasing functions of α for α, $\sigma^2 \in [0, \infty)$ and an invertible $R_{norm,\mathcal{Q}}$,
3) Both ML and LS estimators provide the same estimation error covariance matrices if $\beta = 0$ or ∞.

Proof. Let $\varsigma \in \mathbb{R}^+$. Using Matrix Inversion Lemma, we have:

$$\left(R_{\text{tot},\mathcal{Q}}^{-1} + \varsigma H_{\mathcal{Q}} H_{\mathcal{Q}}^T\right)^{-1} = R_{\text{tot},\mathcal{Q}} - R_{\text{tot},\mathcal{Q}} H_{\mathcal{Q}} \Big(\varsigma^{-1} I_{2\times 2} + H_{\mathcal{Q}}^T R_{\text{tot},\mathcal{Q}} H_{\mathcal{Q}}\Big)^{-1} H_{\mathcal{Q}}^T R_{\text{tot},\mathcal{Q}}. \tag{2.22}$$

By noting that matrix $\left(R_{\text{tot},\mathcal{Q}}^{-1} + \varsigma H_{\mathcal{Q}} H_{\mathcal{Q}}^T\right)^{-1}$ is positive definite, we have,

$$\begin{aligned} & R_{\text{tot},\mathcal{Q}} - R_{\text{tot},\mathcal{Q}} H_{\mathcal{Q}} \Big(\varsigma^{-1} I_{2\times 2} + H_{\mathcal{Q}}^T R_{\text{tot},\mathcal{Q}} H_{\mathcal{Q}}\Big)^{-1} H_{\mathcal{Q}}^T R_{\text{tot},\mathcal{Q}} \succ 0 \\ \overset{\text{Lemma 3}}{\Longrightarrow} & R_{\text{tot},\mathcal{Q}}^{-1} - H_{\mathcal{Q}} \Big(\varsigma^{-1} I_{2\times 2} + H_{\mathcal{Q}}^T R_{\text{tot},\mathcal{Q}} H_{\mathcal{Q}}\Big)^{-1} H_{\mathcal{Q}}^T \succ 0 \end{aligned}$$

Next, let ς go to ∞. We have

$$R_{\text{tot},\mathcal{Q}}^{-1} - H_{\mathcal{Q}}\Big(H_{\mathcal{Q}}^T R_{\text{tot},\mathcal{Q}} H_{\mathcal{Q}}\Big)^{-1} H_{\mathcal{Q}}^T \succeq 0 \overset{\text{Lemma 3}}{\Longrightarrow} H_{\mathcal{Q}}^T R_{\text{tot},\mathcal{Q}}^{-1} H_{\mathcal{Q}} \succeq H_{\mathcal{Q}}^T H_{\mathcal{Q}}\Big(H_{\mathcal{Q}}^T R_{\text{tot},\mathcal{Q}} H_{\mathcal{Q}}\Big)^{-1} H_{\mathcal{Q}}^T H_{\mathcal{Q}}$$
$$\Longrightarrow \Big(H_{\mathcal{Q}}^T R_{\text{tot},\mathcal{Q}}^{-1} H_{\mathcal{Q}}\Big)^{-1} \preceq \Big(H_{\mathcal{Q}}^T H_{\mathcal{Q}}\Big)^{-1}\Big(H_{\mathcal{Q}}^T R_{\text{tot},\mathcal{Q}} H_{\mathcal{Q}}\Big)\Big(H_{\mathcal{Q}}^T H_{\mathcal{Q}}\Big)^{-1} \Longrightarrow C_{\theta,\text{ML}} \preceq C_{\theta,\text{LS}}, \tag{2.23}$$

where for the last line, we are using the property that if $A \succeq B \succ 0$ then $B^{-1} \succeq A^{-1} \succ 0$ and we are assuming $H_{\mathcal{Q}}$ is full-rank.

The second property can be easily confirmed by taking the derivative with respect to σ^2 and α. We next prove the third property. For $\beta = 0$, we have $R_{\text{tot},\mathcal{Q}} = (\alpha + \sigma^2) I_{k\times k}$, resulting in

$$\lim_{\beta\to 0} C_{\theta,\text{LS}} = \lim_{\beta\to 0} C_{\theta,\text{ML}} = \frac{\alpha+\sigma^2}{D_{\mathcal{Q}}^T A_k D_{\mathcal{Q}}} \begin{bmatrix} D_{\mathcal{Q}}^T D_{\mathcal{Q}} & 1_k^T D_{\mathcal{Q}} \\ 1_k^T D_{\mathcal{Q}} & k \end{bmatrix}.$$

For $\beta = \infty$, we have $R_{\text{tot},\mathcal{Q}} = \alpha 1_k 1_k^T + \sigma^2 I_{k\times k}$,

$$C_{\theta,\text{LS}} = \alpha\Big(H_{\mathcal{Q}}^T H_{\mathcal{Q}}\Big)^{-1}\Big(H_{\mathcal{Q}}^T 1_k 1_k^T H_{\mathcal{Q}}\Big)\Big(H_{\mathcal{Q}}^T H_{\mathcal{Q}}\Big)^{-1} + \sigma^2\Big(H_{\mathcal{Q}}^T H_{\mathcal{Q}}\Big)^{-1},$$

and

$$\begin{aligned}
&\Big(H_{\mathcal{Q}}^T H_{\mathcal{Q}}\Big)^{-1}\Big(H_{\mathcal{Q}}^T 1_k 1_k^T H_{\mathcal{Q}}\Big)\Big(H_{\mathcal{Q}}^T H_{\mathcal{Q}}\Big)^{-1} \\
&= \frac{1}{\Big(D_{\mathcal{Q}}^T A_k D_{\mathcal{Q}}\Big)^2} \begin{bmatrix} D_{\mathcal{Q}}^T D_{\mathcal{Q}} & 1_k^T D_{\mathcal{Q}} \\ 1_k^T D_{\mathcal{Q}} & k \end{bmatrix} \begin{bmatrix} k^2 & -k\big(1_k^T D_{\mathcal{Q}}\big) \\ -k\big(1_k^T D_{\mathcal{Q}}\big) & \big(1_k^T D_{\mathcal{Q}}\big)^2 \end{bmatrix} \begin{bmatrix} D_{\mathcal{Q}}^T D_{\mathcal{Q}} & 1_k^T D_{\mathcal{Q}} \\ 1_k^T D_{\mathcal{Q}} & k \end{bmatrix} \\
&= \frac{1}{\Big(D_{\mathcal{Q}}^T A_k D_{\mathcal{Q}}\Big)^2} \begin{bmatrix} k^2\big(D_{\mathcal{Q}}^T D_{\mathcal{Q}}\big) - k\big(1_k^T D_{\mathcal{Q}}\big)^2 & -k\big(D_{\mathcal{Q}}^T D_{\mathcal{Q}}\big)\big(1_k^T D_{\mathcal{Q}}\big) + \big(1_k^T D_{\mathcal{Q}}\big)^3 \\ 0 & 0 \end{bmatrix} \\
&\times \begin{bmatrix} D_{\mathcal{Q}}^T D_{\mathcal{Q}} & 1_k^T D_{\mathcal{Q}} \\ 1_k^T D_{\mathcal{Q}} & k \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.
\end{aligned} \tag{2.24}$$

Therefore, we have

$$\lim_{\beta\to\infty} C_{\theta,\text{LS}} = \alpha \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \frac{\sigma^2}{D_{\mathcal{Q}}^T A_k D_{\mathcal{Q}}} \begin{bmatrix} D_{\mathcal{Q}}^T D_{\mathcal{Q}} & 1_k^T D_{\mathcal{Q}} \\ 1_k^T D_{\mathcal{Q}} & k \end{bmatrix} = \begin{bmatrix} \alpha + \frac{D_{\mathcal{Q}}^T D_{\mathcal{Q}}}{D_{\mathcal{Q}}^T A_k D_{\mathcal{Q}}}\sigma^2 & \frac{1_k^T D_{\mathcal{Q}}}{D_{\mathcal{Q}}^T A_k D_{\mathcal{Q}}}\sigma^2 \\ \frac{1_k^T D_{\mathcal{Q}}}{D_{\mathcal{Q}}^T A_k D_{\mathcal{Q}}}\sigma^2 & \frac{k}{D_{\mathcal{Q}}^T A_k D_{\mathcal{Q}}}\sigma^2 \end{bmatrix}.$$

By comparing this equation to Eq. 2.19, the third property can be verified. □

Remark 3. *Theorem 7 (part 3) shows that the optimum positioning of Eq. 2.17 minimizes the estimation error variance of the LS case too.*

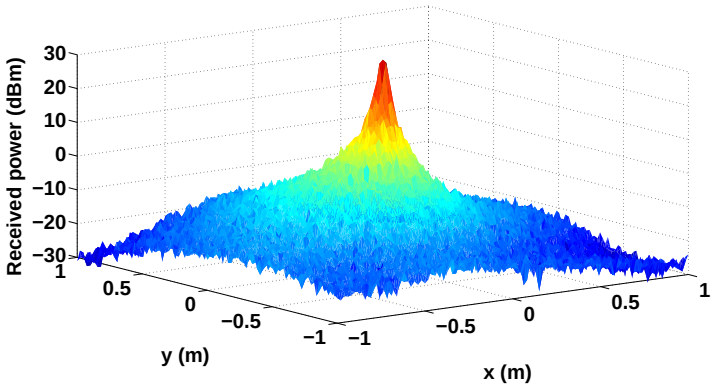

Figure 2.6: A 2D simulated channel at 1GHz frequency with the following underlying parameters: $\theta = [-22 \;\; 3.0]^T$, $\sqrt{\alpha} = \sqrt{8}$dB, $\beta = 1$m and $\sigma = \sqrt{2}$dB. The transmitter is located at $q_b = [0 \;\; 0]^T$.

We next verify the derived theorems, using a simulated channel. Fig. 2.6 shows a simulated channel, generated with our probabilistic channel simulator [104], with the following parameters: frequency of operation of 1GHz, $\theta = [-22 \quad 3.0]^T$, $\sqrt{\alpha} = \sqrt{8}$dB and $\beta = 1$m. As for multipath fading, this channel experiences a correlated Rician fading, with Jakes power spectrum [13], which results in the multipath fading getting uncorrelated after 0.12m. The pdf of a unit-average Rician distribution, with parameter K_{ric}, is given by [11]: $f_{\Upsilon_{\text{MP}}}(x) = (1 + K_{\text{ric}})e^{-K_{\text{ric}}-(1+K_{\text{ric}})x} I_0\Big(2\sqrt{xK_{\text{ric}}(K_{\text{ric}}+1)}\Big)$, where $I_0(.)$ is the modified zeroth-order Bessel function. Note that $K_{\text{ric}} = 0$ results in an exponential distribution, which experiences a considerable amount of channel variations, while $K_{\text{ric}} = \infty$ results in no fading, i.e., we will have a channel with only path loss and shadowing. Multipath power (in dB), σ^2, is related to K_{ric} as follows: $\sigma^2 = \mathbb{E}\{\omega^2(q)\} = 100\int_0^\infty \log_{10}^2(x) f_{\Upsilon_{\text{MP}}}(x)dx - 100\Big(\int_0^\infty \log_{10}(x) f_{\Upsilon_{\text{MP}}}(x)dx\Big)^2$. For the simulated channel of Fig. 2.6, $\sigma = \sqrt{2}$ dB, which corresponds to $K_{\text{ric}} = 19$.

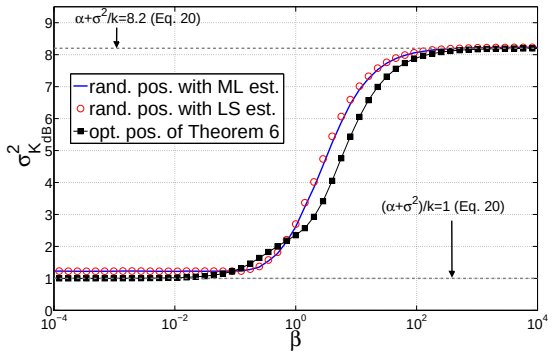

Figure 2.7: Impact of β on the estimation of K_{dB} for both optimum positioning of Theorem 6 and random sampling.

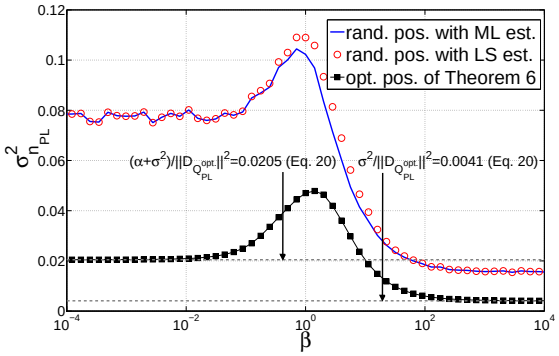

Figure 2.8: Impact of β on the estimation of n_{PL} for both optimum positioning of Theorem 6 and random sampling.

Fig. 2.7 and Fig. 2.8 show the impact of the correlation distance, β, on the estimation variance of K_{dB} and n_{PL} respectively. In this example, the workspace is a ring with an inner radius of 0.3m and an outer radius of 3.3m, superimposed on the simulated channel of Fig. 2.6, such that the centers of the rings are positioned at the transmitter. We consider the case where $k = 8$ samples are taken from the workspace. Furthermore, we compare the performance for the case of random uniformly-distributed samples with the case where samples are optimally positioned based on Theorem 6. For this workspace, enforcing $D_{\mathcal{Q}}^{1_k} = 0$ results in $\max \|D_{\mathcal{Q}}\|_2^2 = 100k \log_{10}^2(\frac{10}{3})$, which can be achieved if and only if

half of the samples are distributed on the inner circle while the other half are on the outer one. Therefore, we assume that four samples are equally-spaced on the inner circle while the other four are equally-spaced on the outer one.[3] The figures show that the optimum positioning of Theorem 6 can reduce the error, especially for the estimation of n_{PL} (which will have the most impact on the overall channel estimation error [6]). It can also be seen that as β approaches 0 or ∞, both estimators have the same quality as predicted by Theorem 7. Finally, the performances of the two estimators are not that different for other values of β in this example.

2.4 Impact of channel parameters on the estimation of the shadowing power

In this section we investigate the impact of different channel parameters on the estimation of the shadowing power (α). Similar to the previous sections, we consider both ML and LS estimations of the shadowing power. While ML is the optimum estimator, it requires the knowledge of β, which may not be available. Thus, the ML analysis provides a benchmark for the estimation performance of the LS approach. We first consider the case where path loss parameters are perfectly estimated and then extend our analysis to the case where path loss parameters are estimated using ML or LS estimator.

As mentioned earlier in Section 2.1, finding a closed-form expression for the ML estimation of the shadowing power in the presence of multipath fading is a challenging problem. To simplify the mathematical analysis, we neglect the multipath fading effects in this part and consider the case where $R_{\mathrm{tot},\mathcal{Q}} = \alpha R_{\mathrm{norm},\mathcal{Q}}$. However, the analysis for the LS estimator can be easily extended to the general case where $\sigma^2 \neq 0$. For this case, the LS approach estimates χ which is the sum of the shadowing and multipath powers in the

[3]Note that the multipath fading components of different sampling points are uncorrelated with 100% probability for the case of optimum positioning and 95% probability for the case of random.

dB domain.

Let $Y_{\mathcal{Q},\text{cent}} = Y_{\mathcal{Q}} - H_{\mathcal{Q}}\theta$ denote the centered version of vector $Y_{\mathcal{Q}}$. Let $\hat{\alpha}_{\text{ML}|\theta}$ and $\hat{\alpha}_{\text{LS}|\theta}$ represent the Maximum Likelihood and Least Square estimators of α respectively, when θ is known. As discussed in Section 2.1, we have[4]

$$\hat{\alpha}_{\text{ML}|\theta} = \frac{Y_{\mathcal{Q},\text{cent}}^T R_{\text{norm},\mathcal{Q}}^{-1} Y_{\mathcal{Q},\text{cent}}}{k} \text{ and } \hat{\alpha}_{\text{LS}|\theta} = \frac{Y_{\mathcal{Q},\text{cent}}^T Y_{\mathcal{Q},\text{cent}}}{k}, \tag{2.25}$$

For $\beta = \infty$, matrix $R_{\text{norm},\mathcal{Q}}$ is not invertible and the ML estimator can not be applied. As β goes to ∞, we have $Y_{\mathcal{Q},\text{cent}} = \varrho 1_k$ where $\varrho \sim \mathcal{N}(0,\alpha)$. Therefore, the optimum estimate of the shadowing power becomes ϱ^2 in this case, which can be attained by the LS estimator with the estimation error variance of $\mathbb{E}\{(\varrho^2 - \alpha)^2\} = 2\alpha^2$. However, in practice, when β goes to ∞, α will become zero, which results in no shadowing variations and hence nothing to estimate beyond the path loss curve.

Lemma 6. *Assume $X \in \mathbb{R}^k \sim \mathcal{N}(0,R)$ is a zero mean multivariate Gaussian vector with the covariance matrix $R = [r_{i,j}] \in \mathbb{R}^{k\times k}$. We have $\mathbb{E}\{(X^TX)^2\} = \big[\text{trace}(R)\big]^2 + 2 \times 1_k^T(R \bullet R)1_k$.*

Proof. We have

$$\begin{aligned}\mathbb{E}\Big\{(X^TX)^2\Big\} &= \sum_{i=1}^{k}\mathbb{E}\{x_i^4\} + \sum_{\substack{1\le i,j\le k\\ i\neq j}}\mathbb{E}\{x_i^2x_j^2\} = 3\sum_{i=1}^{k} r_{i,i}^2 + \sum_{\substack{1\le i,j\le k\\ i\neq j}}\big(r_{i,i}r_{j,j} + 2r_{i,j}^2\big)\\ &= \Big[\sum_{i=1}^{k} r_{i,i}\Big]^2 + 2\sum_{1\le i,j\le k} r_{i,j}^2 = \big[\text{trace}(R)\big]^2 + 2\times 1_k^T(R\bullet R)1_k. \end{aligned}\tag{2.26}$$

□

Theorem 8. *Consider the case where multipath is negligible, i.e. $\sigma^2 = 0$ and $R_{tot,\mathcal{Q}} = \alpha R_{norm,\mathcal{Q}}$. Let $\hat{\alpha}_{ML|\theta}$ and $\hat{\alpha}_{LS|\theta}$ represent the Maximum Likelihood and Least Square estimators of α respectively, when θ is known. We have the following properties:*

[4]We assume perfect knowledge of β when we use this ML estimation in the rest of this section.

1) Both estimators are unbiased,
2) $\sigma^2_{\hat{\alpha}_{ML|\theta}} = \frac{2}{k}\alpha^2$ *and* $\sigma^2_{\hat{\alpha}_{LS|\theta}} = \frac{2}{k^2}\alpha^2 \sum_{1\le i,j\le k} e^{-\frac{\|q_i - q_j\|}{\beta}}$ *where* $q_i \neq q_j$ *if* $i \neq j$*. Thus, the error variance of both estimators is an increasing function of* α*,*
3) $\sigma^2_{\hat{\alpha}_{LS|\theta}}$ *is an increasing function of* β *while* $\sigma^2_{\hat{\alpha}_{ML|\theta}}$ *is independent of* β*.*

Proof. We have: $\mathbb{E}\{\hat{\alpha}_{\text{ML}|\theta}\} = \frac{1}{k}\mathbb{E}\{Y^T_{\mathcal{Q},\text{cent}} R^{-1}_{\text{norm},\mathcal{Q}} Y_{\mathcal{Q},\text{cent}}\} = \frac{1}{k}\text{trace}\Big[R^{-1}_{\text{norm},\mathcal{Q}}\mathbb{E}\{Y_{\mathcal{Q},\text{cent}} Y^T_{\mathcal{Q},\text{cent}}\}\Big] = \alpha$. The LS case can be immediately verified to be unbiased. Next, we prove the second property. Define $U \triangleq R^{-\frac{1}{2}}_{\text{norm},\mathcal{Q}} Y_{\mathcal{Q},\text{cent}} \sim \mathcal{N}(0, \alpha I_{k\times k})$, where $\beta \neq \infty$ (note that the ML estimate is not defined for $\beta = \infty$). Using Lemma 6 we have:

$$\begin{aligned}
\sigma^2_{\hat{\alpha}_{\text{ML}|\theta}} &= \mathbb{E}\left\{\left(\alpha - \frac{Y^T_{\mathcal{Q},\text{cent}} R^{-1}_{\text{norm},\mathcal{Q}} Y_{\mathcal{Q},\text{cent}}}{k}\right)^2\right\} = \mathbb{E}\left\{\left(\frac{Y^T_{\mathcal{Q},\text{cent}} R^{-1}_{\text{norm},\mathcal{Q}} Y_{\mathcal{Q},\text{cent}}}{k}\right)^2\right\} - \alpha^2 \\
&= \frac{1}{k^2}\mathbb{E}\left\{\left(U^T U\right)^2\right\} - \alpha^2 = \frac{\alpha^2}{k^2}\left(k^2 + 2k\right) - \alpha^2 = \frac{2}{k}\alpha^2 \text{ and} \\
\sigma^2_{\hat{\alpha}_{\text{LS}|\theta}} &= \mathbb{E}\left\{\left(\alpha - \frac{Y^T_{\mathcal{Q},\text{cent}} Y_{\mathcal{Q},\text{cent}}}{k}\right)^2\right\} = \mathbb{E}\left\{\left(\frac{Y^T_{\mathcal{Q},\text{cent}} Y_{\mathcal{Q},\text{cent}}}{k}\right)^2\right\} - \alpha^2 \\
&= \frac{1}{k^2}\left(\left[\text{trace}(R_{\text{tot},\mathcal{Q}})\right]^2 + 21^T_k (R_{\text{tot},\mathcal{Q}} \bullet R_{\text{tot},\mathcal{Q}}) 1_k\right) - \alpha^2 = \frac{2}{k^2} 1^T_k (R_{\text{tot},\mathcal{Q}} \bullet R_{\text{tot},\mathcal{Q}}) 1_k \\
&= \frac{2}{k^2}\alpha^2 1^T_k (R_{\text{norm},\mathcal{Q}} \bullet R_{\text{norm},\mathcal{Q}}) 1_k = \frac{2}{k^2}\alpha^2 \sum_{1\le i,j\le k} e^{-\frac{\|q_i - q_j\|}{\beta}}.
\end{aligned}$$

Thus, they are both increasing functions of α. The third property can be easily confirmed from these equations. □

The next theorem characterizes the Cramer-Rao bound for the estimation of the shadowing power and path loss parameters under the assumption that multipath is negligible and correlation distance β is known.

Theorem 9. *Consider the case where multipath power is negligible, i.e.* $R_{tot,\mathcal{Q}} = \alpha R_{norm,\mathcal{Q}}$ *and* β *is known. Define* $\xi \triangleq [\theta^T, \alpha]^T$*. Let* $\hat{\xi}$ *and* $C(\hat{\xi})$ *represent an unbiased estimation of* ξ *and the corresponding covariance matrix respectively. We then have the following*

Cramer-Rao bound,

$$C(\hat{\xi}) \succeq \begin{bmatrix} \alpha\Big(H_{\mathcal{Q}}^T R_{norm,\mathcal{Q}}^{-1} H_{\mathcal{Q}}\Big)^{-1} & 0_{2\times 1} \\ 0_{1\times 2} & \frac{2\alpha^2}{k} \end{bmatrix}, \tag{2.27}$$

where $\succeq$ denotes matrix inequality.

Proof. From Eq. 2.3, we have,

$$\begin{aligned}&\log f_{Y_{\mathcal{Q}}}(Y_{\mathcal{Q}}|\theta,\alpha,\beta,\sigma^2=0)\\ &= -\frac{k}{2}\log 2\pi - \frac{1}{2}\log|R_{\text{norm},\mathcal{Q}}| - \frac{k}{2}\log\alpha - \frac{1}{2\alpha}\big(Y_{\mathcal{Q}} - H_{\mathcal{Q}}\theta\big)^T R_{\text{norm},\mathcal{Q}}^{-1}(\beta)\big(Y_{\mathcal{Q}} - H_{\mathcal{Q}}\theta\big),\end{aligned}$$

and

$$\begin{aligned}\nabla_\theta \log f_{Y_{\mathcal{Q}}}(Y_{\mathcal{Q}}|\theta,\alpha,\beta,\sigma^2=0) &= -\frac{1}{\alpha}\Big[H_{\mathcal{Q}}^T R_{\text{norm},\mathcal{Q}}^{-1}(\beta)H_{\mathcal{Q}}\theta - H_{\mathcal{Q}}^T R_{\text{norm},\mathcal{Q}}^{-1} Y_{\mathcal{Q}}\Big]\\ &= \frac{1}{\alpha}H_{\mathcal{Q}}^T R_{\text{norm},\mathcal{Q}}^{-1}(\beta)\Big(Y_{\mathcal{Q}} - H_{\mathcal{Q}}\theta\Big).\end{aligned}$$

Therefore,

$$\mathbb{E}\Big\{\nabla_\theta \log f_{Y_{\mathcal{Q}}}(Y_{\mathcal{Q}}|\theta,\alpha,\beta,\sigma^2=0)\nabla_\theta^T \log f_{Y_{\mathcal{Q}}}(Y_{\mathcal{Q}}|\theta,\alpha,\beta,\sigma^2=0)\Big\} = \frac{1}{\alpha}H_{\mathcal{Q}}^T R_{\text{norm},\mathcal{Q}}^{-1} H_{\mathcal{Q}}. \tag{2.28}$$

Furthermore,

$$\frac{d}{d\alpha}\log f_{Y_{\mathcal{Q}}}(Y_{\mathcal{Q}}|\theta,\alpha,\beta,\sigma^2=0) = -\frac{k}{2\alpha} + \frac{1}{2\alpha^2}\big(Y_{\mathcal{Q}} - H_{\mathcal{Q}}\theta\big)^T R_{\text{norm},\mathcal{Q}}^{-1}(\beta)\big(Y_{\mathcal{Q}} - H_{\mathcal{Q}}\theta\big).$$

Define $X \triangleq \frac{1}{\sqrt{\alpha}} R_{\text{norm},\mathcal{Q}}^{-\frac{1}{2}}(\beta)\big(Y_{\mathcal{Q}} - H_{\mathcal{Q}}\theta\big)$. We have $X \sim \mathcal{N}(0, I_{k\times k})$. Therefore,

$$\mathbb{E}\Big\{\Big(\frac{d}{d\alpha}\log f_{Y_{\mathcal{Q}}}(Y_{\mathcal{Q}}|\theta,\alpha,\beta,\sigma^2=0)\Big)^2\Big\} = \frac{1}{4\alpha^2}\mathbb{E}\Big\{\big(X^TX - k\big)^2\Big\} \overset{\text{Lemma 6}}{=} \frac{k}{2\alpha^2}. \tag{2.29}$$

For the cross terms, we have,

$$\begin{aligned}&\mathbb{E}\Big\{\frac{d}{d\alpha}\log f_{Y_{\mathcal{Q}}}(Y_{\mathcal{Q}}|\theta,\alpha,\beta,\sigma^2=0)\nabla_\theta \log f_{Y_{\mathcal{Q}}}(Y_{\mathcal{Q}}|\theta,\alpha,\beta,\sigma^2=0)\Big\}\\ &= \frac{1}{2\alpha^{\frac{3}{2}}}\mathbb{E}\Big\{\big(X^TX - k\big)H_{\mathcal{Q}}^T R_{\text{norm},\mathcal{Q}}^{-\frac{1}{2}}(\beta)X\Big\} = 0_{2\times 1}.\end{aligned} \tag{2.30}$$

From Eq. 2.28, Eq. 2.29 and Eq. 2.30, we have the following for the Fisher information:

$$I(\xi) = \begin{bmatrix} \frac{1}{\alpha} H_{\mathcal{Q}}^T R_{\text{norm},\mathcal{Q}}^{-1} H_{\mathcal{Q}} & 0_{2\times 1} \\ 0_{1\times 2} & \frac{k}{2\alpha^2} \end{bmatrix}, \tag{2.31}$$

which proves the theorem. □

In Eq. 2.4 of Section 2.1 we showed that the ML estimation of path loss parameters is an unbiased estimator, with the following error covariance matrix: $C_{\theta,\text{ML}} = \left(H_{\mathcal{Q}}^T R_{\text{tot},\mathcal{Q}}^{-1} H_{\mathcal{Q}}\right)^{-1}$. Therefore, the ML estimation achieves the Cramer-Rao bound, as can be seen from Theorem 9. Furthermore, it can be seen that the ML estimation of the shadowing power of Theorem 8 achieves the Cramer-Rao bound of Theorem 9. We next characterize the ML estimation of the shadowing power when the knowledge of the path loss parameters is not available. We first summarize some properties of projection matrices.

Lemma 7. *A square matrix $P \in \mathbb{R}^{k\times k}$ is a projection matrix iff $P^2 = P$. If P is a projection matrix, then $1_k^T \left(P \bullet P\right) 1_k = \text{trace}(P^2) = \text{trace}(P)$.*

Proof. Readers are referred to [99]. □

Theorem 10. *Consider the case where multipath is negligible, i.e. $\sigma^2 = 0$, and β is bounded. Let $\hat{\alpha}_{ML|\theta=\hat{\theta}_{ML}}$ represent the ML estimation of α when θ is estimated using an ML estimator: $\theta = \hat{\theta}_{ML}$. Then, $\mathbb{E}\left\{\hat{\alpha}_{ML|\theta=\hat{\theta}_{ML}}\right\} = \frac{k-2}{k}\alpha$, which is a biased estimator. Moreover, we have the following for the estimation error variance $\sigma^2_{\hat{\alpha}_{ML|\theta=\hat{\theta}_{ML}}} = \frac{2}{k}\alpha^2$, which is independent of θ, β and an increasing function of α.*

Proof. If $\theta = \hat{\theta}_{\text{ML}}$, we then have the following Maximum Likelihood estimator:

$$\hat{\alpha}_{\text{ML}|\theta=\hat{\theta}_{\text{ML}}} = \frac{Y_{\mathcal{Q}}^T P_{\mathcal{Q},\text{ML}}^T R_{\text{norm},\mathcal{Q}}^{-1} P_{\mathcal{Q},\text{ML}} Y_{\mathcal{Q}}}{k}, \tag{2.32}$$

where $P_{\mathcal{Q},\text{ML}} = I_{k\times k} - H_{\mathcal{Q}}\Big(H_{\mathcal{Q}}^T R_{\text{norm},\mathcal{Q}}^{-1} H_{\mathcal{Q}}\Big)^{-1} H_{\mathcal{Q}}^T R_{\text{norm},\mathcal{Q}}^{-1}$. Hence,

$$\begin{aligned}
&P_{\mathcal{Q},\text{ML}}^T R_{\text{norm},\mathcal{Q}}^{-1} P_{\mathcal{Q},\text{ML}} \\
&= \Big(R_{\text{norm},\mathcal{Q}}^{-1} - R_{\text{norm},\mathcal{Q}}^{-1} H_{\mathcal{Q}}\Big(H_{\mathcal{Q}}^T R_{\text{norm},\mathcal{Q}}^{-1} H_{\mathcal{Q}}\Big)^{-1} H_{\mathcal{Q}}^T R_{\text{norm},\mathcal{Q}}^{-1}\Big) \\
&\times \Big(I_{k\times k} - H_{\mathcal{Q}}\Big(H_{\mathcal{Q}}^T R_{\text{norm},\mathcal{Q}}^{-1} H_{\mathcal{Q}}\Big)^{-1} H_{\mathcal{Q}}^T R_{\text{norm},\mathcal{Q}}^{-1}\Big) \\
&= R_{\text{norm},\mathcal{Q}}^{-1} - R_{\text{norm},\mathcal{Q}}^{-1} H_{\mathcal{Q}}\Big(H_{\mathcal{Q}}^T R_{\text{norm},\mathcal{Q}}^{-1} H_{\mathcal{Q}}\Big)^{-1} H_{\mathcal{Q}}^T R_{\text{norm},\mathcal{Q}}^{-1} \qquad (2.33)
\end{aligned}$$

and $\mathbb{E}\Big\{Y_{\mathcal{Q}} Y_{\mathcal{Q}}^T\Big\} = H_{\mathcal{Q}}\theta\theta^T H_{\mathcal{Q}}^T + \alpha R_{\text{norm},\mathcal{Q}}$. Therefore, using the above equations, we have

$$\begin{aligned}
\mathbb{E}\Big\{\hat{\alpha}_{\text{ML}|\theta=\hat{\theta}_{\text{ML}}}\Big\} &= \frac{1}{k}\text{trace}\left[P_{\mathcal{Q},\text{ML}}^T R_{\text{norm},\mathcal{Q}}^{-1} P_{\mathcal{Q},\text{ML}} \mathbb{E}\Big\{Y_{\mathcal{Q}} Y_{\mathcal{Q}}^T\Big\}\right] \\
&= \frac{\alpha}{k}\text{trace}\left[I_{k\times k} - R_{\text{norm},\mathcal{Q}}^{-1} H_{\mathcal{Q}}\Big(H_{\mathcal{Q}}^T R_{\text{norm},\mathcal{Q}}^{-1} H_{\mathcal{Q}}\Big)^{-1} H_{\mathcal{Q}}^T\right] \\
&= \frac{k-2}{k}\alpha. \qquad (2.34)
\end{aligned}$$

As can be seen, this estimator is biased but it has an unbiased asymptotic behavior. Next we calculate the error variance of this estimator,

$$\begin{aligned}
\sigma^2_{\hat{\alpha}_{\text{ML}|\theta=\hat{\theta}_{\text{ML}}}} &= \mathbb{E}\left\{\Big(\alpha - \frac{Y_{\mathcal{Q}}^T P_{\mathcal{Q},\text{ML}}^T R_{\text{norm},\mathcal{Q}}^{-1} P_{\mathcal{Q},\text{ML}} Y_{\mathcal{Q}}}{k}\Big)^2\right\} \\
&= \alpha^2 - 2\frac{k-2}{k}\alpha^2 + \frac{1}{k^2}\mathbb{E}\left\{\Big(Y_{\mathcal{Q}}^T P_{\mathcal{Q},\text{ML}}^T R_{\text{norm},\mathcal{Q}}^{-1} P_{\mathcal{Q},\text{ML}} Y_{\mathcal{Q}}\Big)^2\right\}. \qquad (2.35)
\end{aligned}$$

Let us define $U \triangleq R_{\text{norm},\mathcal{Q}}^{-\frac{1}{2}} P_{\mathcal{Q},\text{ML}} Y_{\mathcal{Q}}$. We have $\mathbb{E}\{U\} = R_{\text{norm},\mathcal{Q}}^{-\frac{1}{2}} P_{\mathcal{Q},\text{ML}} H_{\mathcal{Q}}\theta = 0$ and

$$\begin{aligned}
C_U = \mathbb{E}\{UU^T\} &= \alpha R_{\text{norm},\mathcal{Q}}^{-\frac{1}{2}} P_{\mathcal{Q},\text{ML}} R_{\text{norm},\mathcal{Q}} P_{\mathcal{Q},\text{ML}}^T R_{\text{norm},\mathcal{Q}}^{-\frac{1}{2}} \\
&= \alpha R_{\text{norm},\mathcal{Q}}^{-\frac{1}{2}}\Big(R_{\text{norm},\mathcal{Q}} - H_{\mathcal{Q}}\Big(H_{\mathcal{Q}}^T R_{\text{norm},\mathcal{Q}}^{-1} H_{\mathcal{Q}}\Big)^{-1} H_{\mathcal{Q}}^T\Big) \\
&\times \Big(I_{k\times k} - R_{\text{norm},\mathcal{Q}}^{-1} H_{\mathcal{Q}}\Big(H_{\mathcal{Q}}^T R_{\text{norm},\mathcal{Q}}^{-1} H_{\mathcal{Q}}\Big)^{-1} H_{\mathcal{Q}}^T\Big) R_{\text{norm},\mathcal{Q}}^{-\frac{1}{2}} \\
&= \alpha R_{\text{norm},\mathcal{Q}}^{-\frac{1}{2}}\Big(R_{\text{norm},\mathcal{Q}} - H_{\mathcal{Q}}\Big(H_{\mathcal{Q}}^T R_{\text{norm},\mathcal{Q}}^{-1} H_{\mathcal{Q}}\Big)^{-1} H_{\mathcal{Q}}^T\Big) R_{\text{norm},\mathcal{Q}}^{-\frac{1}{2}} \\
&= \alpha\Big(I_{k\times k} - R_{\text{norm},\mathcal{Q}}^{-\frac{1}{2}} H_{\mathcal{Q}}\Big(H_{\mathcal{Q}}^T R_{\text{norm},\mathcal{Q}}^{-1} H_{\mathcal{Q}}\Big)^{-1} H_{\mathcal{Q}}^T R_{\text{norm},\mathcal{Q}}^{-\frac{1}{2}}\Big). \qquad (2.36)
\end{aligned}$$

It can be easily confirmed that $\frac{1}{\alpha}C_U$ is a projection matrix. Therefore, using Lemma 6 and 7 we have:

$$\mathbb{E}\Big\{\Big(Y_{\mathcal{Q}}^T P_{\mathcal{Q}}^T R_{\text{norm},\mathcal{Q}}^{-1} P_{\mathcal{Q}} Y_{\mathcal{Q}}\Big)^2\Big\} = \mathbb{E}\Big\{\big(U^T U\big)^2\Big\} = \big[\text{trace}(C_U)\big]^2 + 21_k^T (C_U \bullet C_U) 1_k$$
$$= \alpha^2\Big((k-2)^2 + 21_k^T(\frac{1}{\alpha}C_U \bullet \frac{1}{\alpha}C_U)1_k\Big) = \alpha^2\Big((k-2)^2 + 2\text{trace}(\frac{1}{\alpha}C_U)\Big) = k(k-2)\alpha^2. \tag{2.37}$$

Substituting this value in Eq. 2.35 results in $\sigma^2_{\hat{\alpha}_{\text{ML}|\theta=\hat{\theta}_{\text{ML}}}} = \frac{2}{k}\alpha^2$. □

Remark 4. *Theorem 10 shows that the ML estimation error variance of the shadowing power, for both cases where the path loss parameters are perfectly known or estimated using an ML estimator, is independent of the sampling positions.*

Next we characterize the estimation of the shadowing power when the path loss parameters are estimated using an LS estimator, i.e. $\theta = \hat{\theta}_{\text{LS}}$.

Theorem 11. *Consider the case where multipath is negligible, i.e. $\sigma^2 = 0$. Let $\hat{\alpha}_{LS|\theta=\hat{\theta}_{LS}}$ represent the Least Square estimation of α when $\theta = \hat{\theta}_{LS}$. We then have the following,*

$$\mathbb{E}\Big\{\hat{\alpha}_{LS|\theta=\hat{\theta}_{LS}}\Big\} = \alpha\big[1 - \frac{1}{k}\textit{trace}(E)\big]$$
$$\sigma^2_{\hat{\alpha}_{LS|\theta=\hat{\theta}_{LS}}} = \frac{\alpha^2}{k^2}\Big[2\textit{trace}(R^2_{\textit{norm},\mathcal{Q}}) + \big[\textit{trace}(E)\big]^2 + 2\textit{trace}(E^2) - 4\textit{trace}(F)\Big],$$

where $E = \big(H_{\mathcal{Q}}^T H_{\mathcal{Q}}\big)^{-1} H_{\mathcal{Q}}^T R_{\textit{norm},\mathcal{Q}} H_{\mathcal{Q}}$ and $F = \big(H_{\mathcal{Q}}^T H_{\mathcal{Q}}\big)^{-1} H_{\mathcal{Q}}^T R^2_{\textit{norm},\mathcal{Q}} H_{\mathcal{Q}}$. Therefore, the estimation error variance is independent of θ while it is an increasing function of α.

Proof. We have the following Least Square estimator when $\theta = \hat{\theta}_{\text{LS}}$:

$$\hat{\alpha}_{\text{LS}|\theta=\hat{\theta}_{\text{LS}}} = \frac{Y_{\mathcal{Q}}^T P_{\mathcal{Q},\text{LS}}^T P_{\mathcal{Q},\text{LS}} Y_{\mathcal{Q}}}{k} \tag{2.38}$$

where $P_{\mathcal{Q},\text{LS}} = I_{k\times k} - H_{\mathcal{Q}}\Big(H_{\mathcal{Q}}^T H_{\mathcal{Q}}\Big)^{-1} H_{\mathcal{Q}}^T$ is the projection matrix. We have

$$\begin{aligned}\mathbb{E}\Big\{\hat{\alpha}_{\text{LS}|\theta=\hat{\theta}_{\text{LS}}}\Big\} &= \frac{1}{k}\text{trace}\Big[P_{\mathcal{Q},\text{LS}}\mathbb{E}\Big\{Y_{\mathcal{Q}}Y_{\mathcal{Q}}^T\Big\}\Big] \\ &= \frac{\alpha}{k}\text{trace}\left[R_{\text{norm},\mathcal{Q}} - H_{\mathcal{Q}}\Big(H_{\mathcal{Q}}^T H_{\mathcal{Q}}\Big)^{-1} H_{\mathcal{Q}}^T R_{\text{norm},\mathcal{Q}}\right] \\ &= \alpha\big[1 - \frac{1}{k}\text{trace}(E)\big]. \end{aligned} \tag{2.39}$$

We next characterize the error variance of this estimator:

$$\begin{aligned}\sigma^2_{\hat{\alpha}_{\text{LS}|\theta=\hat{\theta}_{\text{LS}}}} &= \mathbb{E}\left\{\Big(\alpha - \frac{Y_{\mathcal{Q}}^T P_{\mathcal{Q},\text{LS}}^T P_{\mathcal{Q},\text{LS}} Y_{\mathcal{Q}}}{k}\Big)^2\right\} \\ &= \alpha^2 - 2\alpha^2\big[1 - \frac{1}{k}\text{trace}(E)\big] + \frac{1}{k^2}\mathbb{E}\left\{\Big(Y_{\mathcal{Q}}^T P_{\mathcal{Q},\text{LS}}^T P_{\mathcal{Q},\text{LS}} Y_{\mathcal{Q}}\Big)^2\right\}. \end{aligned} \tag{2.40}$$

Define $U \triangleq P_{\mathcal{Q},\text{LS}} Y_{\mathcal{Q}}$. We have $U \sim \mathcal{N}(0, C_U)$ where

$$\begin{aligned} C_U &\triangleq \alpha P_{\mathcal{Q},\text{LS}} R_{\text{norm},\mathcal{Q}} P_{\mathcal{Q},\text{LS}}^T \\ &= \alpha R_{\text{norm},\mathcal{Q}} - \alpha R_{\text{norm},\mathcal{Q}} H_{\mathcal{Q}}\Big(H_{\mathcal{Q}}^T H_{\mathcal{Q}}\Big)^{-1} H_{\mathcal{Q}}^T - \alpha H_{\mathcal{Q}}\Big(H_{\mathcal{Q}}^T H_{\mathcal{Q}}\Big)^{-1} H_{\mathcal{Q}}^T R_{\text{norm},\mathcal{Q}} \\ &+ \alpha H_{\mathcal{Q}}\Big(H_{\mathcal{Q}}^T H_{\mathcal{Q}}\Big)^{-1} H_{\mathcal{Q}}^T R_{\text{norm},\mathcal{Q}} H_{\mathcal{Q}}\Big(H_{\mathcal{Q}}^T H_{\mathcal{Q}}\Big)^{-1} H_{\mathcal{Q}}^T. \end{aligned} \tag{2.41}$$

Therefore, we have $\text{trace}(C_U) = \alpha\big[k - \text{trace}(E)\big]$. After some lines of derivations, it can be confirmed that $1_k^T (C_U \bullet C_U) 1_k = \text{trace}(C_U^2) = \alpha^2\big(\text{trace}(R^2_{\text{norm},\mathcal{Q}}) + \text{trace}(E^2) - 2\text{trace}(F)\big)$. Substituting Eq. 2.41 in Eq. 2.40 results in:

$$\begin{aligned}\sigma^2_{\hat{\alpha}_{\text{LS}|\theta=\hat{\theta}_{\text{LS}}}} &= \alpha^2 - 2\alpha^2\big[1 - \frac{1}{k}\text{trace}(E)\big] \\ &+ \frac{\alpha^2}{k^2}\left[\big[k - \text{trace}(E)\big]^2 + 2\text{trace}(R^2_{\text{norm},\mathcal{Q}}) + 2\text{trace}(E^2) - 4\text{trace}(F)\right] \\ &= \frac{\alpha^2}{k^2}\Big[2\text{trace}(R^2_{\text{norm},\mathcal{Q}}) + \big[\text{trace}(E)\big]^2 + 2\text{trace}(E^2) - 4\text{trace}(F)\Big]. \end{aligned} \tag{2.42}$$

□

As β goes to zero, we have $\lim_{\beta\to 0} E = \lim_{\beta\to 0} F = I_{2\times 2}$. Thus, it can be verified that $\lim_{\beta\to 0}\mathbb{E}\Big\{\hat{\alpha}_{\text{LS}|\theta=\hat{\theta}_{\text{LS}}}\Big\} = \frac{k-2}{k}\alpha$, which is asymptotically unbiased. Moreover, we have

$\lim_{\beta\to 0}\sigma^2_{\hat{\alpha}_{\text{LS}|\theta=\hat{\theta}_{\text{LS}}}} = \frac{2\alpha^2}{k^2}$. For the case of $\beta = \infty$, we have $\lim_{\beta\to\infty} F = k\lim_{\beta\to\infty} E$. Similar to the derivations of Eq. 2.24, we can show that,

$$\lim_{\beta\to\infty} E = (H_{\mathcal{Q}}^T H_{\mathcal{Q}})^{-1}\left(H_{\mathcal{Q}}^T 1_k 1_k^T H_{\mathcal{Q}}\right) = \begin{bmatrix} k & -1_k^T D_{\mathcal{Q}} \\ 0 & 0 \end{bmatrix}. \tag{2.43}$$

Therefore, $\lim_{\beta\to\infty}\mathbb{E}\left\{\hat{\alpha}_{\text{LS}|\theta=\hat{\theta}_{\text{LS}}}\right\} = 0$ and as a result $\lim_{\beta\to\infty}\hat{\alpha}_{\text{LS}|\theta=\hat{\theta}_{\text{LS}}} = 0$. Furthermore, $\lim_{\beta\to\infty}\text{trace}(E^2) = k^2$ which results in $\lim_{\beta\to\infty}\sigma^2_{\hat{\alpha}_{\text{LS}|\theta=\hat{\theta}_{\text{LS}}}} = \alpha^2$. As β goes to ∞, the measurement vector becomes $Y_{\mathcal{Q}} = H\theta + \varrho 1_k$, where $\varrho \sim \mathcal{N}(0,\alpha)$. For $k \geqslant 2$, the path loss slope, n_{PL}, can be perfectly estimated. Then, we have $Y_{\mathcal{Q}} - (-n_{\text{PL}} D_{\mathcal{Q}}) = (K_{\text{dB}} + \varrho)1_k$. In this case, the LS estimation of α will be 0, which corresponds to the error variance of α^2. It should be noted that in this case, the channel can be perfectly predicted for $k \geq 2$. Similarly, in practice, as β goes to ∞, α would become zero. Thus, shadowing power of zero will be correctly estimated in practice.

2.5 Performance analysis using unbiased estimation of path loss parameters

In Section 2.2, we considered the impact of the underlying parameters on the spatial predictability of a wireless channel, where we assumed that the underlying parameters are estimated perfectly. In this section, we extend that analysis to also consider the impact of estimation error in path loss parameters. In [6], authors analyzed the sensitivity of channel predication to the estimation of the underlying parameters. Fig. 2.9 shows the impact of parameter estimation error on the overall channel prediction performance. For each curve, only one parameter is perturbed while the rest are assumed perfectly estimated. It can be seen that the curves attain their minima when there is no parameter estimation error, as expected. We can furthermore observe that uncertainty in the estimation of different parameters impacts the performance differently. As can be seen, the prediction is more

sensitive to path loss parameters (especially path loss exponent n_{PL}). In other words, the effect of an error in the estimation of the shadowing parameters is almost negligible, as compared to the error in path loss estimation. As such, in this section we extend the analysis of Section 2.2 to the case where errors in the estimation of path loss parameters are also considered.

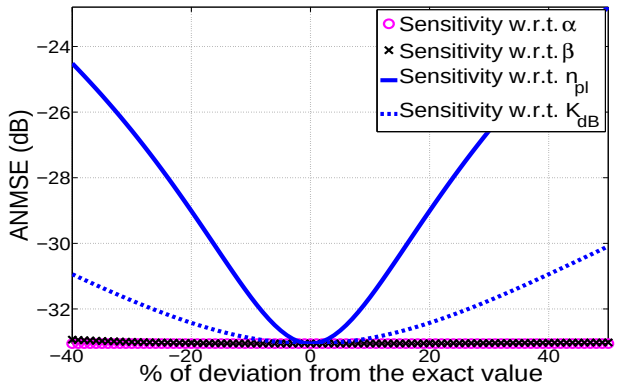

Figure 2.9: Average Normalized Mean Square Error (ANMSE), spatially averaged over different channel realization and random sampling positions, as a function of the % of estimation error in $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\theta}$ [6].

Consider the case where path loss parameters are estimated using an unbiased estimator.[5] We next characterize the error variance of channel prediction for this case, assuming that the error in the estimation of α, β and σ^2 is negligible. Since we are considering both the ML and LS estimators, we assume that $R_{\text{tot},\mathcal{Q}}$ is invertible in the rest of the section (This is naturally implied if $\sigma^2 \neq 0$). Let $\hat{\theta}_{\text{unb}} = SY_{\mathcal{Q}}$ denote an unbiased estimator of θ. We have the following for the error covariance matrix $C_{\theta,\,\text{unb}} = SR_{\text{tot},\mathcal{Q}}S^T$, with $SH_{\mathcal{Q}} = I_{k\times k}$. Let $\hat{\Upsilon}_{\text{dB},\mathcal{Q},\hat{\theta}_{\text{unb}}}(q) = \mathbb{E}\Big\{\Upsilon_{\text{dB}}(q) \;\Big|\; Y_{\mathcal{Q}}, \theta = \hat{\theta}_{\text{unb}}, \alpha, \beta, \sigma^2\Big\}$ denote the estimation of channel at position q, when path loss parameters are estimated using the aforementioned unbiased

[5]The unbiased estimator can be either ML or LS.

estimator:

$$\begin{aligned}\hat{\Upsilon}_{\text{dB},\mathcal{Q},\hat{\theta}_{\text{unb}}}(q) &= h^T(q)\hat{\theta}_{\text{unb}} + \phi_{\mathcal{Q}}^T(q)R_{\text{tot},\mathcal{Q}}^{-1}\big(Y_{\mathcal{Q}} - H_{\mathcal{Q}}\hat{\theta}_{\text{unb}}\big) \\ &= \underbrace{\Big[\Big(h^T(q) - \phi_{\mathcal{Q}}^T(q)R_{\text{tot},\mathcal{Q}}^{-1}H_{\mathcal{Q}}\Big)S + \phi_{\mathcal{Q}}^T(q)R_{\text{tot},\mathcal{Q}}^{-1}\Big]}_{G_{\mathcal{Q}}} Y_{\mathcal{Q}}.\end{aligned}$$

We have the following characterization for the error variance of channel estimation:

$$\begin{aligned}\sigma^2_{\text{dB},\mathcal{Q},\hat{\theta}_{\text{unb}}}(q) &\triangleq \mathbb{E}\Big\{\big(\Upsilon_{\text{dB}}(q) - \hat{\Upsilon}_{\text{dB},\mathcal{Q},\hat{\theta}_{\text{unb}}}(q)\big)^2 \;\Big|\; \alpha,\beta,\sigma^2\Big\} \\ &= E\Big\{\Big(h^T(q)\theta + \Xi_{\{q\}} - G_{\mathcal{Q}}Y_{\mathcal{Q}}\Big)\Big(h^T(q)\theta + \Xi_{\{q\}} - G_{\mathcal{Q}}Y_{\mathcal{Q}}\Big)^T\Big\} \\ &= \mathbb{E}\Big\{\big(\Xi_{\{q\}} - G_{\mathcal{Q}}\Xi_{\mathcal{Q}}\big)\big(\Xi_{\{q\}} - G_{\mathcal{Q}}\Xi_{\mathcal{Q}}\big)^T\Big\} \\ &= \alpha + \sigma^2 + G_{\mathcal{Q}}R_{\text{tot},\mathcal{Q}}G_{\mathcal{Q}}^T - G_{\mathcal{Q}}\phi_{\mathcal{Q}}(q) - \phi_{\mathcal{Q}}^T(q)G_{\mathcal{Q}}^T, \qquad (2.44)\end{aligned}$$

where $\Xi_{\{q\}} = \nu(q) + \omega(q)$ denotes the sum of shadowing and multipath power (in dB domain) at position q. It can be easily confirmed that

$$\begin{aligned}G_{\mathcal{Q}}R_{\text{tot},\mathcal{Q}}G_{\mathcal{Q}}^T = &\Big(h^T(q) - \phi_{\mathcal{Q}}^T(q)R_{\text{tot},\mathcal{Q}}^{-1}H_{\mathcal{Q}}\Big)SR_{\text{tot},\mathcal{Q}}S^T\Big(h^T(q) - \phi_{\mathcal{Q}}^T(q)R_{\text{tot},\mathcal{Q}}^{-1}H_{\mathcal{Q}}\Big)^T \\ &- \phi_{\mathcal{Q}}^TR_{\text{tot},\mathcal{Q}}^{-1}\phi_{\mathcal{Q}} + G_{\mathcal{Q}}\phi_{\mathcal{Q}}(q) + \phi_{\mathcal{Q}}^T(q)G_{\mathcal{Q}}^T. \qquad (2.45)\end{aligned}$$

Therefore, we have

$$\begin{aligned}\sigma^2_{\text{dB},\mathcal{Q},\hat{\theta}_{\text{unb}}}(q) = &\underbrace{\alpha + \sigma^2 - \phi_{\mathcal{Q}}^TR_{\text{tot},\mathcal{Q}}^{-1}\phi_{\mathcal{Q}}}_{\text{initial ch. est. error var. if path loss is perfectly known}} \\ &+ \underbrace{\Big(h^T(q) - \phi_{\mathcal{Q}}^T(q)R_{\text{tot},\mathcal{Q}}^{-1}H_{\mathcal{Q}}\Big)C_{\theta,\,\text{unb}}\Big(h^T(q) - \phi_{\mathcal{Q}}^T(q)R_{\text{tot},\mathcal{Q}}^{-1}H_{\mathcal{Q}}\Big)^T}_{\text{increase in error var. due to error propagation from est. of }\theta}. \qquad (2.46)\end{aligned}$$

The initial uncertainty of channel estimation at q can be represented by $\alpha+\sigma^2-\phi_{\mathcal{Q}}^TR_{\text{tot},\mathcal{Q}}^{-1}\phi_{\mathcal{Q}}$ if path loss is perfectly removed. Then, the second term of Eq. 2.46 is an increase in the error due to error propagation in the estimation of θ. As can be seen, $\sigma^2_{\text{dB},\mathcal{Q},\hat{\theta}_{\text{unb}}}(q)$ is not a function of θ since $\hat{\Upsilon}_{\text{dB},\mathcal{Q},\hat{\theta}_{\text{unb}}}(q)$ is an unbiased estimator of channel at position q.

In the previous sections, we showed that as α and σ^2 increase, the estimation of path loss parameters as well as channel prediction quality become worse. Thus, we expect

to have the same trend, when considering both path loss estimation error and channel predictability. We next study the effect of correlation distance on the overall performance. Similar to Section 2.3, we consider two cases of $\beta = 0, \infty$ and characterize the overall channel estimation error variance. Moreover, we propose an optimum positioning scheme that minimizes the overall channel estimation error for these two cases.

Theorem 12. *Let* $\mathcal{Q}^{opt}_{\sigma^2_{dB},\mathcal{Q},\hat{\theta}_{ML/LS}(q),\beta=0,\infty,\sigma^2\neq 0}$ [6] *denote the optimum positioning which minimizes the overall estimation error variance at* q *for both* $\beta = 0$ *and* ∞ *and* $\sigma^2 \neq 0$*, considering path loss estimation through either ML or LS. We have* $\frac{1}{k}1_k^T D_{\mathcal{Q}^{opt}_{\sigma^2_{dB},\mathcal{Q},\hat{\theta}_{ML/LS}(q),\beta=0,\infty,\sigma^2\neq 0}} = D_{\{q\}}$.

Proof. For $\beta = \infty$, we have $\lim_{\beta\to\infty}\phi_{\mathcal{Q}}(q) = \frac{\rho}{1+\rho}\chi 1_k, \forall q \in \mathcal{K}$, where $\rho = \frac{\alpha}{\sigma^2}$ and $\chi = \alpha + \sigma^2$. If $\sigma^2 \neq 0$, then using properties 3 and 4 of Section 2.3-B, we get $\lim_{\beta\to\infty} h^T(q) - \phi^T_{\mathcal{Q}}(q)R^{-1}_{\text{tot},\mathcal{Q}}H_{\mathcal{Q}} = \left[1 - \frac{\rho}{1+\rho}\chi \lim_{\beta\to\infty} 1_k^T R^{-1}_{\text{tot},\mathcal{Q}} 1_k \quad -D_{\{q\}} + \frac{\rho}{1+\rho}\chi \lim_{\beta\to\infty} 1_k^T R^{-1}_{\text{tot},\mathcal{Q}} D_{\mathcal{Q}}\right] = \left[\frac{1}{1+\rho k} \quad \frac{\rho}{1+\rho k}1_k^T D_{\mathcal{Q}} - D_{\{q\}}\right]$ and $\lim_{\beta\to\infty}\phi^T_{\mathcal{Q}}R^{-1}_{\text{tot},\mathcal{Q}}\phi_{\mathcal{Q}} = \frac{\rho^2 k}{(1+\rho)(1+\rho k)}\chi$. Moreover, from Eq. 2.19 and Theorem 7, we have $C_{\theta,\text{ML/LS},\beta=\infty} \triangleq \lim_{\beta\to\infty} C_{\theta,\text{ML}} = \lim_{\beta\to\infty} C_{\theta,\text{LS}} = \frac{\chi}{1+\rho}\begin{bmatrix} \rho + \frac{D^T_{\mathcal{Q}}D_{\mathcal{Q}}}{D^T_{\mathcal{Q}}A_k D_{\mathcal{Q}}} & \frac{1_k^T D_{\mathcal{Q}}}{D^T_{\mathcal{Q}}A_k D_{\mathcal{Q}}} \\ \frac{1_k^T D_{\mathcal{Q}}}{D^T_{\mathcal{Q}}A_k D_{\mathcal{Q}}} & \frac{k}{D^T_{\mathcal{Q}}A_k D_{\mathcal{Q}}} \end{bmatrix}$, where $A_k = kI_{k\times k} - 1_k 1_k^T$. After some lines of derivations, it can be shown that

$$\begin{aligned}
&\lim_{\beta\to\infty}\Big(h^T(q) - \phi^T_{\mathcal{Q}}(q)R^{-1}_{\text{tot},\mathcal{Q}}H_{\mathcal{Q}}\Big)C_{\theta,\text{ML/LS},\beta=\infty}\Big(h^T(q) - \phi^T_{\mathcal{Q}}(q)R^{-1}_{\text{tot},\mathcal{Q}}H_{\mathcal{Q}}\Big)^T \\
&= \frac{D^T_{\mathcal{Q}}D_{\mathcal{Q}} + \rho\big(1_k^T D_{\mathcal{Q}}\big)^2 - 2(1+\rho k)1_k^T D_{\mathcal{Q}} D_{\{q\}} + k(1+\rho k)D^2_{\{q\}}}{(1+\rho)(1+\rho k)D^T_{\mathcal{Q}}A_k D_{\mathcal{Q}}}\chi \\
&= \frac{\big(1_k^T D_{\mathcal{Q}} - kD_{\{q\}}\big)^2\rho + D^T_{\mathcal{Q}}D_{\mathcal{Q}} - 21_k^T D_{\mathcal{Q}}D_{\{q\}} + kD^2_{\{q\}}}{(1+\rho)(1+\rho k)D^T_{\mathcal{Q}}A_k D_{\mathcal{Q}}}\chi. \qquad (2.47)
\end{aligned}$$

[6] The notation $\hat{\theta}_{\text{ML/LS}}$ denotes that the estimation of path loss parameters can be either $\hat{\theta}_{\text{ML}|\alpha,\beta,\sigma^2}$ or $\hat{\theta}_{\text{LS}}$.

Thus, we have the following for Eq. 2.46:

$$\begin{aligned}
&\lim_{\beta\to\infty} \sigma^2_{\text{dB},\mathcal{Q},\hat{\theta}_{\text{ML/LS}}}(q) \\
&= \chi\left[1 - \frac{\rho^2 k}{(1+\rho)(1+\rho k)} + \frac{\left(1_k^T D_{\mathcal{Q}} - kD_{\{q\}}\right)^2\rho + D_{\mathcal{Q}}^T D_{\mathcal{Q}} - 21_k^T D_{\mathcal{Q}} D_{\{q\}} + kD_{\{q\}}^2}{(1+\rho)(1+\rho k) D_{\mathcal{Q}}^T A_k D_{\mathcal{Q}}}\right] \\
&= \frac{Q_1\rho + Q_2}{(1+\rho)(1+\rho k) D_{\mathcal{Q}}^T A_k D_{\mathcal{Q}}}\chi, \qquad (2.48)
\end{aligned}$$

where $Q_1 = (k+1)D_{\mathcal{Q}}^T A_k D_{\mathcal{Q}} + \left(1_k^T D_{\mathcal{Q}} - kD_{\{q\}}\right)^2$ and $Q_2 = D_{\mathcal{Q}}^T A_k D_{\mathcal{Q}} + D_{\mathcal{Q}}^T D_{\mathcal{Q}} - 21_k^T D_{\mathcal{Q}} D_{\{q\}} + kD_{\{q\}}^2$. It can be easily confirmed that $Q_1 = kQ_2$. Therefore, we have $\lim_{\beta\to\infty} \sigma^2_{\text{dB},\mathcal{Q},\hat{\theta}_{\text{ML/LS}}}(q) = \frac{Q_2}{(1+\rho) D_{\mathcal{Q}}^T A_k D_{\mathcal{Q}}}\chi = \frac{Q_2}{D_{\mathcal{Q}}^T A_k D_{\mathcal{Q}}}\sigma^2$. Moreover, we have:

$$\lim_{\beta\to 0} \sigma^2_{\text{dB},\mathcal{Q},\hat{\theta}_{\text{ML/LS}}}(q) = \left(1 + h^T(q)\left(H_{\mathcal{Q}}^T H_{\mathcal{Q}}\right)^{-1} h(q)\right)\chi = \frac{Q_2}{D_{\mathcal{Q}}^T A_k D_{\mathcal{Q}}}\chi. \qquad (2.49)$$

Thus, the optimum positioning which minimizes channel estimation error variance for both $\beta = 0$ and ∞ is $\mathcal{Q}^{\text{opt}}_{\sigma^2_{\text{dB},\mathcal{Q},\hat{\theta}_{\text{ML/LS}}}(q),\beta=0,\infty,\sigma^2\neq 0} = \arg\min_{\mathcal{Q}} \frac{Q_2}{D_{\mathcal{Q}}^T A_k D_{\mathcal{Q}}}$. We have,

$$\begin{aligned}
\frac{Q_2}{D_{\mathcal{Q}}^T A_k D_{\mathcal{Q}}} &= 1 + \frac{D_{\mathcal{Q}}^T D_{\mathcal{Q}} - 21_k^T D_{\mathcal{Q}} D_{\{q\}} + kD_{\{q\}}^2}{D_{\mathcal{Q}}^T A_k D_{\mathcal{Q}}} = 1 + \frac{1}{k}\frac{D_{\mathcal{Q}}^T A_k D_{\mathcal{Q}} + \left(1_k^T D_{\mathcal{Q}} - kD_{\{q\}}\right)^2}{D_{\mathcal{Q}}^T A_k D_{\mathcal{Q}}} \\
&= 1 + \frac{1}{k} + \frac{1}{k}\frac{\left(1_k^T D_{\mathcal{Q}} - kD_{\{q\}}\right)^2}{D_{\mathcal{Q}}^T A_k D_{\mathcal{Q}}}.
\end{aligned}$$

It can be easily confirmed that A_k is positive-semidefinite. Thus, under full rank assumption of $H_{\mathcal{Q}}$ (as discussed in Section 2.1), we have $D_{\mathcal{Q}}^T A_k D_{\mathcal{Q}} > 0$. Therefore, to minimize the estimation error variance for both $\beta = 0$ and ∞, we need $1_k^T D_{\mathcal{Q}} = kD_{\{q\}}$ for $D_{\mathcal{Q}} \in \mathbb{R}^k \backslash \text{span}\{1_k\}$. □

Case of $\sigma^2 = 0$ and $\beta = 0$ can be treated the same as Eq. 2.49 with $\chi = \alpha$, which results in the same optimum positioning scheme. However, if $\sigma^2 = 0$ and $\beta = \infty$, for $k \geq 2$, the channel variations can be perfectly estimated at each point. Theorem 12 shows that the optimum positioning results in $\|q - q_b\| = \left(\prod_{i=1}^k \|q - q_i\|\right)^{\frac{1}{k}}$. This suggests that the optimum measurement positions should be chosen such that the distance of q to the

transmitter be the geometric average of the distances of the measurement points to the transmitter. Let $\mathcal{Q}' = \mathcal{Q}^{\text{opt}}_{\sigma^2_{\text{dB},\mathcal{Q},\hat{\theta}_{\text{ML/LS}}}(q),\beta=0,\infty,\sigma^2\neq 0}$ denote the optimum positioning. We have,

$$\lim_{\beta\to 0}\sigma^2_{\text{dB},\mathcal{Q}',\hat{\theta}_{\text{ML/LS}}}(q) = \alpha+\sigma^2+\frac{\alpha+\sigma^2}{k} \text{ and } \lim_{\beta\to\infty}\sigma^2_{\text{dB},\mathcal{Q}',\hat{\theta}_{\text{ML/LS}}}(q) = \sigma^2+\frac{\sigma^2}{k}. \quad (2.50)$$

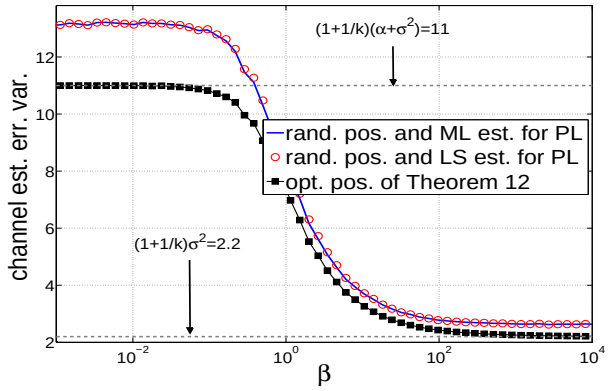

Figure 2.10: Impact of β on the channel predictability, when considering path loss estimation error.

For $\beta = 0$, as k goes to ∞, the estimation error of path loss parameters goes to 0 and the estimation error variance becomes $\alpha+\sigma^2$. This value is an initial uncertainty assuming known path loss parameters. For the case of $\beta = \infty$, on the other hand, the estimation error variance becomes σ^2 as k goes to ∞. Fig. 2.10 shows the impact of correlation distance β on the estimation performance when path loss parameters are estimated using an ML/LS estimator. The impact of optimum positioning of Theorem 12 can also be seen from the figure. For this example, the workspace is a 2m $\times$ 2m square with $\sqrt{\alpha} = \sqrt{8}$dB and $\sigma = \sqrt{2}$dB, where $k = 10$ samples are taken from the workspace (either randomly or optimally). The y axis then represents the estimation performance after averaging over several runs of channel realization and sampling patterns. As can be seen, ML and LS estimators provide very similar performance in this case. Furthermore, optimizing the position of the samples, according to Theorem 12 can improve the performance considerably.

2.6 Numerical analysis on real channel measurements

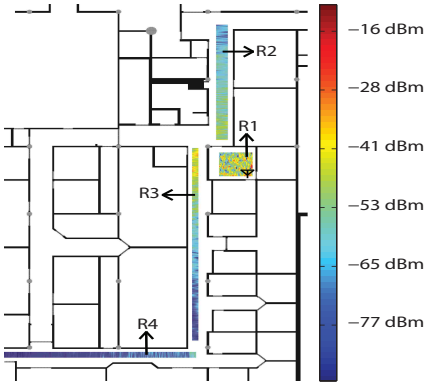

Figure 2.11: Blueprint of the portion of the basement of the ECE bldg. where channel measurements are collected. A colormap of the measured received signal power is superimposed on the map. R1 denotes the main room, where the transmitter is located (as marked on the figure). R2, R3 and R4 correspond to different hallways at the basement. See the PDF file for a colored version of the map.

In this section, we show the impact of different environments, and their underlying parameters, on channel predictability, using real channel measurements. Fig. 2.11 shows the blueprint of a portion of the basement of the Electrical and Computer Engineering building. We used a Pioneer-AT robot to make several measurements along different routes in the basement, in order to map the received signal strength (each route is a straight line). The robot is equipped with an 802.11g wireless card, with transmission at 2.4 GHz. It uses the MadWiFi device driver to measure the received signal power [96]. The figure also shows a color-map of our measured received signal power. In order to see the impact of different underlying parameters on channel predictability, the area is divided into four regions of $R1, \cdots, R4$, as can be seen from the figure. Since we are dealing with real data, we can not check the accuracy of the estimation of the underlying parameters. As such, we use all the measurements in each region, to estimate the underlying parameters of that region, which are then used to understand channel predictability of each region. We use the LS estimator of Section 2.1, in order to estimate channel parameters of each region.

As can be seen, as the distance to the transmitter increases, n_{PL} (the slope of path loss) increases. This phenomena has previously been reported in the literature as well [105]. Another interesting phenomenon is the shadowing behavior. As can be seen, correlation distance (β) increases as we get farther from the transmitter and move to the hallways. This makes sense as shadowing is the result of the transmitted signal being possibly blocked by a number of obstacles before reaching the receiver. Finally, for region R1 (the main room), multipath fading is the dominant term, as can be seen. This is expected since that room is rich in scatterers and reflectors, with no major obstacle. Next, we consider channel predictability of different regions and relate the observed behaviors to the underlying parameters of Table 2.1 .

Table 2.1: Channel Parameters for Different Regions and Routes

Region	K_{dB}	n_{PL}	α	β	σ^2
R1	-20.8870	1.2272	negligible	negligible	22.1238
R2	-21.4677	2.3878	10.7772	0.0979	2.8862
R3	-17.9694	2.9795	8.6385	0.3231	7.6628
R4	68.7836	9.9392	2.0157	1.4377	7.5687
A1	-	-	8.2164	0.0809	2.9721
A2	-	-	11.6332	0.0860	2.9313
B1	-	-	11.7535	0.2858	6.3979
B2	-	-	11.6029	0.5832	6.1956
C1	-	-	10.4193	0.2258	5.1696
C2	-	-	10.3451	0.2396	7.2873

Fig. 2.12 compares channel prediction quality of different regions (measured by Average Normalized MSE as defined in Section 2.1), given the parameters of Table 2.1. As can be seen, region R4 has the best performance, as compared to the other regions. From Table 2.1, region R4 has the smallest $\alpha + \sigma^2$ and good amount of correlation, which result in better predictability. On the other hand, region R1 experiences considerable multipath fading and negligible shadowing, which results in the worst predictability. Regions R2 and R3 have similar performances, since one has a higher shadowing correlation while the other experiences lower $\alpha + \sigma^2$. As was shown earlier, path loss parameters do not impact

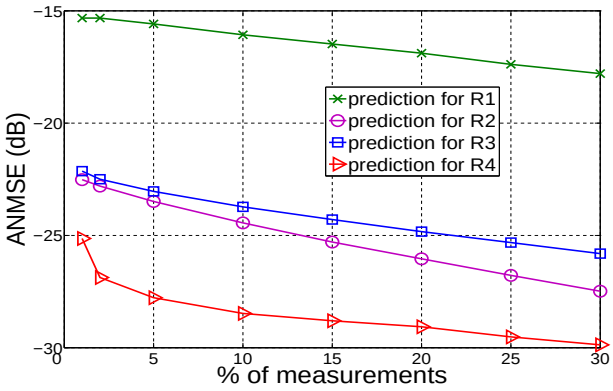

Figure 2.12: Comparison of channel predictability for different regions of Fig. 2.11.

channel predictability. We next study the impact of each individual channel parameter on the estimation performance more closely.

Table 2.1 also shows channel parameters corresponding to three pairs of routes in the basement of ECE building (pairs A, B and C). Each pair is chosen such that only one parameter changes and the rest are almost the same. Fig. 2.13 (top) shows the impact of the shadowing power on the estimation performance. As can be seen, for A1/A2 pair, the correlation distance and multipath power are almost the same. However, A1 has a smaller shadowing power, which results in a better estimation performance. Fig. 2.13 (middle) and Fig. 2.13 (bottom) show the impact of correlation distance and multipath power on the estimation performance respectively. For each case, other channel parameters are almost the same. As can be seen, B2 with its higher correlation distance and C1 with its smaller multipath power provide better predictability.

2.7 Summary

In this chapter, we developed a probabilistic channel prediction framework for predicting the spatial variations of a wireless channel, based on a small number of measurements.

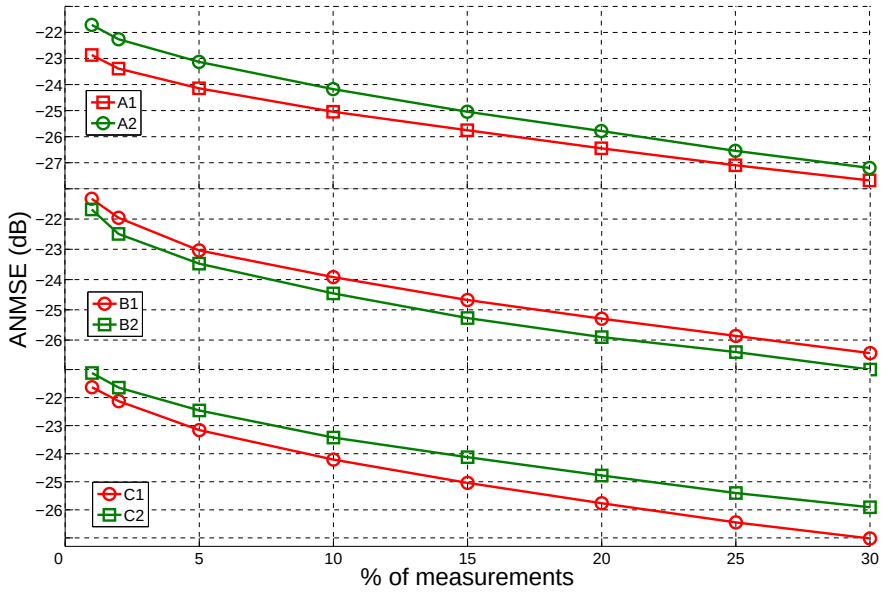

Figure 2.13: Impact of (top) shadowing power, (middle) correlation distance and (bottom) multipath power on channel prediction performance, using real channel measurements of Fig. 2.11.

We then proposed a mathematical foundation for understanding the spatial predictability of wireless channels. More specifically, we characterized the impact of different environments, in terms of their underlying parameters, on wireless channel predictability. We furthermore showed how sampling positions can be optimized to improve the prediction quality. Finally, we showed the performance of the proposed framework in predicting (and justifying the predictability of) the spatial variations of real channels, using several measurements in our building. Overall, the proposed probabilistic framework of this chapter can be integrated with motion planning algorithms in robotic networks applications and improve the connectivity of the mobile robots while accomplishing their task.

Chapter 3

An integrated sparsity and model-based framework for channel prediction

In the previous chapter, we proposed a probabilistic framework to predict the spatial variations of the channel and characterized the corresponding channel predictability in different environments. In this chapter, we start by showing how the sparsity of the wireless channels in the frequency domain can be utilized for channel prediction based on sparse measurements. We then propose an integrated sparsity and model-based framework that can keep the strengths of both approaches in order to design an estimator with even a better performance.

Consider the workspace $\mathcal{K}$. We assume that the workspace is discretized into an ordered set of points $\mathcal{P}$. Let vector $x \in \mathbb{R}^N$ represent the corresponding received signal strength over $\mathcal{P}$, where $N = |\mathcal{P}|$. Consider the case where the received signal strength to the base station is sparsely sampled at positions $\mathcal{Q} = \{q_1, q_2, \cdots, q_K\} \subset \mathcal{P}$ over the workspace, with K representing the total number of gathered measurements. The channel measurements can be gathered by one or a number of cooperative robots, making measurements along their trajectories. Thus, the measurements can be collected at the same time

or at different time instants over the workspace.[1] Define $y_i \triangleq \Upsilon_{\text{dB}}(q_i)$ for $1 \leq i \leq K$. Let $Y_{\mathcal{Q}} = [y_1, \cdots, y_K]^T \in \mathbb{R}^K$ represent the vector of all the gathered channel measurements (in dB). We then have $Y_{\mathcal{Q}} = \Phi_{\mathcal{Q}} x$, where $\Phi_{\mathcal{Q}}$ represents a $K \times N$ sampling matrix that corresponds to $\mathcal{Q}$. More specifically, the ith row of $\Phi_{\mathcal{Q}}$ has all zero entries, except for the entry that corresponds to q_i, which becomes one. In this chapter, it is our goal to predict the received signal strength[2], at unvisited locations (set $\mathcal{P} \setminus \mathcal{Q}$), i.e. estimate vector x from $Y_{\mathcal{Q}}$, where $K \ll N$.

The chapter is organized as follows. In Section 3.1, we provide a brief introduction to the theory of compressed sensing. In Section 3.2 we show how channel sparsity in the frequency domain can be utilized for estimating the spatial variations of a wireless channel, based on a small number of measurements. In Section 3.3 we analyze the impact of the underlying channel parameters on the sparsity of the channel in the frequency domain. In Section 3.4 we show the underlying tradeoffs between the sparsity-based approach and probabilistic framework of Chapter 2. Section 3.5 then proposes an integrated framework that combines the strengths of both and shows its superior performance, using real channel measurements. A summary of the results of the chapter is provided in Section 3.6.

3.1 An Overview of Compressive sampling theory [1–4]

The new theory of sampling is based on the fact that real-world signals typically have a sparse representation in a certain transformed domain. However, in most of the signal processing applications, the signal of interest is first fully sampled, after which a transformation is applied and only the coefficients above a certain threshold are saved. This

[1] As mentioned earlier in Section 2.1, we assume that the channel field is not changing with time, in our modeling and prediction. Thus, we only need to consider the spatial variations of the measurements.

[2] While we pose our framework based on the prediction of the received signal strength (or power), we use the terms "channel prediction" and "received signal strength prediction" interchangeably in Chapter 2 and 3.

is not efficient as it puts a heavy burden on sampling the entire signal while only a small percentage of the transformed coefficients were needed to represent it. The new theory of compressive sampling, on the other hand, allows us to sense the signal in a compressed manner to begin with.

A *sparse* signal is a signal that can be represented with a small number of non-zero coefficients. A *compressible* signal is a signal that has a transformation where most of its energy is in a very few coefficients, making it possible to approximate the rest with zero. The new theory of compressive sampling [106] shows that, under certain conditions, a compressible signal can be reconstructed using very few observations. Most natural signals are indeed compressible. The best sparse representation of a signal depends on the application and can be inferred from analyzing similar data. Consider a scenario where we are interested in recovering a vector $x \in \mathbb{R}^N$. In our case, x represents the received signal strength over the field of interest. We refer to the domain of x as the primal domain. For 2D signals, vector x can represent the columns of the matrix of interest stacked up to form a vector. Let $z \in \mathbb{R}^K$, where $K \ll N$, represent the incomplete linear measurements of vector x obtained by the sensors. We will have

$$z = \Phi x, \tag{3.1}$$

where we refer to Φ as the observation matrix. Clearly, solving for x based on the observation set z is an ill-posed problem as the system is severely under-determined ($K \ll N$). However, suppose that x has a sparse representation in another domain, i.e., it can be represented as a linear combination of a small set of vectors: $x = \Gamma X$, where Γ is an invertible matrix and X is S-sparse, i.e., $|\text{supp}(X)| = S \ll N$, where supp($X$) refers to the set of indices of the non-zero elements of X and $|\cdot|$ denotes its cardinality. This means that the number of non-zero elements in X is considerably smaller than N. Then we will have

$$z = \Psi X, \tag{3.2}$$

where $\Psi = \Phi \times \Gamma$. We refer to the domain of X as the sparse domain (or transform domain). If $S \leq K$ and we knew the positions of the non-zero coefficients of X, we could

solve this problem with traditional techniques like least-squares. In general, however, we do not know anything about the structure of X except for the fact that it is sparse (which we can validate by analyzing similar data). The new theory of compressive sensing allows us to solve this problem.

Theorem 1. *(see [106] for details and the proof): If $K \geq 2S$ and under specific conditions, the desired X is the solution to the following optimization problem:*

$$min||X||_0, \text{ such that } z = \Psi X, \tag{3.3}$$

where $||X||_0 = |supp(X)|$ represents the zero norm of vector X.

Theorem 1 states that we only need $2 \times S$ measurements to recover X and therefore x fully. This theorem, however, requires solving a non-convex combinatorial problem, which is not practical.

Instead, consider the following ℓ_1 relaxation of the aforementioned ℓ_0 optimization problem:

$$\text{min}||X||_1, \text{ subject to } z = \Psi X. \tag{3.4}$$

Theorem 2. *(see [1, 107, 108] for details) Assume that X is S-sparse. The ℓ_1 relaxation can exactly recover X from measurement z if matrix Ψ satisfies the Restricted Isometry Condition (RIC) for $(2S, \sqrt{2} - 1)$.*

Restricted Isometry Condition (RIC) [109, 110]: Matrix Ψ satisfies the RIC with parameters (Z, ϵ) for $\epsilon \in (0, 1)$ if

$$(1 - \epsilon)||c||_2 \leq ||\Psi c||_2 \leq (1 + \epsilon)||c||_2 \tag{3.5}$$

for all Z-sparse vector c.

While it is not possible to define all the classes of matrices Ψ that satisfy RIC, it is shown that random partial Fourier matrices [111] satisfy RIC with the probability $1 -$

$O(N^{-M})$ if $K \geq B_M S \times \log^{O(1)} N$, where B_M is a constant, M is an accuracy parameter and O(·) is Big-O notation [106]. This shows that the number of required measurements could be considerably less than N. While the recovery of sparse signals is important, in practice signals may rarely be sparse. Most signals, however, will be compressible. In practice, the observation vector z will also be corrupted by noise. The ℓ_1 relaxation and the corresponding required RIC condition can be easily extended to the case of noisy observations with compressible signals [107].

The ℓ_1 optimization problem of Eq. 3.4 can be posed as a linear programming problem [112]. The compressive sensing algorithms that reconstruct the signal based on ℓ_1 optimization are typically referred to as "Basis Pursuit" [1]. The Restricted Isometry Condition also implies that the columns of matrix Ψ should have a certain near-orthogonality property. Matching Pursuit (MP) approaches, on the other hand, are another class of algorithms that use this property to iteratively reconstruct the signal with less computational complexity. Readers are referred to [110, 113] for more details on this. Next, we summarize sparsity-based channel estimation using a small number of measurements.

3.2 Sparsity-based prediction of the channel spatial variations

In this part, we are interested in predicting the spatial variations of the received signal strength at unvisited locations, based on sparse measurements. Our analysis of several channel measurements has shown that wireless channels are compressible in the frequency domain for several scenarios. For instance, the solid curve of Fig. 3.1 (left) shows a sample channel measurement across a street in San Francisco (data is courtesy of Mark Smith [5]). The dashed curve shows the sparsified version of this channel, where only 3% of its ordered Fourier coefficients are retained (ordered decreasingly) while the rest are zeroed.

As can be seen, only 3% of the Fourier coefficients can capture the spatial variations of the channel well. Fig. 3.1 (right) measures the sparsity of the channel in the frequency domain, for different percentages of the retained Fourier coefficients. The y-axis shows -10log(NMSE), where NMSE denotes the Normalized Mean Square Error of the difference between the channel and its sparsified version. Then, the plot characterizes how compressible this channel is. As can be seen, this channel is fairly compressible, i.e., a small percentage of the Fourier coefficients suffices for capturing the signal.

We have also investigated the sparsity of the wireless channels, using other basis, such as wavelet and Legendre [114]. While the channel can possibly be very compressible in the wavelet domain, sampling in the spatial domain and reconstructing based on the wavelet transformation results in a poor quality. This is due to the fact that spatial point-sampling and wavelet basis are not incoherent, resulting in the corresponding Ψ not satisfying the RIC condition [41]. In [114], the authors show that reconstruction based on the random sampling and Legendre basis meets the RIC condition, suggesting a possible recovery strategy if the signal can be compressible in the Legendre basis. Our analysis of several real channel measurements, however, shows that Fourier domain provides a considerably

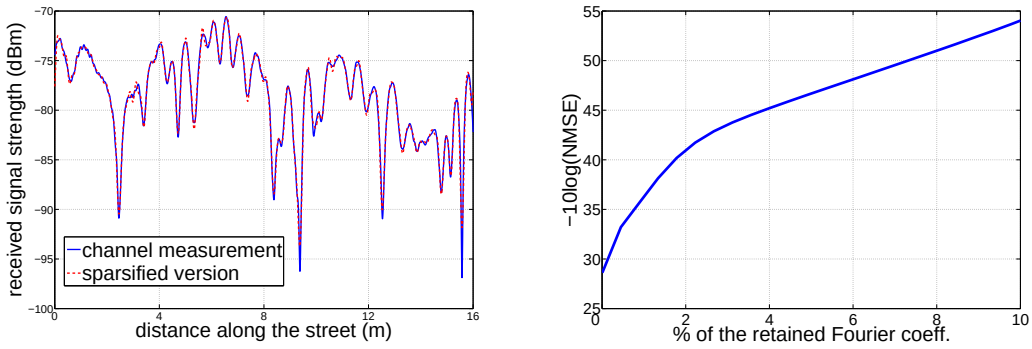

Figure 3.1: (left) channel measurement across a street in San Francisco [5], along with its sparsified version, when only 3% of its ordered Fourier coefficients are retained, and (right) measuring the compressibility of the left channel.

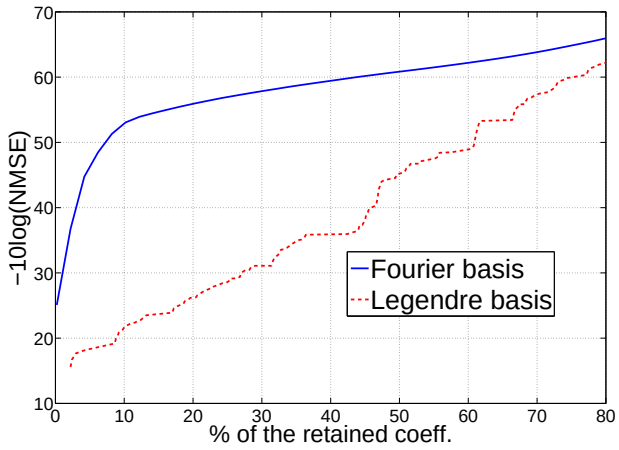

Figure 3.2: Comparison of channel compressibility in Fourier and Legendre basis, for a chunk of the outdoor channel of Fig. 3.1 (left). It can be seen that the channel is considerably more compressible in Fourier basis.

more compressible representation than Legendre basis. Fig. 3.2, for instance, compares the compressibility of the channel measurement gathered in San Francisco, based on both Fourier and Legndre basis. It can be seen that a wireless signal is considerably more compressible in the Fourier domain. This, accompanied with the fact that the computational complexity of Fourier transformation is also considerably lower, makes Fourier an appropriate domain for our sparsity-based channel reconstruction.

Consider the workspace $\mathcal{K}$. As introduced earlier in this chapter, $x \in \mathbb{R}^N$ and $Y_{\mathcal{Q}}$ represent the vector of the spatial variations of the channel, corresponding to $\mathcal{P}$, and the vector of all the available channel measurements respectively (both in dB). We have $Y_{\mathcal{Q}} = \Phi_{\mathcal{Q}} x$, where $\Phi_{\mathcal{Q}}$ represents a $K \times N$ sampling matrix that corresponds to $\mathcal{Q}$. Then, in the context of compressive sensing (Section 3.1), vector x of Eq. 3.1 represents the spatial variations of the channel and $z = Y_{\mathcal{Q}}$ denotes the vector of the sparse available channel measurements. Therefore, the sparsity-based estimation of the channel spatial variations,

using compressibility in the frequency domain, can be posed as follows:

$$\tilde{X}_{\text{sparse}} = \arg\min ||X||_1 \quad \text{s.t. } Y_{\mathcal{Q}} = \Phi_{\mathcal{Q}} \Gamma X, \tag{3.6}$$

where Γ denotes the inverse Fourier matrix and X represents the Fourier transformation of x. For a 1D case, we have $[\Gamma_{\text{1D}}]_{m,n} = \frac{1}{\sqrt{N}} e^{j\frac{2\pi}{N}(m-1)(n-1)}$ for $1 \leq m, n \leq N$. For a 2D scenario, consider the case where the size of the discretized workspace is $J_x \times J_y$, where $J_x \times J_y = N$. Define $\kappa_1(M_1, M_2) \triangleq \lceil \frac{M_1}{M_2} \rceil$, where $\lceil . \rceil$ denotes the smallest integer greater than or equal to the argument and $\kappa_2(M_1, M_2) \triangleq M_1 - M_2\big[\kappa_1(M_1, M_2) - 1\big]$, for arbitrary variables M_1 and $M_2 \in \mathbb{N}$. We have, $\Big[\Gamma_{\text{2D}}\Big]_{m,n} = \frac{1}{\sqrt{J_x J_y}} e^{j\frac{2\pi}{J_y}\left(\kappa_1(m,J_x)-1\right)\left(\kappa_1(n,J_x)-1\right) + j\frac{2\pi}{J_x}\left(\kappa_2(m,J_x)-1\right)\left(\kappa_2(n,J_x)-1\right)}$, for $1 \leq m, n \leq N$.

3.3 Impact of the channel underlying parameters on the variations of channel frequency response

So far, we established that a wireless channel can be considerably compressible in the Fourier domain, based on examining real channel measurements. In this section, our goal is to mathematically characterize channel compressibility in the frequency domain as this directly impacts the performance of our sparsity-based estimator. More specifically, we characterize the impact of the underlying channel parameters, i.e. parameters of path loss, shadowing and multipath, on the variations of the frequency components of a wireless channel. This analysis shows how different parameters can impact the compressibility of the channel and the resulting performance of the sparsity-based estimator. For this analysis, consider the case where a wireless channel is measured along a route. Without loss of generality, we assume that the channel is sampled across x-axis at equally-distanced positions, where the distance between two consecutive sampling positions is d and the base station is located at the origin. Let $\mathcal{T}$ denote the set of all the sampled positions. We have

$\mathcal{T} = \{q_1 = d, q_2 = 2d, \cdots, q_N = Nd\}$ and $D_{\mathcal{T}} = \left[10\log_{10}(d), \cdots, 10\log_{10}(Nd)\right]^T$. The measurement vector, $Y_{\mathcal{T}}$, can then be represented by the following, for this 1D case, based on Eq. 2.2,

$$Y_{\mathcal{T}} = K_{\text{dB}} 1_N - n_{\text{PL}} D_{\mathcal{T}} + \vartheta_{\mathcal{T}} + \Omega_{\mathcal{T}}, \tag{3.7}$$

where similar to the Section 2.1, $\vartheta_{\mathcal{T}}$ (impact of shadowing) is a zero-mean Gaussian random vector with the covariance matrix $R_{\mathcal{T}} \in \mathbb{R}^{N\times N}$, where $\left[R_{\mathcal{T}}\right]_{i,j} = \alpha\, e^{-|i-j|d/\beta}$ for $1 \leq i, j \leq N$. Furthermore, for this analysis, we assume that $\Omega_{\mathcal{T}}$ (impact of multipath fading) is a zero-mean Gaussian vector with the covariance matrix $\sigma^2 I_{N\times N}$ (as discussed earlier in Section 2.1). Next, we characterize the impact of different channel parameters on the variations of the frequency response of a wireless channel. Let ${\Gamma_{\text{1D}}}^{-1}$ denote the 1D Fourier transform matrix with entries $[{\Gamma_{\text{1D}}}^{-1}]_{m,n} = \frac{1}{\sqrt{N}} e^{-j\frac{2\pi}{N}(m-1)(n-1)}$ for $1 \leq m, n \leq N$. We have the following for the frequency response of the channel:

$$\begin{aligned} Y_{\mathcal{T},\mathcal{F}} \triangleq {\Gamma_{\text{1D}}}^{-1} Y_{\mathcal{T}} &= K_{\text{dB}} {\Gamma_{\text{1D}}}^{-1} 1_N - n_{\text{PL}} {\Gamma_{\text{1D}}}^{-1} D_{\mathcal{T}} + {\Gamma_{\text{1D}}}^{-1}\vartheta_{\mathcal{T}} + {\Gamma_{\text{1D}}}^{-1}\Omega_{\mathcal{T}} \\ &= \underbrace{\sqrt{N} K_{\text{dB}} e_1 - n_{\text{PL}} {\Gamma_{\text{1D}}}^{-1} D_{\mathcal{T}}}_{\text{impact of path loss}} + \underbrace{{\Gamma_{\text{1D}}}^{-1}\vartheta_{\mathcal{T}}}_{\text{impact of shadowing}} + \underbrace{{\Gamma_{\text{1D}}}^{-1}\Omega_{\mathcal{T}}}_{\text{impact of multipath}}, \end{aligned} \tag{3.8}$$

where e_1 denotes a unit vector in $\mathbb{R}^N$, with all entries zero except for the first one. We have, $Y_{\mathcal{T},\mathcal{F}} \sim \mathcal{N}\Big(\sqrt{N} K_{\text{dB}} e_1 - n_{\text{PL}} {\Gamma_{\text{1D}}}^{-1} D_{\mathcal{T}}, \alpha {\Gamma_{\text{1D}}}^{-1} R_{\text{norm},\mathcal{T}} {\Gamma_{\text{1D}}}^{-H} + \sigma^2 I_{N\times N}\Big)$. As can be seen, K_{dB} only affects the dc component of the frequency domain. Moreover, as n_{PL} increases, the absolute value of each frequency coefficient corresponding to the second part of the path loss term in Eq. 3.8 increases. This implies that as n_{PL} increases, on average, channel frequency response will become less compressible. As α and/or σ^2 increase, the variation of each of the Fourier components around its mean increases, as expected, implying lower chance of compressibility.

In order to understand the impact of β, we next characterize the variations of each component of ${\Gamma_{\text{1D}}}^{-1}\vartheta_{\mathcal{T}}$ as a function of β. For $1 \leq k \leq N$, we have the following for the

normalized variance of the kth element of ${\Gamma_{\text{1D}}}^{-1}\vartheta_{\mathcal{T}}$,

$$\begin{aligned}
&\Big[{\Gamma_{\text{1D}}}^{-1}R_{\text{norm},\mathcal{T}}{\Gamma_{\text{1D}}}^{-H}\Big]_{k,k}\\
&=\frac{1}{N}\sum_{1\le m,n\le N} e^{-j\frac{2\pi}{N}(k-1)(m-1)}e^{-\frac{|m-n|d}{\beta}}e^{j\frac{2\pi}{N}(k-1)(n-1)}=\frac{1}{N}\sum_{0\le m,n\le N-1} e^{-j\frac{2\pi}{N}(k-1)(m-n)}\rho^{|m-n|}\\
&=1+\frac{1}{N}\sum_{i=1}^{N-1}(N-i)\big[e^{-j\frac{2\pi}{N}(k-1)i}+e^{j\frac{2\pi}{N}(k-1)i}\big]\rho^i=1+\frac{1}{N}\sum_{i=1}^{N-1}(N-i)\big[(\frac{\rho}{\lambda_k})^i+(\rho\lambda_k)^i\big],
\end{aligned}$$

where $\rho=e^{-\frac{d}{\beta}}$ and $\lambda_k=e^{j\frac{2\pi}{N}(k-1)}$.

Lemma 1. *For $\varsigma\in(-1,1)$, we have,*

$$\sum_{i=1}^{N-1}(N-i)\varsigma^i=\frac{N\varsigma}{1-\varsigma}-\frac{\varsigma-\varsigma^{N+1}}{(1-\varsigma)^2}. \tag{3.9}$$

Proof. We have,

$$\begin{aligned}
\sum_{i=1}^{N-1}(N-i)\varsigma^i&=N\sum_{i=1}^{N-1}\varsigma^i-\sum_{i=1}^{N}i\varsigma^i=N\frac{\varsigma-\varsigma^N}{1-\varsigma}-\Big[\frac{\varsigma-\varsigma^N}{(1-\varsigma)^2}-\frac{(N-1)\varsigma^N}{1-\varsigma}\Big]\\
&=\varsigma\frac{N-1-N\varsigma+\varsigma^N}{(1-\varsigma)^2}.
\end{aligned} \tag{3.10}$$

□

Using Lemma 1 and the fact that $\rho\neq 1$ and $\lambda_k^N=(\frac{1}{\lambda_k})^N=1$, we have:

$$\begin{aligned}
&\Big[{\Gamma_{\text{1D}}}^{-1}R_{\text{norm},\mathcal{T}}{\Gamma_{\text{1D}}}^{-H}\Big]_{k,k}\\
&=1+\frac{1}{N}\sum_{i=1}^{N-1}(N-i)(\frac{\rho}{\lambda_k})^i+\frac{1}{N}\sum_{i=1}^{N-1}(N-i)(\rho\lambda_k)^i\\
&=1+\frac{(1-\frac{1-\rho^N}{N})\frac{\rho}{\lambda_k}-(\frac{\rho}{\lambda_k})^2}{(1-\frac{\rho}{\lambda_k})^2}+\frac{(1-\frac{1-\rho^N}{N})\rho\lambda_k-(\rho\lambda_k)^2}{(1-\rho\lambda_k)^2}\\
&=1+\rho(1-\frac{1-\rho^N}{N})\Big[\frac{\frac{1}{\lambda_k}}{(1-\frac{\rho}{\lambda_k})^2}+\frac{\lambda_k}{(1-\rho\lambda_k)^2}\Big]-\rho^2\Big[\frac{\frac{1}{\lambda_k^2}}{(1-\frac{\rho}{\lambda_k})^2}+\frac{\lambda_k^2}{(1-\rho\lambda_k)^2}\Big]\\
&=1+\rho(1-\frac{1-\rho^N}{N})\frac{(1+\rho^2)\Lambda_k-4\rho}{(1+\rho^2-\rho\Lambda_k)^2}-\rho^2\frac{\Lambda_k^2-2\rho\Lambda_k-2(1-\rho^2)}{(1+\rho^2-\rho\Lambda_k)^2},
\end{aligned}$$

where $\Lambda_k = \lambda_k + \frac{1}{\lambda_k} = 2\cos\big(\frac{2\pi}{N}(k-1)\big)$. Let $f_\rho : [-2,2] \to \mathbb{R}^+$ be defined as follows

$$f_\rho(\Lambda) \triangleq 1 + \rho(1 - \frac{1-\rho^N}{N})\frac{(1+\rho^2)\Lambda - 4\rho}{(1+\rho^2-\rho\Lambda)^2} - \rho^2\frac{\Lambda^2 - 2\rho\Lambda - 2(1-\rho^2)}{(1+\rho^2-\rho\Lambda)^2} \text{ for } \rho \in (0,1). \tag{3.11}$$

Then, the following theorem characterizes the variance of the Fourier transformation of the shadowing term ($\Gamma_{1\mathrm{D}}{}^{-1}\vartheta_{\mathcal{T}}$).

Theorem 3. *Let* $g_k \triangleq \Big[\Gamma_{1D}{}^{-1} R_{\mathcal{T}} \Gamma_{1D}{}^{-H}\Big]_{k,k} = \alpha f_\rho(\Lambda_k)$ *denote the variance of the* k*th Fourier transform component of* $\vartheta_{\mathcal{T}}$*, where* $\beta \in (0,\infty)$*. We have* $g_k \geq g_{k+1}$ *for* $1 \leq k \leq \lceil\frac{N-1}{2}\rceil$.

Proof. For $1 \leq k \leq \lceil\frac{N-1}{2}\rceil$, it can be easily confirmed that $\Lambda_k \geq \Lambda_{k+1}$. To prove the Theorem, it suffices to show that $f_\rho(\Lambda)$ is an increasing function of Λ. Taking the derivative with respect to Λ results in:

$$\begin{aligned}
\frac{d}{d\Lambda}f_\rho(\Lambda) &= \rho(1-\frac{1-\rho^N}{N})\frac{(1+\rho^2)(1+\rho^2-\rho\Lambda) + 2\rho\big(\Lambda(\rho^2+1-\rho\Lambda) + \rho(\Lambda^2-4)\big)}{(1+\rho^2-\rho\Lambda)^3} \\
&\quad - \rho^2\frac{2(\Lambda+\rho)(1+\rho^2-\rho\Lambda) + 2\rho(\Lambda^2-4)}{(1+\rho^2-\rho\Lambda)^3} \\
&= \frac{(1+\rho^2-\rho\Lambda)\big(\rho(1+\rho^2+2\rho\Lambda) - \rho^2\times 2(\Lambda+\rho)\big)}{(1+\rho^2-\rho\Lambda)^3} \\
&\quad - \frac{1-\rho^N}{N}\rho\frac{(1+\rho^2-\rho\Lambda)(1+\rho^2+2\rho\Lambda) + 2\rho^2(\Lambda^2-4)}{(1+\rho^2-\rho\Lambda)^3} \\
&= \frac{(\rho-\rho^3)(1+\rho^2-\rho\Lambda) + \frac{1-\rho^N}{N}\Big[(\rho+\rho^3)(1+\rho^2-\rho\Lambda) - 2\rho(1-\rho^2)^2\Big]}{(1+\rho^2-\rho\Lambda)^3}.
\end{aligned} \tag{3.12}$$

For $\Lambda \in [-2,2]$, we have $\Lambda^2 - 4 \leq 0$, resulting in $\rho^2 - \rho\Lambda + 1 \geq 1 - \rho^2 > 0$ for all $\rho \in (0,1)$. Therefore, it suffices to show that the numerator is positive. We have, $(\rho-\rho^3)(1+\rho^2-\rho\Lambda) + \frac{1-\rho^N}{N}\Big[(\rho+\rho^3)(1+\rho^2-\rho\Lambda) - 2\rho(1-\rho^2)^2\Big] \geq \rho(1-\rho^2)^2 + \frac{1-\rho^N}{N}\rho(1-\rho^2)(3\rho^2-1)$. For $\frac{1}{\sqrt{3}} \leq \rho < 1$, it can easily confirmed that the right side of the

above inequality is greater than zero. For $0 < \rho < \frac{1}{\sqrt{3}}$, we have,

$$\rho(1-\rho^2)^2 + \frac{1-\rho^N}{N}\rho(1-\rho^2)(3\rho^2-1) \geq \rho(1-\rho^2)^2 + \rho(1-\rho^2)(3\rho^2-1)$$
$$= 2\rho^3(1-\rho^2) > 0, \tag{3.13}$$

which proves the theorem. □

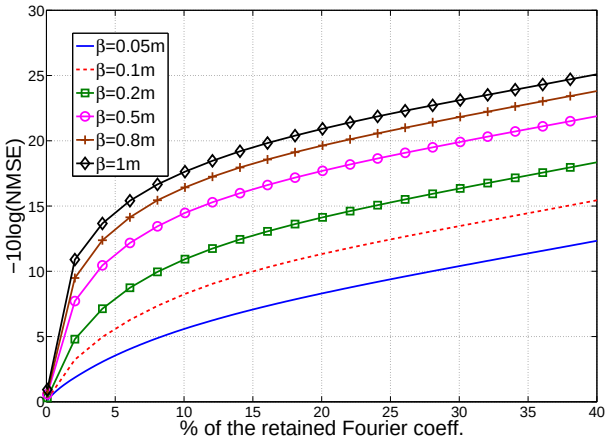

Figure 3.3: Characterizing the sparsity of the shadowing component, for different values of β, for $N = 1500$ and $d = 0.01$m. The y-axis shows the inverse of the Normalized Mean Square Error (in dB) between the shadowing component and its sparsified version, as a function of the % of the retained Fourier coefficients

For $\beta = 0$, it can be easily confirmed that $g_k = \alpha$ for $1 \leq k \leq N$. Moreover, for $\beta = \infty$, we have $g_1 = N\alpha$ and $g_k = 0$ for $2 \leq k \leq N$. In summary, the analysis of this part implies the following: as α and/or σ^2 increase, the probability of having a less compressible channel increases. As n_{PL} increases, channel becomes less sparse (in the frequency domain) on average. As for the impact of β, while our derivations are towards establishing that as β increases, the shadowing component becomes more compressible, more analysis is required to complete the proof. Thus, we complement this part with a simulation result. Fig. 3.3 characterizes the sparsity of the shadowing component, for different correlation distances (βs). The y-axis measures the sparsity, i.e. the inverse of

the Normalized Mean Square Error (in dB) between the shadowing component and its sparsified version, as a function of the percentage of the retained Fourier coefficients. The figure shows that as the correlation distance increases, the shadowing component becomes more compressible.

3.4 Channel prediction and the underlying tradeoffs

In this section, we compare the performance of the proposed sparsity-based framework of this chapter with the probabilistic framework of Chapter 2. As we shall see, each approach has its own strength that can result in a better reconstruction, depending on the scenario.

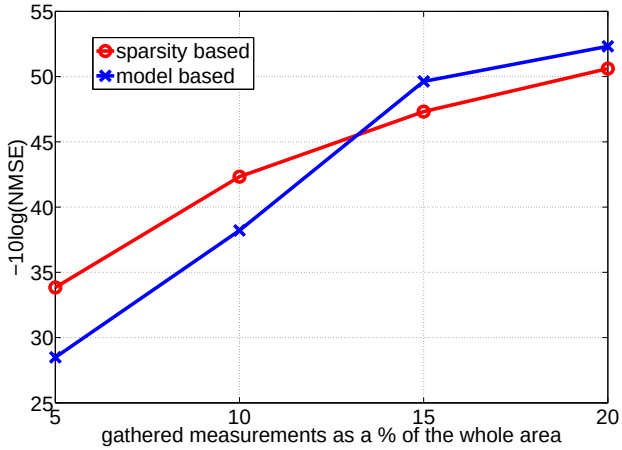

Figure 3.4: Comparison of the sparsity-based and model-based approaches in estimating the spatial variations of the channel of Fig. 3.1 (left).

Fig. 3.4 shows the performance of the sparsity-based approach of Section 3.2 and the model-based approach of Chapter 2 for the reconstruction of the channel in Fig. 3.1 (left), where the x-axis shows the percentage of the measurements gathered (as a % of the whole area of interest). In this case, the gathered measurements are randomly distributed over the workspace. For the model-based approach, our LS estimator of Section 2.1 is used for es-

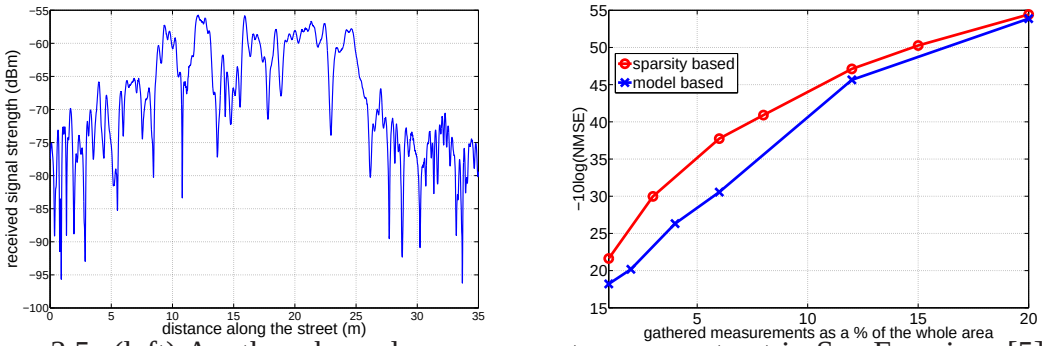

Figure 3.5: (left) Another channel measurement across a street in San Francisco [5] and (right) comparison of the sparsity-based and model-based approaches in estimating the spatial variations of this channel.

timating the underlying parameters. This is then followed by utilizing Eq. 2.8 for channel prediction. It can be seen that when the number of measurements is small (less than 13.5% in this figure), the sparsity-based approach outperforms the model-based one. This makes sense as the model-based approach needs to estimate the underlying parameters. For a very small number of measurements, the error in the estimation of these parameters can be high, resulting in a performance degradation in the overall estimation. As the number of measurements increases, the model-based approach then outperforms the sparsity-based one in this case.

The model-based approach is also sensitive to the accuracy of the underlying model. In order to see this, Fig. 3.5 (left) shows another channel measurement in San Francisco [5]. It can be seen that this channel can not be well characterized by only one path loss trend. As a result, we expect that the performance of our model-based approach degrade since it assumes only one path loss trend in the area of interest (See Eq. 2.2). Fig. 3.5 (right) shows the performance of channel reconstruction in this case. It can be seen that the sparsity-based approach outperforms the model-based one in this case, due to a modeling inaccuracy of the model-based approach.

The performance of the sparsity-based approach, on the other hand, depends heavily

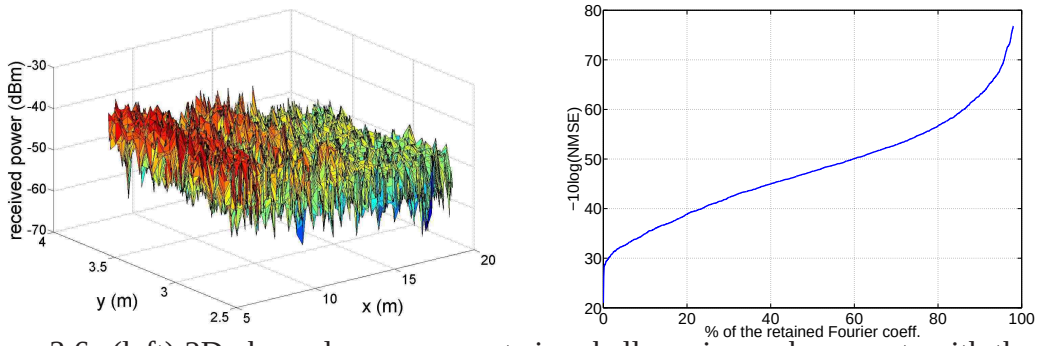

Figure 3.6: (left) 2D channel measurements in a hallway in our basement, with the base station at $(0, 0)$, and (right) characterizing the sparsity of this channel. See the PDF file for a colored version.

on the compressibility of the channel in the frequency domain. For both Fig. 3.4 and 3.5 (right), the channel is considerably compressible in the Fourier domain, which is evident from the good performance of the sparsity-based approach. There could, however, be cases where the spatial variations of the channel are not that compressible in the area of interest. Fig. 3.6 (left) shows a 2D channel, in a hallway in our basement. Fig. 3.6 (right) shows the sparsity of this channel in the same way that we measured the sparsity for Fig. 3.1. As can be seen, this channel is not that sparse. Fig. 3.7 shows the 2D reconstruction of this channel. It can be seen that the performance of both approaches degrades considerably, as compared to the previous channels. This area experiences considerable multipath fading and negligible shadowing, which reduces channel compressibility. Thus, the model-based approach outperforms the sparsity one unless almost half of the area is sampled. In summary, both approaches have their strengths and can be useful in estimating a wireless channel, based on a small number of measurements. However, depending on the scenario and the percentage of the available measurements, one of the approaches may outperform the other one. Thus, in the next section, we propose an integrated approach which takes advantage of both sparsity in the frequency domain and probabilistic characterization in the spatial domain.

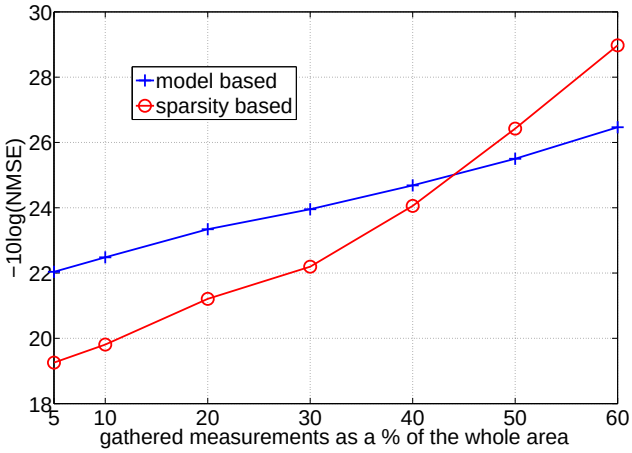

Figure 3.7: Performance of the model-based and sparsity-based approaches for the 2D channel of Fig. 3.6 (left).

3.5 An integrated sparsity and model-based framework for estimating channel spatial variations

So far, we discussed a sparsity-based and a model-based approach for estimating the spatial variations of a wireless channel. In Section 3.4, we showed the underlying tradeoffs between the two approaches and discussed the strengths and weaknesses of each. In this section, we propose a framework that integrates the strengths of both approaches, in order to achieve a more robust channel estimator with a better performance.

For the model-based approach, its performance is directly affected by the estimation of the underlying model parameters, as we saw in Section 3.4. Thus, in this section we also show how the sparsity of the channel in frequency domain can further be utilized to improve the estimation of the underlying model parameters.

3.5.1 An integrated model and sparsity-based estimator

Define $\mathcal{Q}^C$ as the set of all the positions of $\mathcal{P}$ where channel needs to be estimated: $\mathcal{Q}^C \triangleq \mathcal{P} \setminus \mathcal{Q} = \{q'_1, q'_2, \cdots, q'_{N-K}\}$. Let $Y_{\mathcal{Q}^C}$ denote the channel values at positions corresponding to $\mathcal{Q}^C$, which are not directly measured. Based on Eq. 2.8, we have the following for the probability distribution of $Y_{\mathcal{Q}^C}$, conditioned on all the gathered measurements and the underlying parameters: $f(Y_{\mathcal{Q}^C}|Y_{\mathcal{Q}}, \theta, \alpha, \beta, \sigma^2) \sim \mathcal{N}\big(\tilde{Y}_{\mathcal{Q}^C}, R_{\text{tot},\mathcal{Q}^C|Y_{\mathcal{Q}},\theta,\alpha,\beta,\sigma^2}\big)$ with

$$\begin{aligned}
\tilde{Y}_{\mathcal{Q}^C} &\triangleq \mathbb{E}\Big\{ Y_{\mathcal{Q}^C} \;\Big|\; Y_{\mathcal{Q}}, \theta, \alpha, \beta, \sigma^2 \Big\} = H_{\mathcal{Q}^C}\theta + \Sigma^T_{\mathcal{Q},\mathcal{Q}^C} R^{-1}_{\text{tot},\mathcal{Q}} \big(Y_{\mathcal{Q}} - H_{\mathcal{Q}}\theta\big) \\
R_{\text{tot},\mathcal{Q}^C|Y_{\mathcal{Q}},\theta,\alpha,\beta,\sigma^2} &\triangleq \mathbb{E}\Big\{ \big(Y_{\mathcal{Q}^C} - \tilde{Y}_{\mathcal{Q}^C}\big)\big(Y_{\mathcal{Q}^C} - \tilde{Y}_{\mathcal{Q}^C}\big)^T \;\Big|\; \theta, \alpha, \beta, \sigma^2 \Big\} \\
&= R_{\text{tot},\mathcal{Q}^C} - \Sigma^T_{\mathcal{Q},\mathcal{Q}^C} R^{-1}_{\text{tot},\mathcal{Q}} \Sigma_{\mathcal{Q},\mathcal{Q}^C}, \qquad (3.14)
\end{aligned}$$

where $\Sigma_{\mathcal{Q},\mathcal{Q}^C} = \text{COV}(Y_{\mathcal{Q}}, Y_{\mathcal{Q}^C}) \in \mathbb{R}^{K\times(N-K)}$ with entries $\Big[\Sigma_{\mathcal{Q},\mathcal{Q}^C}\Big]_{i,j} = \alpha e^{-\frac{\|q_i - q'_j\|}{\beta}}$ for $1 \le i \le K$ and $1 \le j \le N-K$. Let $\tilde{Y}_{\mathcal{Q}^C,\text{ML}}$ denote the ML estimation of $Y_{\mathcal{Q}^C}$. We then have :

$$\begin{aligned}
\tilde{Y}_{\mathcal{Q}^C,\text{ML}} &= \arg\max f(Y_{\mathcal{Q}^C}|Y_{\mathcal{Q}}, \theta, \alpha, \beta, \sigma^2) = \arg\max \ln\Big(f(Y_{\mathcal{Q}^C}|Y_{\mathcal{Q}}, \theta, \alpha, \beta, \sigma^2)\Big) \\
&= \arg\min \frac{1}{2}(Y_{\mathcal{Q}^C} - \tilde{Y}_{\mathcal{Q}^C})^T R^{-1}_{\text{tot},\mathcal{Q}^C|Y_{\mathcal{Q}},\theta,\alpha,\beta,\sigma^2}(Y_{\mathcal{Q}^C} - \tilde{Y}_{\mathcal{Q}^C}). \qquad (3.15)
\end{aligned}$$

Clearly, we have $\tilde{Y}_{\mathcal{Q}^C,\text{ML}} = \tilde{Y}_{\mathcal{Q}^C}$ if no information on channel sparsity is utilized. Eq. 3.15 is equivalent to the following optimization problem, as a function of the Fourier coefficients of the channel (X):

$$\begin{aligned}
&\tilde{X}_{\text{ML}} = \arg\min (\Phi_{\mathcal{Q}^C}\Gamma X - \tilde{Y}_{\mathcal{Q}^C})^T R^{-1}_{\text{tot},\mathcal{Q}^C|Y_{\mathcal{Q}},\theta,\alpha,\beta,\sigma^2}(\Phi_{\mathcal{Q}^C}\Gamma X - \tilde{Y}_{\mathcal{Q}^C}) \\
&\text{s.t. } Y_{\mathcal{Q}} = \Phi_{\mathcal{Q}}\Gamma X,
\end{aligned}$$

where $\Phi_{\mathcal{Q}}$ denotes the corresponding sampling matrix, as defined in Section 2.1 and Γ is the inverse Fourier matrix (see Eq. 3.6). $\Phi_{\mathcal{Q}^C}$ is then defined in a similar manner. By integrating this estimator with the sparsity-based one of Eq. 3.6, we have,

$$\begin{aligned}
&\tilde{X}_{\text{integrated}} = \arg\min \left(\tau\|X\|_1 + \left\| R^{-\frac{1}{2}}_{\text{tot},\mathcal{Q}^C|Y_{\mathcal{Q}},\theta,\alpha,\beta,\sigma^2}\big(\Phi_{\mathcal{Q}^C}\Gamma X - \tilde{Y}_{\mathcal{Q}^C}\big) \right\|_2^2 \right) \\
&\text{s.t. } Y_{\mathcal{Q}} = \Phi_{\mathcal{Q}}\Gamma X, \qquad (3.16)
\end{aligned}$$

where τ is a weighting coefficient. This estimator has optimality in both ℓ_1 and Maximum Likelihood senses. Eq. 3.16 can also be posed as the following unconstrained optimization problem to account for measurement noise:

$$\tilde{X}_{\text{integrated}} = \arg\min \left(\tau' \|X\|_1 + \left\| A_{\text{integrated}} X - b_{\text{integrated}} \right\|_2^2 \right) \tag{3.17}$$

where

$$A_{\text{integrated}} = \begin{bmatrix} R_{\text{tot},\mathcal{Q}^C|Y_\mathcal{Q},\theta,\alpha,\beta,\sigma^2}^{-\frac{1}{2}} \Phi_{\mathcal{Q}^C} \Gamma \\ \Phi_\mathcal{Q} \Gamma \end{bmatrix}, b_{\text{integrated}} = \begin{bmatrix} R_{\text{tot},\mathcal{Q}^C|Y_\mathcal{Q},\theta,\alpha,\beta,\sigma^2}^{-\frac{1}{2}} \tilde{Y}_{\mathcal{Q}^C} \\ Y_\mathcal{Q} \end{bmatrix},$$

and τ' is a weighting coefficient. τ' can be assigned to give more or less emphasis to the sparsity part (ℓ_1 optimization). Furthermore, in [115], the authors show that for a general $\ell_1 - \ell_2$ problem, i.e. $\arg\min \left(\tau' \|X\|_1 + \|AX - b\|_2^2 \right)$, we should have $\tau' < 2\|A^T b\|_\infty$. Otherwise, the unique solution will be the zero vector. This gives us a range for valid values of τ'. In some of the optimization literature and papers that have such an $\ell_1 - \ell_2$ problem, a pre-determined coefficient is found, by assuming some a priori information about the signal [116]. However, we do not assume any a priori information to optimize τ'. We can simply choose τ' to be a fraction of the maximum allowed value. Alternatively, an adaptive weight, based on the percentage of available channel samples and the estimated underlying parameters, can also be utilized.

3.5.2 Estimation of the underlying model parameters using channel sparsity in the frequency domain

In the previous part, we assumed that the underlying parameters of the probabilistic model are estimated, using the ML or LS approach of Section 2.1. As noted in Section 3.4, if enough channel samples are collected, the underlying parameters can be estimated with a good enough accuracy. However, at low enough sampling rates, the error in the estimation of the underlying parameters may not be negligible. Thus, in this part, we show how the

sparsity of the channel can further be used to improve the estimation of the underlying parameters.

Case of negligible multipath fading – $\sigma^2 \approx 0$: Depending on the environment, multipath fading can be negligible, as compared to the shadowing and path loss terms. We start by considering this case. Under this assumption, we can apply the ML estimator of channel parameters, as developed in Section 2.1 (Eq. 2.5), and write an expression for the ML estimation of the channel, only as a function of the correlation distance β, as follows:

$$\begin{aligned}
&\hat{Y}_{\mathcal{Q}^C,\text{ML},\sigma^2=0}(\beta) \triangleq \mathbb{E}\Big\{Y_{\mathcal{Q}^C}\Big|Y_{\mathcal{Q}}, \theta = \hat{\theta}_{\text{ML},\sigma^2=0}, \alpha = \hat{\alpha}_{\text{ML},\sigma^2=0}, \beta, \sigma^2 = 0\Big\} \\
&= H_{\mathcal{Q}^C}\hat{\theta}_{\text{ML},\sigma^2=0} + \Sigma^T_{\text{norm},\mathcal{Q},\mathcal{Q}^C}(\beta)R^{-1}_{\text{norm},\mathcal{Q}}(\beta)\Big(Y_{\mathcal{Q}} - H_{\mathcal{Q}}\hat{\theta}_{\text{ML},\sigma^2=0}\Big) \\
&= \Bigg(\Big[H_{\mathcal{Q}^C} - \Sigma^T_{\text{norm},\mathcal{Q},\mathcal{Q}^C}(\beta)R^{-1}_{\text{norm},\mathcal{Q}}(\beta)H_{\mathcal{Q}}\Big]\big(H^T_{\mathcal{Q}}R^{-1}_{\text{norm},\mathcal{Q}}(\beta)H_{\mathcal{Q}}\big)^{-1}H^T_{\mathcal{Q}}R^{-1}_{\text{norm},\mathcal{Q}}(\beta) \\
&+ \Sigma^T_{\text{norm},\mathcal{Q},\mathcal{Q}^C}(\beta)R^{-1}_{\text{norm},\mathcal{Q}}(\beta)\Bigg)Y_{\mathcal{Q}},
\end{aligned}$$

where $\Sigma_{\text{norm},\mathcal{Q},\mathcal{Q}^C} = \frac{1}{\alpha}\Sigma_{\mathcal{Q},\mathcal{Q}^C}$, with entries $\Big[\Sigma_{\text{norm},\mathcal{Q},\mathcal{Q}^C}\Big]_{i,j} = e^{-\frac{\|q_i - q'_j\|}{\beta}}$ for $1 \leq i \leq K$ and $1 \leq j \leq N - K$. By considering the channel over the whole field, including both measured and estimated points, we will have: $\hat{x}(\beta) = \Phi^T_{\mathcal{Q}}Y_{\mathcal{Q}} + \Phi^T_{\mathcal{Q}^C}\hat{Y}_{\mathcal{Q}^C,\text{ML},\sigma^2=0}(\beta)$, where $\hat{x}$ is a vector of channel signal strengths over the whole field. Next, we utilize the sparsity of x in the frequency domain in order to estimate β. We have,

$$\hat{\beta}_{\text{sparsity},\sigma^2=0} = \arg\min \|\Gamma^{-1}\hat{x}(\beta)\|_1 = \arg\min \big\|\Gamma^{-1}\Phi^T_{\mathcal{Q}}Y_{\mathcal{Q}} + \Gamma^{-1}\Phi^T_{\mathcal{Q}^C}\hat{Y}_{\mathcal{Q}^C,\text{ML},\sigma^2=0}(\beta)\big\|_1. \tag{3.18}$$

No closed-form expression, however, exists for the optimum β in this case. Once β is estimated from Eq. 3.18, α and θ can be immediately estimated as follows (see Section 2.1):

$$\begin{aligned}
\hat{\theta}_{\text{ML,sparsity},\sigma^2=0} &= \Big(H^T_{\mathcal{Q}}R^{-1}_{\text{norm},\mathcal{Q}}(\hat{\beta}_{\text{sparsity},\sigma^2=0})H_{\mathcal{Q}}\Big)^{-1}H^T_{\mathcal{Q}}R^{-1}_{\text{norm},\mathcal{Q}}(\hat{\beta}_{\text{sparsity},\sigma^2=0})Y_{\mathcal{Q}}, \\
\hat{\alpha}_{\text{ML,sparsity},\sigma^2=0} &= \frac{1}{K}\big(Y_{\mathcal{Q}} - H_{\mathcal{Q}}\hat{\theta}_{\text{ML,sparsity},\sigma^2=0}\big)^T R^{-1}_{\text{norm},\mathcal{Q}}(\hat{\beta}_{\text{sparsity},\sigma^2=0})\big(Y_{\mathcal{Q}} - H_{\mathcal{Q}}\hat{\theta}_{\text{ML,sparsity},\sigma^2=0}\big).
\end{aligned}$$

Once the underlying parameters are estimated, we can apply the integrated estimator of Eq. 3.17.

Case of non-negligible multipath: If $\sigma^2 \neq 0$, there is no closed-form expression that can express all the underlying parameters as a function of one of them, as was done in the previous part. Furthermore, the ML estimation of the underlying parameters was for the case of $\sigma^2 = 0$ in Section 2.1. Thus, in this case we consider the LS estimator of the underlying parameters. We can write the following for the estimated channel, as a function of α, β, σ^2 and the LS estimation of the path loss parameters:

$$\begin{aligned}\hat{Y}_{\mathcal{Q}^C,\text{LS}}(\alpha,\beta,\sigma^2) &\triangleq \mathbb{E}\Big\{Y_{\mathcal{Q}^C}\Big|Y_{\mathcal{Q}},\theta=\hat{\theta}_{\text{LS}},\alpha,\beta,\sigma^2\Big\}\\ &= \Bigg(\Big[H_{\mathcal{Q}^C}-\alpha\Sigma^T_{\text{norm},\mathcal{Q},\mathcal{Q}^C}(\beta)\big[\alpha R_{\text{norm},\mathcal{Q}}(\beta)+\sigma^2 I_{K\times K}\big]^{-1}H_{\mathcal{Q}}\Big]\big(H^T_{\mathcal{Q}}H_{\mathcal{Q}}\big)^{-1}H^T_{\mathcal{Q}}\\ &\quad+\alpha\Sigma^T_{\text{norm},\mathcal{Q},\mathcal{Q}^C}(\beta)\big[\alpha R_{\text{norm},\mathcal{Q}}(\beta)+\sigma^2 I_{K\times K}\big]^{-1}\Bigg)Y_{\mathcal{Q}}. \qquad (3.19)\end{aligned}$$

Similar to the previous part, this results in the following for the sparsity-based estimation of β, assuming that α and σ^2 are known:

$$\hat{\beta}_{\text{sparsity}}(\alpha,\sigma^2)=\arg\min\big\|\Gamma^{-1}\Phi^T_{\mathcal{Q}}Y_{\mathcal{Q}}+\Gamma^{-1}\Phi^T_{\mathcal{Q}^C}\hat{Y}_{\mathcal{Q}^C,\text{LS}}(\alpha,\beta,\sigma^2)\big\|_1. \qquad (3.20)$$

Assuming an estimated β, we can then estimate α and σ^2, using an LS estimator:

$$\Big[\hat{\alpha}_{\text{LS}}(\beta),\hat{\sigma}^2_{\text{LS}}(\beta)\Big]=\arg\min_{\alpha,\sigma^2>0}\sum_{l\in\mathcal{L}'_{\mathcal{Q}}}w'(l)\Big[\alpha e^{-l/\beta}+\sigma^2\delta(l)-\hat{r}_{\mathcal{Q}}(l)\Big]^2, \qquad (3.21)$$

where $\delta(.)$ denotes the dirac delta function and $\mathcal{L}'_{\mathcal{Q}}=\{l|0<\hat{r}_{\mathcal{Q}}(l)\leq\hat{\chi}_{\text{LS}|\theta=\hat{\theta}_{\text{LS}}}\}$. The weights $w'(l)$ can be chosen based on our assessment of the accuracy of the estimation of $\hat{r}_{\mathcal{Q}}(l)$. By iteratively solving the equations given by 3.20 and 3.21, we can estimate the underling channel parameters. After estimating the parameters, the integrated estimator of Eq. 3.17 can be applied to reconstruct the channel.

Next, we compare the performance of the integrated approach to that of the sparsity-based and model-based ones. To solve the convex problem of Eq. 3.17, we use SpaRSA

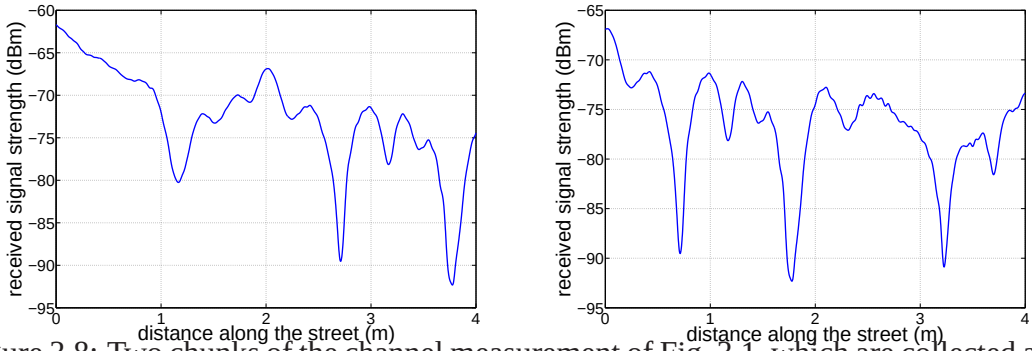

Figure 3.8: Two chunks of the channel measurement of Fig. 3.1, which are collected across a street in San Francisco [5] – (left) channel A and (right) channel B.

[117]. SpaRSA is an efficient iterative solver for minimizing an objective function that is a weighted sum of a quadratic (ℓ_2) error term and a sparsity regularizer (ℓ_1). In each iteration, it solves an optimization subproblem, involving a quadratic term with a diagonal Hessian, in combination with the original sparsity regularizer. The readers are referred to [117] for more details.

First, we show the performance of the integrated approach for two chunks of a channel across a street in San Francisco [5]. The first chunk, channel A, is as shown in Fig. 3.8 (left) whereas channel B is shown in Fig. 3.8 (right). Fig. 3.9 compares the sparsity level of these two chunks. As can be seen, channel B is more sparse than channel A. Fig. 3.10 shows the performance of the integrated approach for both channel A (left) and B (right) and compares them with that of the original sparsity and model-based approaches. As for estimating the underlying parameters, if the number of gathered measurements is high enough such that channel spatial correlation can be properly estimated, then the LS estimator of Section 2.1 is used to estimate all the underlying parameters. If the number of available channel measurements is very low, on the other hand, the proposed sparsity approach of this section is utilized to estimate the underlying parameters. For this channel, we assumed that multipath is negligible, when estimating the underlying parameters. Furthermore, an adaptive weight, inversely proportional to the number of available chan-

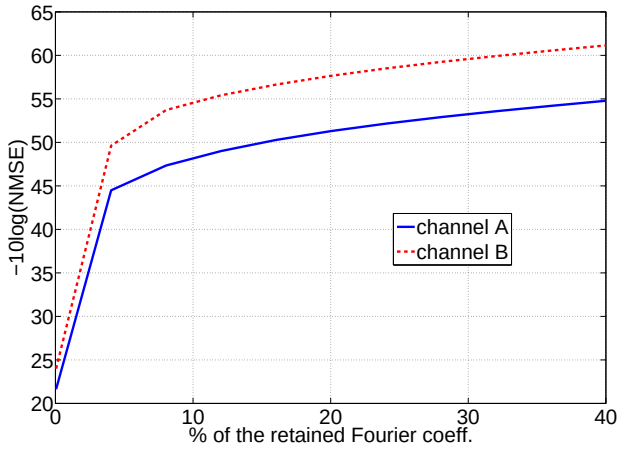

Figure 3.9: Characterizing the sparsity of channel A and B. The y-axis shows the inverse of the Normalized Mean Square Error (in dB) between the channel and its sparsified version, as a function of the % of the retained Fourier coefficients.

nel samples, is used for the ℓ_1 term. As can be seen, the integrated approach outperforms the original approaches considerably and can provide more than 10dB performance improvement depending on the % of available measurements. By comparing the left and right figures, it can furthermore be seen that the sparsity-based approach provides a better performance for channel B since it is sparser. Thus, channel A benefits more from the integrated approach.

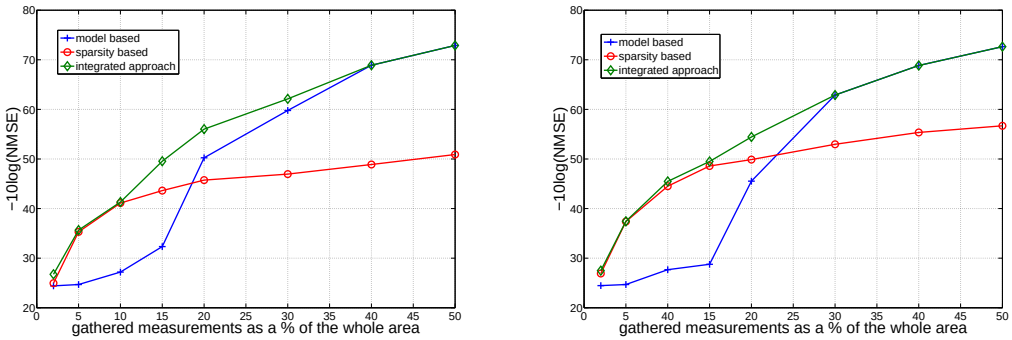

Figure 3.10: Performance of the proposed integrated sparsity and model-based approach for (left) channel A (Fig. 3.8 (left)) and (right) channel B (Fig. 3.8 (right)).

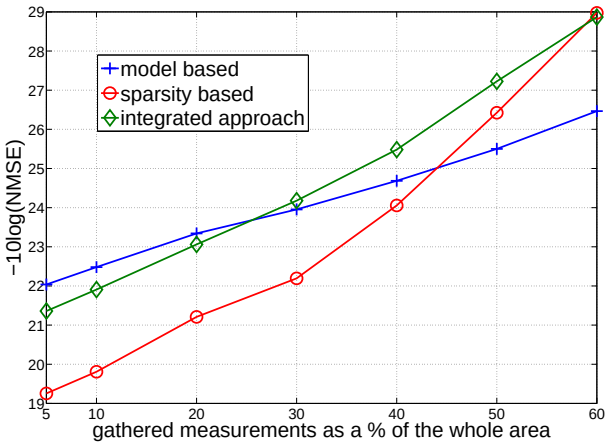

Figure 3.11: Performance of the proposed integrated sparsity and model-based approach for the 2D channel of Fig. 3.6 (left).

Next consider the 2D channel of Fig. 3.6 (left). Fig. 3.11 shows the performance of our integrated approach for this channel. This is an indoor channel that experiences considerable multipath. Thus, we can not assume that $\sigma^2 \approx 0$. For low sampling rates, where the underlying parameters can not be accurately estimated by the LS approach of Section 2.1, we use our proposed integrated approach, by solving Eq. 3.20 and Eq. 3.21 iteratively. This is then followed by applying our proposed estimator of Eq. 3.17. As can be seen, the integrated approach can improve the performance by properly combining the benefits of both approaches. For this indoor result, we used a fixed weight (τ'). More performance improvements can be achieved by properly adapting the weight. In [24, 28], approaches similar to our model-based approach are presented. Based on the these results, we expect that our proposed integrated approach outperforms these approaches as well, especially for low sampling rates.

3.6 Summary

In this chapter, we showed how to utilize the sparsity of the channel in the frequency domain in order to estimate the spatial variations of a wireless channel, based on a small number of measurements. We also characterized the underlying tradeoffs between the probabilistic approach of Chapter 2 and the sparsity-based estimator of this chapter. We showed that the probabilistic framework performs well, when the channel underlying parameters are estimated correctly. However, if the channel underlying parameters can not be estimated correctly, for instance due to the very small number of the available measurements or a modeling mismatch, the sparsity-based approach will then outperforms the probabilistic framework. Motivated by our analysis of the underlying tradeoffs between these two approaches, we then proposed an integrated channel prediction framework. In this framework, we showed how to utilize both channel sparsity in the frequency domain and probabilistic characterization in the spatial domain, in order to build a channel estimator that can keep the benefits of both approaches. We furthermore validate the applicability of this framework using both outdoor and indoor channel measurements.

Chapter 4

Binary consensus over fading channels

In Chapters 2 and 3, we considered connectivity-related issues by proposing a framework for understanding channel predictability. Our proposed framework will allow a network of cooperating nodes to more intelligently and efficiently maintain connectivity.

In this chapter, we then focus on sensing-related issues where each agent has a possibly erroneous perception of a measured parameter in its environment. Our goal is then to mitigate sensing errors through group cooperation and consensus. More specifically, we consider the case where each agent starts with a binary decision of a parameter of interest it has measured. We then study the *binary consensus* problem over fading channels, where the goal of every node is to reach the majority of the initial votes of all the agents (in order to increase immunity to local detection errors). For instance, in a cooperative fire detection scenario, each node has an initial opinion as to if there is a fire or not. However, as a network they may act based only on the majority vote. Therefore, the goal of the network is for each node to reach consensus over the majority of initial votes. Another application of *binary consensus* is in the cooperative spectrum sensing in cognitive radio networks. In this scenario, the secondary users communicate with each other in order to reach consensus on busy or idle status of the primary user, which is a binary value.

In this chapter, we first consider the binary consensus problem over a fixed and fully-connected network topology, where each link experiences fading and receiver noise. To improve the performance and robustness of the network cooperation, we propose novel consensus-seeking protocols that utilize information of link qualities and noise variances. We show that the proposed approach can improve consensus performance drastically. We then model the network state as a Markov chain and characterize the transient behavior of the network probabilistically. In particular, we show that, in the presence of channel uncertainties, the network state is asymptotically memoryless, depending on the utilized decision-making strategy. This is undesirable since the group agreement is not related to the initial state of the system and is merely a function of channel errors. While channel uncertainty can result in undesirable asymptotic behavior, depending on the utilized decision-making strategy, we show that the network can still be in consensus for a long period of time (enough for practical purposes) with high probability. In order to characterize the transient behavior, we derive a tight approximation for the second largest eigenvalue of the average of the underlying linear dynamical system. The derived expressions show how channel uncertainty and network topology affect binary consensus and shed light on the underlying tradeoffs.

In realistic scenarios, however, some links may not exist due to poor quality. Furthermore, the underlying communication topology could be time-varying. Therefore, we extend the binary consensus scenario to the not fully-connected and rapidly-changing network topologies. We start by considering the case where if a link exists, it is perfect in order to solely focus on the impact of not fully-connected graphs. We then consider binary consensus over not fully-connected rapidly-changing topologies with fading channels. Furthermore, we consider two different decision-making strategies, in terms of using the available transmissions: fusion and diversity. In the first approach, the given resources are used to increase the flow of information in the network whereas the second strategy aims to increase robustness to link errors by channel coding. We then characterize the underlying tradeoffs between these two approaches in terms of the speed of convergence and

asymptotic behavior for binary consensus over rapidly-changing network topology. For instance, we show that by fusing the received information, the fusion strategy propagates the information faster in the network and results in a better transient behavior. However, it can lack asymptotic accurate consensus depending on the form of the local decision function. On the other hand, the diversity strategy can provide a better asymptotic performance. The main contributions of this chapter are summarized as follows:

1. Proposing a novel consensus-seeking protocol that utilizes information of link qualities and noise variances to improve the performance and robustness of network cooperation.

2. Mathematical characterization of the impact of fading, noise, network connectivity and time-varying topology on consensus performance, which becomes challenging due to all the introduced uncertainties.

3. Characterizing the underlying tradeoffs between our proposed diversity-based and fusion-based approaches in terms of speed of convergence and asymptotic behavior.

The rest of this chapter is organized as follows: Section 4.1 introduces the problem and describes our system model. In Section 4.2, we develop the foundations of binary consensus over a fully connected time-invariant network topology with fading channels, in order to focus solely on the impact of fading. More specifically, we propose a novel consensus-seeking protocol that utilizes information of link qualities and noise variances. We then mathematically analyze the performance of the proposed protocol in terms of the transient and asymptotic behavior. In Section 4.3, we extend our analysis to time-varying not fully connected network topologies. In Section 4.3.1, we consider binary consensus over time-varying not fully connected network topologies with ideal links in order to build an understanding of the impact of rapidly-changing topologies. We mathematically analyze both fusion and diversity decision-making strategies in terms of the transient and asymp-

totic behavior. Then, we extend that analysis to the case of fading channels in Section 4.3.2. A summary of the results of the chapter is provided in Section 4.4.

4.1 Problem formulation

Consider a cooperative network of M nodes[1] that are trying to reach consensus over the occurrence of an event. Each agent has its own initial decision, based on its one-time sensing. The goal of the network is for each node to reach a decision that is equal to the majority of the initial votes. For instance, in a cooperative fire detection scenario, each node has an initial opinion as to if there is a fire or not. However, as a network, they may act only based on the majority vote. Therefore, it is desirable that every node reaches the majority of the initial votes without a group leader. As it may happen in realistic scenarios, the nodes may not have any information on the sensing quality of themselves or others. Therefore, the main goal is that each node reaches the majority of the initial votes.

In order to achieve this, each node will transmit its current decision to other nodes. The transmissions occur over fading channels and are furthermore corrupted by the receiver noise. Each node will then revise its current vote based on the received information. This process will go on for a while. We say that *accurate consensus* is achieved if each agent reaches the majority of the initial votes. The network can also be in a state of consensus while the information of the initial state is lost. We refer to this state as memoryless consensus. More specifically, if the probability of consensus (or equivalently the probability of being in all the states of the system) is independent of the initial condition, we say that any consensus, if achieved, is memoryless. This is undesirable since the group agreement is not related to the initial state of the system and is merely a function of channel errors.

[1]We also use the term “agent” to refer to each node.

Let $b_i(0) \in \{0, 1\}$ represent the initial vote of the ith node, at time step $k = 0$, where $b_i(k) = 1$ indicates that the ith agent decides that the event occurred whereas $b_i(k) = 0$ denotes otherwise. Each agent will send its binary vote (only one bit of information) to the rest over fading channels. Let $r_{j,i}(k)$ represent the fading coefficient of the link from node j to node i. The receiver then learns $r_{j,i}$s and uses it in the detection process. Let $n_{j,i}(k)$ represent the receiver noise at the kth time step in the transmission from the jth node to the ith one. $n_{j,i}(k)$ is zero-mean Gaussian with the variance of $\sigma^2_{j,i}$. We take the receiver noise of the receptions of different nodes to be uncorrelated. Let $b_{j,i}(k)$ represent the reception of the ith node from the transmission of the jth one at the kth time step. We have the following if there exists a link from the jth node to the ith one:[2]

$$b_{j,i}(k) = r_{j,i}(k)b_j(k) + n_{j,i}(k) \text{ for } 1 \leq i, j \leq M, \tag{4.1}$$

where $n_{i,i}(k) = 0$ and $r_{i,i}(k) = 1$. We assume that each receiver can learn the fading coefficient of each of its receptions and undo its effect. Furthermore, we consider the case where the network experiences rapidly-changing but stationary fading channels. This means that the nodes move fast enough such that $r_{j,i}$s change and become uncorrelated from one time step to the next. However, their movements are limited to a given area such that fading channels can be considered stationary.

Let $\mathcal{N}_i(k)$ represent the set of indices of those agents that can communicate to the ith one (excluding itself) at time step k. We have

$$\mathcal{N}_i(k) = \{o^i_1(k), \cdots, o^i_{|\mathcal{N}_i(k)|}(k)\}, \tag{4.2}$$

for $o^i_j(k) \in \{1, 2, \ldots, M\} \setminus \{i\}$ where $o^i_j(k) \neq o^i_{j'}(k)$ for $j \neq j'$, and $|\mathcal{N}_i(k)|$ represents the size of $\mathcal{N}_i(k)$: $0 \leq |\mathcal{N}_i(k)| < M$. Each agent will then update its vote based on its past vote and the received information as follows:

$$b_i(k+1) = \mathcal{F}\left(b_i(k), b_{o^i_1(k),i}(k), b_{o^i_2(k),i}(k), \ldots, b_{o^i_{|\mathcal{N}_i(k)|}(k),i}(k)\right), \tag{4.3}$$

[2]Note that, without loss of generality, we assumed that the modulation is on-off keying.

for $o_j^i(k) \in \mathcal{N}_i(k)$ and $1 \leq j \leq |\mathcal{N}_i(k)|$ where $\mathcal{F}(.)$ represents a decision-making function. We show how to optimize this function in the next section. Let $D(k) = [b_1(k)\, b_2(k) \cdots b_M(k)]$ represent the state of the system at the kth time step and $S(k) = \sum_{i=1}^{M} b_i(k)$ denote the corresponding sum of all the votes. Let $\Xi(k)$ represent a $2^M \times 1$ vector that contains the probabilities of being in different possible states. We have [89]:

$$\Xi(k) = \begin{bmatrix} \text{prob}\big[D(k) = [00\cdots0]\big] \rightarrow S(k) = 0 \\ \left.\begin{array}{c} \text{prob}\big[D(k) = [00\cdots1]\big] \\ \vdots \\ \text{prob}\big[D(k) = [10\cdots0]\big] \end{array}\right\} \rightarrow S(k) = 1 \\ \vdots \\ \text{prob}\big[D(k) = [11\cdots1]\big] \rightarrow S(k) = M \end{bmatrix}. \tag{4.4}$$

Without loss of generality, possible states are ordered such that $S(k)$ increases. Within each group where $S(k)$ is constant, the states are ordered increasingly. Then $\Xi_n(k) = \text{prob}[D(k) = \phi^{(n)}]$ for $0 \leq n \leq 2^M - 1$, where $\phi^{(n)} = [b_1^n\, b_2^n \cdots b_M^n]$ is the nth state chosen from the ordered list. We have

$$\Xi(k+1) = P^T(k)\Xi(k), \tag{4.5}$$

where $P(k) = [P_{m,n}(k)]$ represents a $2^M \times 2^M$ state transition matrix at time k with $P_{m,n}(k) = \text{prob}[D(k+1) = \phi^{(n)} | D(k) = \phi^{(m)}]$ for $0 \leq m, n \leq 2^M - 1$. Let $\nu_{\phi^{(m)}}^{(i)}(k)$ represent the probability that the ith agent votes one at $(k+1)$th time step, given that the current state is $\phi^{(m)}$:

$$\nu_{\phi^{(m)}}^{(i)}(k) \triangleq \text{prob}\big[b_i(k+1) = 1 | D(k) = \phi^{(m)}\big]. \tag{4.6}$$

Then, the probability of going from state m to state n, at time k, will be as follows:

$$\begin{aligned} P_{m,n}(k) &= \text{prob}\left[D(k+1) = \phi^{(n)} | D(k) = \phi^{(m)}\right] \\ &= \prod_{i=1}^{M} \text{prob}[b_i(k+1) = b_i^n | D(k) = \phi^{(m)}] \\ &= \prod_{i=1}^{M} [b_i^n \nu_{\phi^{(m)}}^{(i)}(k) + (1 - b_i^n)(1 - \nu_{\phi^{(m)}}^{(i)}(k))]. \end{aligned}$$

4.2 Binary consensus over a fixed fully-connected network topology with fading channels

In this section, we develop a mathematical framework for binary consensus over a fixed and fully-connected network topology, where each link experiences fading and receiver noise, as denoted by Eq. 4.1. A network is called fully connected, if there exists a direct link between any two nodes. Our goal, in this section, is to solely focus on the impact of fading. The analysis of this section will then serve as a base for the derivations of the subsequent sections, where we consider binary consensus over not fully-connected time-varying topologies.

4.2.1 Design of the local decision-making function – A Best Affine Estimation (BAE) approach

As denoted in Eq. 4.3, each node updates its vote using function $\mathcal{F}$. In general, this function should be designed based on the optimum detector of the majority vote. However, the computational complexity of this receiver is exponential and thus prohibitive. Another possibility is to design an estimator of $S(k)$, the sum of all the votes. Each node can then easily translate its estimation of $S(k)$ to a detection of majority vote as follows. If

$S(k)$ is estimated to be above $\frac{M}{2}$, then the majority vote is one. Otherwise, it is zero. Therefore, we design the local decision-making function based on the estimation of $S(k)$. We furthermore focus on the Best Affine Estimation (BAE) of $S(k)$. There are two reasons for considering the Best Affine Estimator. First, it was shown that such a receiver has a performance considerably close to the one that is based on the optimum nonlinear estimation of $S(k)$ [87]. Second, we can mathematically characterize the transient behavior of this receiver. In this section, we mathematically characterize the BAE-based local decision-making function.

For a fully-connected graph, we have $|\mathcal{N}_i| = M-1$ for all $1 \leq i \leq M$. Let o_j^i for $1 \leq j \leq M-1$ be as defined in Eq. 4.2 for the i^{th} node, where we dropped index k since $\mathcal{N}_i(k)$ is a time-invariant set in this case. Define the following variables: $G_i(k) = [b_{o_1^i}(k), \cdots, b_{o_{M-1}^i}(k)]^T$, $B_{i,r}(k) = [b_{o_1^i,i}(k), \cdots, b_{o_{M-1}^i,i}(k)]^T$, $R_i(k) = \text{diag}\big(r_{o_1^i,i}(k), \cdots, r_{o_{M-1}^i,i}(k)\big)$ $\omega_i(k) = [n_{o_1^i,i}(k), \cdots, n_{o_{M-1}^i,i}(k)]^T$, where diag($z$) is a diagonal matrix with the elements of vector z on its main diagonal. Then we have the following, considering all the receptions of the ith node:

$$B_{i,r}(k) = R_i(k)G_i(k) + \omega_i(k). \tag{4.7}$$

Let $\psi_i(k)$ represent the sum of the votes of all the nodes except for the ith node: $\psi_i(k) = \vec{1}^T G_i(k) = S(k) - b_i(k)$, where $\vec{1}$ denotes a column vector with all elements of one. Then the ith node estimates $\psi_i(k)$ by using the best affine unbiased function of the received information: $\hat{\psi}_i(k) = \alpha_i^T(k)B_{i,r}(k) + \beta_i(k)$, where $\hat{\psi}_i(k)$ is the ith node's estimate of $\psi_i(k)$. To ensure an unbiased estimator, we should have $E[\hat{\psi}_i(k)] = E[\psi_i(k)] \Rightarrow \alpha_i^T(k)E[B_{i,r}(k)] + \beta_i(k) = \vec{1}^T E[G_i(k)]$. We use $E[z]$, $E\{z\}$ and $\bar{z}$ to denote the average of random variable z. Then we have the following optimization problem,

$$\begin{aligned} \alpha_i(k) &= \text{arg } \min \; E[\big(\hat{\psi}_i(k) - \psi_i(k)\big)^2], \\ \text{subject to } \beta_i(k) &= \big(\vec{1} - R_i(k)\alpha_i(k)\big)^T \Lambda_i(k), \end{aligned} \tag{4.8}$$

where $\Lambda_i(k) \triangleq E[G_i(k)] = \text{prob}\big[G_i(k) = \vec{1}\big]$. $\Lambda_i(k)$ characterizes the voting patterns of

different nodes. We then have,

$$
\begin{aligned}
&E[\big(\hat{\psi}_i(k)-\psi_i(k)\big)^2]\\
&=E\Big[\Big(\alpha_i^T(k)\big(R_i(k)G_i(k)+\omega_i(k)\big)+\big(\vec{1}-R_i(k)\alpha_i(k)\big)^T\Lambda_i(k)-\vec{1}^TG_i(k)\Big)^2\Big]\\
&=\alpha_i^T(k)\Big(R_i(k)C_i(k)R_i(k)+\Omega_i(k)\Big)\alpha_i(k)-2\alpha_i^T(k)R_i(k)C_i(k)\vec{1}+\vec{1}^TC_i(k)\vec{1}, \quad (4.9)
\end{aligned}
$$

where $C_i(k)=E\big[(G_i(k)-\Lambda_i(k))(G_i(k)-\Lambda_i(k))^T\big]$ is the covariance matrix of $G_i(k)$ and $\Omega_i(k)=E[\omega_i(k)\omega_i^T(k)]$. By noting that Eq. 4.9 is a convex function of $\alpha_i(k)$, we have

$$
\begin{aligned}
&\alpha_i(k)=\Big(R_i(k)C_i(k)R_i(k)+\Omega_i(k)\Big)^{-1}R_i(k)C_i(k)\vec{1} \text{ and}\\
&\beta_i(k)=\big(\vec{1}-R_i(k)\alpha_i(k)\big)^T\Lambda_i(k). \quad (4.10)
\end{aligned}
$$

Then, the ith node can detect the majority of the votes, using the estimation of $S(k)$, as follows:

$$
b_i(k+1)=\text{Dec}\Big(\frac{1}{M}\big[b_i(k)+\hat{\psi}_i(k)\big]\Big), \quad (4.11)
$$

where $\text{Dec}(z)=\begin{cases}1 & z\geq 0.5\\ 0 & z<0.5\end{cases}$. As can be seen, to update its vote, the ith node needs to calculate $C_i(k)$, the covariance matrix of $G_i(k)$. This requires the ith node to calculate the correlation between the votes of any two nodes in the network, which could be computationally prohibitive. Therefore, the ith node assumes that the votes of different nodes are uncorrelated when updating its decision. This means that while different votes can be correlated, the ith node considers $C_i(k)$ to be diagonal. This simplification then facilitates the mathematical characterizations of the rest of this chapter. If the ith node assumes that votes of different nodes are uncorrelated, i.e. $E[b_j(k)b_l(k)]=E[b_j(k)]E[b_l(k)]$ for $j\neq l$,

then we have

$$\alpha_{j,i}(k) = \frac{r_{j,i}(k)}{r_{j,i}^2(k) + \frac{\sigma_{j,i}^2}{q_j(k)(1-q_j(k))}} = \frac{1/r_{j,i}(k)}{1 + \frac{1}{q_j(k)(1-q_j(k))\text{CNR}_{j,i}(k)}} \text{ and}$$
$$\beta_i(k) = \sum_{j=1, j\neq i}^{M} \beta_{j,i}(k), \tag{4.12}$$

where $\text{CNR}_{j,i}(k) = \frac{r_{j,i}^2(k)}{\sigma_{j,i}^2}$, $\beta_{j,i}(k) = \big(1 - \alpha_{j,i}(k) r_{j,i}(k)\big) q_j(k)$ and $q_j(k) = E[b_j(k)]$. Therefore, the ith node will update its decision as follows:

$$b_i(k+1) = \text{Dec}\Big(\frac{1}{M}\big[b_i(k) + \hat{\psi}_i(k)\big]\Big) = \text{Dec}\left(\frac{1}{M}\Big[b_i(k) + \sum_{\substack{j=1 \\ j\neq i}}^{M} \Big(\alpha_{j,i}(k) b_{j,i}(k) + \beta_{j,i}(k)\Big)\Big]\right). \tag{4.13}$$

Note that for non-zero noise variances, the probability of the argument of the Dec(.) function of Eq. 4.13 being 0.5 is zero. Therefore, for the sake of the analysis of this section, the value of $\text{Dec}(z = 0.5)$ is chosen one, without loss of generality, as it does not impact network behavior.

It can be seen that Eq. 4.13 assumes that the knowledge of $q_j(k)$ is available at the receiver. If the ith node does not have an estimate of $q_j(k)$, it will assume that $q_j(k) = \frac{1}{2}$. We refer to this case as ***basic BAE***. Then, ***learning BAE*** refers to the case where $q_j(k)$ is statistically learned in the receiver. In order to do so, node i will pass $b_{j,i}(k)$ through a hard decision function to estimate the number of times that b_j becomes one in a given time interval. We then mathematically characterize the asymptotic and transient behavior of the basic BAE case. The mathematical characterization of the performance of learning BAE requires considering the error in the estimation of $q_j(k)$s, which is a challenging problem. Therefore, in Section 4.2.5 we show the performance of learning BAE through simulation.

4.2.2 Asymptotic behavior of basic BAE

For the basic BAE case, Eq. 4.13 can be simplified to the following:

$$b_i(k+1) = \text{Dec}\left(\frac{1}{M}\Big[b_i(k) + \sum_{\substack{j=1\\ j\neq i}}^{M} \widetilde{\alpha}_{j,i}(k) b_j(k) + \widetilde{\beta}_i(k) + \widetilde{n}_i(k)\Big]\right), \tag{4.14}$$

where $\widetilde{\alpha}_{j,i}(k) = \frac{1}{1+\frac{4}{\text{CNR}_{j,i}(k)}}$ and $\widetilde{\beta}_i(k) = \sum_{j=1,j\neq i}^{M} \frac{1-\widetilde{\alpha}_{j,i}(k)}{2}$. Furthermore, $\widetilde{n}_i(k)$ is a zero-mean Gaussian random variable with the variance of $\widetilde{\sigma}_i^2(k) = \frac{1}{M^2}\sum_{j\neq i} \frac{\frac{1}{\overline{CNR_{j,i}(k)}}}{(1+\frac{4}{\overline{CNR_{j,i}(k)}})^2}$. Given the current state of $\phi^{(m)}$, the probability of node i voting one will be as follows:

$$\nu_{\phi^{(m)}}^{(i)}(k) = \text{prob}[b_i(k+1) = 1 | D(k) = \phi^{(m)}] = Q\left(\frac{0.5 - \frac{1}{M}[b_i^m + \sum_{j=1,j\neq i}^{M} \widetilde{\alpha}_{j,i}(k) b_j^m + \widetilde{\beta}_i(k)]}{\widetilde{\sigma}_i(k)}\right), \tag{4.15}$$

where $\phi^{(m)} = [b_1^m\ b_2^m \cdots b_M^m]$ and $Q(z) = \int_z^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$. Since $\nu_{\phi^{(m)}}^{(i)}(k)$, and subsequently $P(k)$ of Eq. 4.5, are functions of $\text{CNR}_{j,i}(k)$s, then matrix $P(k)$ is time-varying. For such cases, the average dynamical system should instead be considered where the average is taken over the fading coefficients to get $\overline{P}$. Assuming that the fading coefficients are stationary, we have $\overline{P} = [\overline{P}_{m,n}]$ with

$$\overline{P}_{m,n} = \prod_{i=1}^{M} \Big[b_i^n \overline{\nu_{\phi^{(m)}}^{i}} + (1 - b_i^n)(1 - \overline{\nu_{\phi^{(m)}}^{i}})\Big] \tag{4.16}$$

and $E\big[\Xi(k+1)\big] = \overline{P}^T E\big[\Xi(k)\big]$. Let $\varrho = 2^M$ denote the number of possible states of the system and $\lambda_{0,\overline{P}}, \lambda_{1,\overline{P}} \ldots, \lambda_{\varrho-1,\overline{P}}$ represent ordered eigenvalues of $\overline{P}$, where $|\lambda_{0,\overline{P}}| \geq |\lambda_{1,\overline{P}}| \geq \ldots \geq |\lambda_{\varrho-1,\overline{P}}|$. The following can be easily confirmed for the case that $\forall i\ \widetilde{\sigma}_i(k) \neq 0$ [79]:

Property 1: Matrix $\overline{P}$ is stochastic and positive (element-wise). A stochastic matrix is a matrix in which the sum of each row is one. It is thus clear that matrix $P(k)$, for

any k, and therefore $\overline{P}$ is stochastic. For the positive part, note that if $\forall i\ \widetilde{\sigma}_i(k) \neq 0$, then $0 < \nu^{(i)}_{\phi^{(m)}}(k) < 1$ from Eq. 4.15 and therefore $P_{m,n}(k) > 0$ and $\overline{P} > 0$. This means that if there is any noise in any of the receptions of each node, then there is a non-zero probability of going from any state to any other.

Property 2: $\lambda_{0,\overline{P}} = 1$, $|\lambda_{i,\overline{P}}| < 1$ for $1 \leq i \leq \varrho - 1$,

Property 3: $\lim_{k\to\infty}(\overline{P}^T)^k \to xy^T$ where $x = \overline{P}^T x$, $y = \overline{P}y$, and $x^T y = 1$,

where 2 and 3 can be easily deduced from Property 1 using Perron and Gershgorin disk theorems [99]. Then from Property 3, we know that the average dynamic of the network reaches a steady state asymptotically. Furthermore, we will have $\lim_{k\to\infty} \overline{\Xi(k)} = xy^T\Xi(0)$ where x and y are as defined in Property 3. Consider $xy^T\Xi(0)$, the asymptotic value of vector $\overline{\Xi(k)}$. $\Xi(0)$ has exactly one element equal to one and the rest zero. Since $\overline{P}$ is stochastic from Property 1, vector y is a vector whose elements are all the same. Then, $y^T\Xi(0)$ loses the information of the initial state. Therefore, the asymptotic value will be independent of the initial state and is proportional to x, the right eigenvector of $\overline{P}^T$. It can be seen that the network loses its memory of the initial state asymptotically due to the impact of link errors, which is undesirable (see [79] for more details). It should be noted that for any amount of non-zero link noise, the asymptotic behavior will be memoryless. The network, however, can still be in consensus for a long period of time (enough for practical purposes) with high probability, which necessitates characterizing the transient behavior. Since the asymptotic behavior is memoryless, it is desirable that the network gets there with a slower rate. In general $\lambda_{1,\overline{P}}, \ldots, \lambda_{\varrho-1,\overline{P}}$ determine the transient behavior of the network. Among these eigenvalues, the second largest one ($\lambda_{1,\overline{P}}$) typically has the most impact on the transient behavior (see [118] for more details). The closer the second eigenvalue is to the unit circle, the longer the network is in consensus. In the limit, it can be easily confirmed that if $\forall i, j \neq i$, $\sigma_{j,i} = 0$, we will have $\lambda_{1,\overline{P}} = 1$ and $\lambda_{i,\overline{P}} = 0$ for $2 \leq i \leq \varrho - 1$ [79]. In [79], we considered binary consensus over AWGN channels and derived an approximated expression for the second largest eigenvalue. We next extend that analysis to the case of fast fading channels.

4.2.3 Transient behavior and the second largest eigenvalue for basic BAE

In this part, we characterize the transient behavior of the basic BAE case in fast-fading environments. In general, finding an exact expression for the second largest eigenvalue of $\overline{P}$ is considerably challenging. Instead, we derive a tight approximation for it, based on the linearization of the $Q(.)$ function.

Assumption 1: For a small arbitrary z, the linearization of the $Q(.)$ function around the origin results in $Q(z) \approx Q(0) - \frac{z}{\sqrt{2\pi}}$. We next prove the following general theorem, which will be used throughout this chapter.

Theorem 1. *Let $\overline{\mathcal{P}} = [\overline{\mathcal{P}}_{m,n}]$ represent the state transition matrix for a time-invariant average dynamical system where $\overline{\mathcal{P}}_{m,n} = E\Big[prob\left[\mathcal{D}(k+1) = \phi^{(n)} | \mathcal{D}(k) = \phi^{(m)}\right]\Big]$. Let $\overline{\mathcal{V}_{\phi^{(m)}}^{(i)}} \triangleq E\Big[prob\big[b_i(k+1) = 1 | \mathcal{D}(k) = \phi^{(m)}\big]\Big]$, where $\mathcal{D}(k)$ denotes the state of the network at time k and $\phi^{(m)}$ is as defined in Section 4.1. Let sum(z') represent the sum of the elements of z' for any arbitrary vector z'. If $\overline{\mathcal{V}_{\phi^{(m)}}^{(i)}} = \frac{S_{i,\phi^{(m)}}}{M} + (1 - \frac{2S'_{i,\phi^{(m)}}}{M})\mathcal{C}$, where $\mathcal{C}$ is a constant (not a function of m or i), and $S_{i,\phi^{(m)}}$ and $S'_{i,\phi^{(m)}}$ are any positive numbers such that $\sum_{i=1}^{M} S_{i,\phi^{(m)}} = \sum_{i=1}^{M} S'_{i,\phi^{(m)}} = Msum(\phi^{(m)})$, then $1 - 2\mathcal{C}$ is one of the eigenvalues of $\overline{\mathcal{P}}$.*

Proof. For any $0 \leq m \leq 2^M - 1$ and $S(k) = \sum_{i=1}^{M} b_i(k)$, we have

$$E\Big[\sum_{j=0}^{M} j\text{prob}[S(k+1) = j | \mathcal{D}(k) = \phi^{(m)}]\Big] = \sum_{i=1}^{M} \overline{\mathcal{V}_{\phi^{(m)}}^{(i)}}.$$

If $\overline{\mathcal{V}_{\phi^{(m)}}^{(i)}} = \frac{S_{i,\phi^{(m)}}}{M} + (1 - \frac{2S'_{i,\phi^{(m)}}}{M})\mathcal{C}$ and $\sum_{i=1}^{M} S_{i,\phi^{(m)}} = \sum_{i=1}^{M} S'_{i,\phi^{(m)}} = M\text{sum}(\phi^{(m)})$, we can extend [89] to show that

$$\sum_{j=0}^{M} (\frac{M}{2} - j) E\Big[\text{prob}\big[S(k+1) = j | \mathcal{D}(k) = \phi^{(m)}\big]\Big] = (1 - 2\mathcal{C})\big(\frac{M}{2} - \text{sum}(\phi^{(m)})\big). \tag{4.17}$$

Let ζ represent a $2^M \times 1$ vector, where $\zeta_d = \frac{M}{2} - \text{sum}(\phi^{(d)})$ and $\xi = \overline{\mathcal{P}}\zeta$. Then, we have $\xi_d = \sum_{j=0}^{M}(\frac{M}{2} - j)E\Big[\text{prob}\big[S(k+1) = j|\mathcal{D}(k) = \phi^{(d)}\big]\Big] = \zeta_d(1-2\mathcal{C})$. Therefore, $1-2\mathcal{C}$ is an eigenvalue of $\overline{\mathcal{P}}$. □

By applying Assumption 1 and averaging Eq. 4.15 over fading, we have:

$$\overline{\nu^{(i)}_{\phi^{(m)}}} \approx \frac{1}{2} - \frac{1}{\sqrt{2\pi}\delta_i}\Big(.5 - \frac{1}{M}\big(b_i^m + \sum_{j\neq i}\gamma_{j,i}\delta_i b_j^m + \sum_{j\neq i}\frac{1-\gamma_{j,i}\delta_i}{2}\big)\Big), \tag{4.18}$$

where $\gamma_{j,i} = E_{\text{fading}}\{\frac{\widetilde{\alpha}_{j,i}(k)}{\widetilde{\sigma}_i(k)}\}$ and $\delta_i = \frac{1}{E_{\text{fading}}\{\frac{1}{\widetilde{\sigma}_i(k)}\}}$ with $E_{\text{fading}}\{.\}$ indicating averaging over fading. We can then write $\overline{\nu^{(i)}_{\phi^{(m)}}}$ as a function of $\overline{\nu^{(i)}_{\phi^{(0)}}}$ as follows:

$$\overline{\nu^{(i)}_{\phi^{(m)}}} \approx \frac{S_{i,\phi^{(m)}}}{M} + (1 - \frac{2S_{i,\phi^{(m)}}}{M})\overline{\nu^{(i)}_{\phi^{(0)}}}, \tag{4.19}$$

where $S_{i,\phi^{(m)}} = \frac{M(b_i^m + \sum_{j\neq i}\gamma_{j,i}\delta_i b_j^m)}{1+\sum_{j\neq i}\gamma_{j,i}\delta_i}$. Note that we dropped index k from $\overline{\nu^{(i)}_{\phi^{(m)}}}$ due to the stationarity assumption.

Theorem 2. *Assume that CNR$_{j,i}$s are i.i.d. exponential random variables with $\mu = \overline{CNR}_{j,i}$. Then we have $\gamma_{j,i} = \gamma$ and $\delta_i = \delta$. Let $\overline{P}_{approx}$ and $\lambda_{1,\overline{P}_{approx}}$ represent the approximation of matrix $\overline{P}$ and its second largest eigenvalue under Assumption 1 respectively. We have $\lambda_{1,\overline{P}_{approx}} = 1 - 2Q\Big(\frac{0.5-\frac{M-1}{2M}(1-\gamma\delta)}{\delta}\Big)$.*

Proof. If $\mu = \overline{\text{CNR}}_{j,i}$, we have the following under Assumption 1: $\overline{\nu^{(i)}_{\phi^{(m)}}} \approx \frac{S_{i,\phi^{(m)}}}{M} + (1 - \frac{2S_{i,\phi^{(m)}}}{M})\overline{\nu_{\phi^{(0)}}}$, where $S_{i,\phi^{(m)}} = \frac{M(b_i^m + \sum_{j\neq i}\gamma\delta b_j^m)}{1+(M-1)\gamma\delta}$ and $\overline{\nu_{\phi^{(0)}}} = \overline{\nu^{(i)}_{\phi^{(0)}}}$. From Eq. 4.18 and by using the approximation of Assumption 1, we have $\overline{\nu_{\phi^{(0)}}} \approx Q\Big(\frac{0.5-\frac{M-1}{2M}(1-\gamma\delta)}{\delta}\Big)$. Moreover, it can be easily confirmed that $\sum_{i=1}^{M} S_{i,\phi^{(m)}} = M\text{sum}(\phi^{(m)})$. Then, by applying Theorem 1, we have $1 - 2\overline{\nu_{\phi^{(0)}}} = 1 - 2Q\Big(\frac{0.5-\frac{M-1}{2M}(1-\gamma\delta)}{\delta}\Big)$ as one of the eigenvalues of $\overline{P}_{\text{approx}}$. As all CNR$_{j,i}$s go to infinity, this eigenvalue goes to one. Consider all the eigenvalues of $\overline{P}$ except for the first one. As mentioned earlier, only the second largest one goes to one as CNR$_{j,i}$s go to infinity. The same can be confirmed for $\overline{P}_{\text{approx}}$ [79]. Therefore, $\lambda_{1,\overline{P}_{\text{approx}}} = 1 - 2Q\Big(\frac{0.5-\frac{M-1}{2M}(1-\gamma\delta)}{\delta}\Big)$. □

4.2.4 Special case: unknown $\sigma_{j,i}$

As can be seen, the decision-making function of Eqs. 4.12 and 4.13 uses information of $\sigma_{j,i}$s. If such information is not available, then each receiver will use a zero-forcing equalizer to undo the impact of the channel and use the equalized received information assuming they are correct. We then have the following:

$$b_i(k+1) = \text{Dec}\left(\frac{1}{M}\left(b_i(k) + \sum_{j=1,j\neq i}^{M} \frac{b_{j,i}(k)}{r_{j,i}(k)}\right)\right) = \text{Dec}\left(\frac{S(k)}{M} + n_i(k)\right), \quad (4.20)$$

where $n_i(k) = \frac{1}{M}\sum_{j=1,j\neq i}^{M} \frac{n_{j,i}(k)}{r_{j,i}(k)}$ with the variance of $\sigma_i^2(k) = \frac{1}{M^2}\sum_{j=1,j\neq i}^{M} \frac{1}{\text{CNR}_{j,i}(k)}$. This receiver can be considered as a special case of the decision-making function of Eq. 4.14 where $\widetilde{\alpha}_{j,i}(k) = 1$ and $\widetilde{\beta}_i(k) = 0$.

Theorem 3. *Consider the case where knowledge of the noise variances is not available at the nodes. Take* $\text{CNR}_{j,i}(k) = \frac{r_{j,i}^2(k)}{\sigma_{j,i}^2}$ *to be i.i.d. exponentially-distributed random variables with* $\mu = \overline{\text{CNR}}_{j,i}$ *representing their average. We have* $\lambda_{1,\overline{P}_{approx}} = 1 - 2Q\Big(\frac{M}{2}E_{fading}\{\frac{1}{\sqrt{\sum_{j\neq i}\frac{1}{\text{CNR}_{j,i}(k)}}}\}\Big)$.

Proof. This case is a special case of basic BAE where $\widetilde{\alpha}_{j,i}(k) = 1$, $\widetilde{\beta}_i(k) = 0$ and $\widetilde{\sigma}_i^2(k) = \sigma_i^2(k) = \frac{1}{M^2}\sum_{j\neq i}\frac{1}{\text{CNR}_{j,i}(k)}$. Therefore, in this case $\delta = \delta_i = \frac{1}{E_{\text{fading}}\{\frac{1}{\sigma_i(k)}\}}$ and $\gamma = \gamma_{j,i} = E_{\text{fading}}\{\frac{1}{\sigma_i(k)}\}$ due to the stationarity assumption. Then using Theorem 2 with the aforementioned parameters will result in $\lambda_{1,\overline{P}_{\text{approx}}} = 1 - 2Q\Big(\frac{1}{2\delta}\Big) = 1 - 2Q\big(\frac{M}{2}E_{\text{fading}}\{\frac{1}{\sqrt{\sum_{j\neq i}\frac{1}{\text{CNR}_{j,i}(k)}}}\}\big)$. □

While Theorem 3 relates the transient behavior of the network to the link qualities, finding a closed-form expression for $E_{\text{fading}}\{\frac{1}{\sqrt{\sum_{j\neq i}\frac{1}{\text{CNR}_{j,i}(k)}}}\}$ is challenging for exponentially-distributed $\text{CNR}_{j,i}$s. Alternatively, we can derive another approximation for the second largest eigenvalue as follows. From Theorem 2, we have $\lambda_{1,\overline{P}_{\text{approx}}} = 1 - 2\overline{\nu_{\phi^{(0)}}}$. Instead of finding an expression for $\overline{\nu_{\phi^{(0)}}}$ using Assumption 1 as we did before, we can directly use Eq. 4.15 as follows: $\overline{\nu_{\phi^{(0)}}} = E_{\text{fading}}\Big\{Q\big(\frac{0.5}{\sigma_i(k)}\big)\Big\}$. Then from the definition of $\sigma_i(k)$,

we have $\frac{1}{2\sigma_i(k)} = \sqrt{\frac{.25M^2}{\sum_{j=1,j\neq i}^{M}\frac{1}{\text{CNR}_{j,i}(k)}}} \leq \sqrt{0.25M^2\text{CNR}_{i,\min}(k)}$, where $\text{CNR}_{i,\min}(k) = \min(\{\text{CNR}_{j,i}(k)|1 \leq j \leq M, j \neq i\})$. The minimum of $M-1$ i.i.d. exponential random variables, $\text{CNR}_{i,\min}(k)$, can be easily shown to have an exponential distribution with $\frac{\mu}{M-1}$ representing its average. For an arbitrary exponentially-distributed random variable u, with the average $\overline{u}$, we have $\overline{Q(\sqrt{au})} = \frac{1}{2}(1-\sqrt{\frac{0.5a\overline{u}}{1+0.5a\overline{u}}})$ for an arbitrary $a > 0$. Therefore,

$$\lambda_{1,\overline{P}_{\text{approx}}} \leq \sqrt{\frac{0.125M^2\mu/(M-1)}{1+0.125M^2\mu/(M-1)}}. \tag{4.21}$$

4.2.5 Consensus performance

In this section, our simulation results will confirm the theoretical derivations of the previous parts and show the performance of group consensus over fading channels. We start by considering the special case where $\sigma_{j,i}$s are unknown. We characterized the performance of this case in Section 4.2.4 Fig. 4.1 (left) and Fig. 4.1 (right) show the performance of a network of 4 nodes that is trying to reach consensus by communicating over AWGN and fast fading channels respectively. The nodes do not have any knowledge of noise variances in this case. Initially 3 out of 4 nodes are voting one. Then it is desirable that all four nodes vote one through communication. For the fading case, the figure shows $E\big[\text{prob}[S(k)=4]\big]$, i.e. the average probability of accurate consensus (averaged over fading), for different noise variances. Both figures show the performance for cases with poor link qualities. For instance, average $\text{CNR}_{j,i} = 0$dB means that average SNR per link is -3dB if sending 0 and 1 are equiprobable. Two observations can be made from the figure. First, it can be seen that at the earlier iterations, the probability of accurate consensus increases. However, after a while, communication is not beneficial anymore as it results in error propagation in the network, a decrease in the probability of accurate consensus and an eventual memoryless consensus. This is as expected from Section 4.2.2, where we characterized the asymptotic behavior of group consensus and showed that it is memoryless in

the presence of any amount of link uncertainty. Second, it can be seen that fading ruins the performance drastically by reducing the probability of accurate consensus. It therefore becomes considerably important to mathematically characterize the transient behavior and propose algorithms to improve the overall performance, as was done in the previous sections. In order to see how well the approximation of Theorem 3 works, Fig. 4.2 shows the 2nd largest eigenvalue of the average transition matrix ($\lambda_{1,\overline{P}}$), its approximation from Theorem 3 ($\lambda_{1,\overline{P}_{\text{approx}}}$), and the upper bound of Eq. 4.21. It can be seen that the approximation and its upper bound are considerably close to the true eigenvalue. As link qualities get worse (lower CNR), the linearization of the Q function provides a better approximation, resulting in the derivation of Theorem 3 getting closer to the true eigenvalue (see [118] for more details on this). It should be noted that the upper bound of Eq. 4.21 is derived for $\lambda_{1,\overline{P}_{\text{approx}}}$ and not for $\lambda_{1,\overline{P}}$.

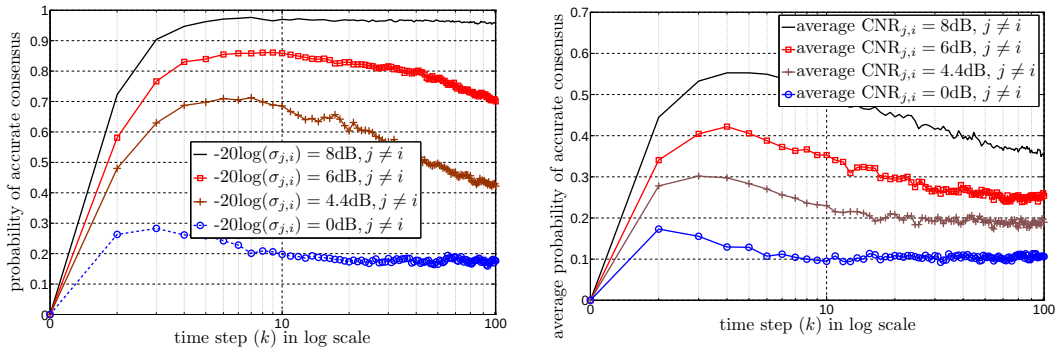

Figure 4.1: Comparison of the performance of binary consensus over AWGN and fading communication channels for $M = 4$ and the case where knowledge of $\sigma_{j,i}$s is not available (see Section 4.2.4) – (left) binary consensus over AWGN communication channels, (right) binary consensus over fading communication channels with $E[r_{j,i}^2] = 1, \forall i, j \neq i$ (averaging is done over several runs).

Next we consider the case where knowledge of $\sigma_{j,i}$s is available at each node and can be used in the decision-making process, as discussed in Section 4.2.1. Fig. 4.3 shows the probability of accurate consensus for the proposed BAE approaches of Section 4.2.1, where all the channels experience the same noise variance ($\sigma_{j,i}^2 = 1$) and the average

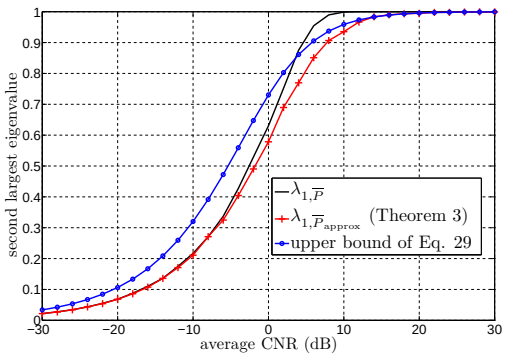

Figure 4.2: Characterization of the 2nd largest eigenvalue for the case where knowledge of $\sigma_{j,i}$s is not available (see Section 4.2.4) with $M = 4$, $\sigma^2_{j,i} = 1$ and $E[r^2_{j,i}] = 1\ \forall i, j \neq i$.

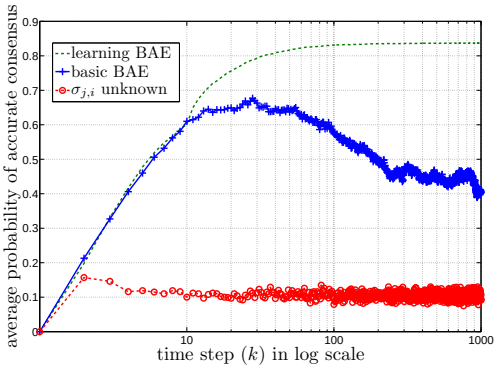

Figure 4.3: Average probability of accurate consensus for $M = 4$, $\sigma^2_{j,i} = 1$ and $E[r^2_{j,i}] = 1$ $\forall i, j \neq i$.

power of fading coefficients is equal to one ($E[r^2_{j,i}] = 1$). Therefore, if sending 0 and 1 is equiprobable, the average SNR of each link is -3dB, which is very low. The figure shows the performance of both basic (solid line with plus markers) and learning (dashed line) BAE approaches. As discussed in Section 4.2.1 for the basic BAE case, $q_j(k)$ is not estimated and is assumed to be 0.5. On the other hand, for the learning case, each node tries to estimate $q_j(k)$s, the voting patterns of other nodes. The performance for the case where knowledge of $\sigma_{j,i}$ is not available (special case) is also shown for comparison (dashed line with circle markers). It can be seen that using the knowledge of noise variances

can improve the performance drastically, as expected. Still the basic-BAE case has an undesirable memoryless asymptotic behavior as shown in Section 4.2.2, i.e. after a certain time the probability of accurate consensus starts to decrease. It can then be seen that by incorporating the online learning of $q_j(k)$, learning BAE can improve the performance and avoid the memoryless asymptotic behavior. However, characterizing the asymptotic and transient behavior of the learning BAE case, is a challenging problem since it requires considering the error in the online estimation of $q_j(k)$. Furthermore, learning BAE is more computationally complex, as compared to the basic BAE. Therefore, in the rest of this chapter, we only consider the basic BAE case.

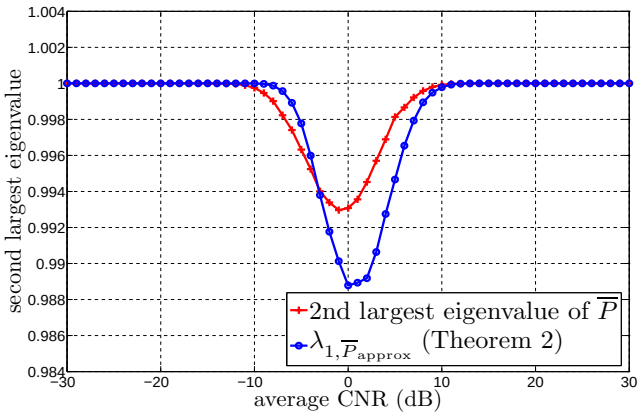

Figure 4.4: Characterization of the 2nd largest eigenvalue for the basic BAE case with $M = 4$, $\sigma^2_{j,i} = 1$ and $E[r^2_{j,i}] = 1\ \forall i, j \neq i$.

To see how well Theorem 2 approximates $\lambda_{1,\overline{P}}$ for the basic BAE case, Fig. 4.4 shows the second largest eigenvalue of $\overline{P}$ as well as $\lambda_{1,\overline{P}_{\text{approx}}}$. It can be seen that the approximation of Theorem 2 works well. At higher average CNRs, as average CNR increases, $\lambda_{1,\overline{P}}$ increases, as expected (similar to Fig. 4.2). However, at lower average CNR, $\lambda_{1,\overline{P}}$ increases as CNR decreases. This is due to the fact that the BAE approach weighs the received information based on link qualities. Therefore, at considerably low average CNR, the received information is almost ignored (as it should be), which results in $\lambda_{1,\overline{P}}$ approaching one. In other words, for very low CNRs, each node keeps its initial opinion resulting in an

identity P matrix. This makes the proposed approach more robust to the receiver noise. On the other hand, if the knowledge of $\sigma_{j,i}$ is not available, as is the case in Section 4.2.4, $\lambda_{1,\overline{P}}$ becomes a non-decreasing function of average CNR (see Fig. 4.2).

4.3 Binary consensus over not fully-connected rapidly-changing network topologies

In Section 4.2, we considered binary consensus over fully-connected fixed network topologies, in order to focus on the impact of fading and noise. In realistic scenarios, however, some links may not exist due to poor quality. Furthermore, the underlying communication topology could be time-varying. Therefore, in this part we relax those assumptions and consider the dynamics of binary consensus over not fully-connected and rapidly-changing network topologies. We model the communication network as a graph, where $\{1, \cdots, M\}$ represents the vertex set and $\mathcal{E}(k)$ is the link set at time k. $\mathcal{N}_i(k)$ then denotes the neighbor set of node i at time k (excluding itself), as introduced earlier. In a not fully-connected graph, there exists a link from node j to node i at time k if $\text{CNR}_{j,i}(k) > \text{CNR}_{\text{Th}}$, i.e. the link quality is above a minimum acceptable threshold. We take $\text{CNR}_{j,i}$s to be i.i.d. random variables with $\mu = \overline{\text{CNR}}_{j,i}$. Let p represent the probability that a link exists from node j to node i at a given time. In exponentially-distributed fading environments, we have: $p = \text{prob}\Big\{\text{CNR}_{j,i}(k) > \text{CNR}_{\text{Th}}\Big\} = e^{-\frac{\text{CNR}_{\text{Th}}}{\mu}}$. The case of $p = 0$ corresponds to the static empty graph and is not of interest to us. In this section, it is our goal to characterize the impact of time-varying not fully-connected graph topologies on consensus.

We start by considering the case where if a link exists, it is perfect, i.e. there is no error in that transmission. Studying this case allows us to solely focus on the impact of not fully-connected graphs and could correspond to the case where the threshold, CNR_{Th}, is chosen very high. We then consider binary consensus over not fully-connected rapidly-changing

topologies with fading channels. We also consider two decision-making approaches: *fusion* and *diversity*. So far in the previous section, we considered the case where each node fuses its received information in every time step, to form its new opinion, which it will then send to other nodes. In this manner, each node helps to propagate the information of other nodes in the network. This strategy is suitable, in particular, when the graph connectivity is low as it creates virtual links between nodes. We refer to it as *fusion strategy* in this section to differentiate it from the alternative diversity approach described next. Each node can also use its transmissions to repeat its initial vote, without fusing its received information. This strategy, to which we refer to as *diversity* strategy, on the other hand can be more robust to link errors. In our previous work, we introduced these two strategies in the context of binary consensus over fixed AWGN networks [118]. In this section, we consider both approaches in the context of time-varying network topologies. As we shall see, a time-varying graph with fading channels requires a new formulation and approach, which we address here.

4.3.1 Binary consensus over a not fully-connected and rapidly-changing topology with ideal links

In this section, we consider the case where if a link exists, its quality is perfect. We also assume an undirected graph in this part, which means that if $(i, j) \in \mathcal{E}(k)$, then $(j, i) \in \mathcal{E}(k)$. Then p denotes the probability that a link exists between two nodes at a given time.

Fusion case

For the fusion case, the decision-making function of Eq. 4.13 will be as follows ($\alpha_{j,i}(k) = 1$ and $\beta_{j,i}(k) = 0$ if there is a link from node j to node i at time k):

$$b_i(k+1) = \text{Dec}\left(\frac{1}{1+|\mathcal{N}_i(k)|}\big(b_i(k) + \sum_{j\in\mathcal{N}_i(k)} b_j(k)\big)\right), \tag{4.22}$$

if $\frac{1}{1+|\mathcal{N}_i(k)|}\big(b_i(k) + \sum_{j\in\mathcal{N}_i(t)} b_j(k)\big) \neq 0.5$. Otherwise $b_i(k+1) = b_i(k)$. The following theorem characterizes the performance of this decision-making function:

Theorem 4. *Consider binary consensus over a rapidly-changing graph with $M > 2$. Let p represent the probability that a link exists at a given time step. Consider the case where if a link exists, its quality is perfect. Then the decision-making function of Eq. 4.22 will have the following properties:*

1- The states with all votes 0 or all votes 1 are absorbing states.

2- If less than $M-1$ nodes vote the same initially, then there is no guarantee that the network converges to the accurate consensus.

3- If $M-1$ nodes vote the same initially, we have asymptotic accurate consensus.

4- The network asymptotically reaches consensus if $p \neq 1$.

Proof. First part can be easily confirmed. We next prove the second part. Assume that at $k = 0$, we have M nodes with some initial votes, where less than $M-1$ nodes vote the same. Let $\mathcal{I}^0(k)$ represent the set of indices of all those nodes that vote 0 at time step k. $\mathcal{I}^1(k)$ is defined in a similar manner for all those nodes that vote 1 at time step k. Then, $\mathcal{I}^0(0) = \{i_1^0, i_2^0, \cdots, i_{|\mathcal{I}^0(0)|}^0\}$ and $\mathcal{I}^1(0) = \{i_1^1, i_2^1, \cdots, i_{|\mathcal{I}^1(0)|}^1\}$ denote two mutually exclusive sets of indices at time $k = 0$, where all the corresponding nodes vote zero and one respectively. Without loss of generality, we assume that[3] $|\mathcal{I}^1(0)| > |\mathcal{I}^0(0)|$. Since less than $M-1$ nodes vote the same initially, we know that $|\mathcal{I}^0(0)| > 1$. Consider the

[3]Note that if $|\mathcal{I}^1(0)| = |\mathcal{I}^0(0)|$, i.e. $S(0) = \frac{M}{2}$, then there will be clearly no consensus.

case where at $k = 0$, there is only a fully-connected graph among nodes in $\mathcal{I}^0(0) \cup \{i_1^1\}$ with no other links in the network. Such a topology would occur with the probability of $p^{\binom{|\mathcal{I}^0(0)|+1}{2}}(1-p)^{\binom{M}{2}-\binom{|\mathcal{I}^0(0)|+1}{2}}$. Then we have the following at the next time step: $\mathcal{I}^0(1) = \mathcal{I}^0(0) \cup \{i_1^1\}$ and $\mathcal{I}^1(1) = \mathcal{I}^1(0) \setminus \{i_1^1\}$. Similarly, the probability that a fully-connected graph exists only among the nodes in $\mathcal{I}^0(1) \cup \{i_2^1\}$ (with no other link in the network) is $p^{\binom{|\mathcal{I}^0(1)|+1}{2}}(1-p)^{\binom{M}{2}-\binom{|\mathcal{I}^0(1)|+1}{2}}$. Then, we have $\mathcal{I}^0(2) = \mathcal{I}^0(1) \cup \{i_2^1\}$ and $\mathcal{I}^1(2) = \mathcal{I}^1(1) \setminus \{i_2^1\}$. By continuing the same procedure, we have all the nodes in the network voting zero after $|\mathcal{I}^1(0)|$ time steps, with a non-zero probability, while $|\mathcal{I}^1(0)| > |\mathcal{I}^0(0)|$. Since an all-zero state is an absorbing state (see Property 1), then the network will stay in inaccurate consensus. This example shows that there is no guarantee of convergence to an accurate consensus state. Next, we prove the third part. Without loss of generality, assume that $M-1$ nodes vote one initially. In the next time step, either no node changes its vote or all the nodes vote one, depending on the network topology. The probability that no node changes its vote can be characterized as $(M-1)p(1-p)^{M-2} + (1-p)^{M-1}$. Therefore, the probability that asymptotic accurate consensus is not reached goes to zero for $M > 2$: $\lim_{k\to\infty}\left[(M-1)p(1-p)^{M-2} + (1-p)^{M-1}\right]^k \to 0$. Finally, we have the following proof for part 4. Let $S(k)$ represent the sum of the votes at time k, as defined in Section 4.1. Define $\varsigma(k) = \begin{cases} 1, S(k) \geq \frac{M}{2} \\ 0, S(k) < \frac{M}{2} \end{cases}$ and $\mathcal{M}(k) = \max(S(k), M - S(k))$. We have $|\mathcal{I}^{\varsigma(k)}(k)| \geq |\mathcal{I}^{1-\varsigma(k)}(k)|$. Since $p \neq 1$, with the non-zero probability of $p^{\binom{\mathcal{M}(k)}{2}}(1-p)^{\binom{M-\mathcal{M}(k)}{2}}p^{\mathcal{M}(k)(M-\mathcal{M}(k))}$, we have a graph which is fully connected in $\mathcal{I}^{\varsigma(k)}(k)$, fully disconnected in $\mathcal{I}^{1-\varsigma(k)}(k)$ and with each node in $\mathcal{I}^{1-\varsigma(k)}(k)$ connected to all the nodes in $\mathcal{I}^{\varsigma(k)}(k)$. Since $M > 2$, having such a graph results in the network reaching a consensus, at time $k+1$, that corresponds to $\varsigma(k)$. Let $p^{\text{consensus}}(k+1)$ denote the probability that the network is in consensus at time $k+1$. We have $p^{\text{consensus}}(k+1) \geq p^{\binom{\mathcal{M}(k)}{2}}(1-p)^{\binom{M-\mathcal{M}(k)}{2}}p^{\mathcal{M}(k)(M-\mathcal{M}(k))} > 0$ for $k \geq 0$. Then the probability that the network does not hit the consensus state goes to zero asymptotically: $\lim_{k\to\infty}\prod_{t=0}^{k}\left[1 - p^{\text{consensus}}(t)\right] \to 0$. □

It should be noted that if $p = 1$, the network will reach accurate consensus in one step iff $S(0) \neq \frac{M}{2}$.

Diversity case

In this part, we consider another strategy in which each node uses its transmissions to repeat its initial vote. Consider the case where the network is given $K + 1$ time steps (K transmissions) to reach an agreement. Each node can use all its transmissions to repeat its initial vote and only fuses the received information afterwards. This strategy can, in particular, be useful in reducing the impact of link error, as we see in the subsequent sections. We will have the following for this case:

$$\begin{aligned} b_i(k) &= b_i(k-1) \text{ for } 1 \leq k \leq K-1, \\ b_i(K) &= \text{Dec}\Big(\frac{1}{M}\big(b_i(0) + \frac{1}{Kp}\sum_{t=0}^{K-1}\sum_{j\in\mathcal{N}_i(t)} b_j(t)\big)\Big). \end{aligned} \tag{4.23}$$

Theorem 5. *Consider binary consensus over a rapidly-changing network topology where p represents the probability that a link exists at a given time step. Consider the case where if a link exists, its quality is perfect. Then the decision-making function of Eq. 4.23 results in asymptotic accurate consensus almost surely iff $S(0) \neq \frac{M}{2}$. Furthermore, the probability of accurate consensus at time step K can be approximated by $\prod_{i=1}^{M} Q\big(\frac{\frac{M}{2}-S(0)}{\sqrt{\big(S(0)-b_i(0)\big)\frac{1-p}{p}}}\sqrt{K}\big)$, for M sufficiently large and $S(0) > \frac{M}{2}$.*

Proof. For all $1 \leq j \leq M$ and $j \neq i$, define the following sequence of independent random variables for node i:

$$X_j^i(t) = \begin{cases} b_j(0) & \text{if there is a link between node } i \\ & \text{and } j \text{ at time step } t \\ 0 & \text{else} \end{cases} \tag{4.24}$$

Then we have $\frac{1}{Kp}\sum_{t=0}^{K-1}\sum_{j\in\mathcal{N}_i(t)} b_j(0) = \frac{1}{Kp}\sum_{t=0}^{K-1}\sum_{j=1,j\neq i}^{M} X_j^i(t) = \frac{1}{p}\sum_{j=1,j\neq i}^{M}\frac{1}{K}\sum_{t=0}^{K-1} X_j^i(t)$. The strong law of large numbers states that the sample average converges almost surely (a.s.) to the expected value: $\text{prob}\Big(\lim_{K\to\infty}\frac{1}{K}\sum_{t=0}^{K-1} X_j^i(t) = pb_j(0)\Big) = 1$, which results in

$$\frac{1}{Kp}\sum_{t=0}^{K-1}\sum_{j\in\mathcal{N}_i(t)} b_j(0) \xrightarrow{a.s.} \sum_{j=1,j\neq i}^{M} b_j(0). \tag{4.25}$$

Then the argument inside the decision function of Eq. 4.23 almost surely converges to the average of the initial votes. Therefore, if $S(0) \neq \frac{M}{2}$, the decision-making function of Eq. 4.23 results in asymptotic accurate consensus with probability of 1. Next we characterize the probability of accurate consensus for the case that the majority of initial votes is one.[4] Let $Y^i(K) = b_i(0) + \frac{1}{p}\sum_{j=1,j\neq i}^{M}\frac{1}{K}\sum_{t=0}^{K-1} X_j^i(t)$. For large enough M, we can evoke the Central Limit Theorem to approximate the distribution of $Y^i(K)$ with a Gaussian, with the following average and variance: $\mu_{Y^i(K)} = S(0)$ and $\sigma^2_{Y^i(K)} = \frac{S(0)-b_i(0)}{Kp}(1-p)$. The probability that $Y^i(K)$ is exactly $\frac{M}{2}$ for most values of p, K and M is negligible. Therefore, we have the following approximation by excluding this case:

$$\text{prob}(b_i(K)=1) = \text{prob}\big(\frac{Y^i(K)}{M} > \frac{1}{2}\big) \approx Q\big(\frac{\frac{M}{2}-S(0)}{\sqrt{\big(S(0)-b_i(0)\big)\frac{1-p}{p}}}\sqrt{K}\big), \tag{4.26}$$

resulting in the following for $S(0) > \frac{M}{2}$:

$$\text{prob(accurate consensus at time } K) \approx \prod_{i=1}^{M} Q\big(\frac{\frac{M}{2}-S(0)}{\sqrt{\big(S(0)-b_i(0)\big)\frac{1-p}{p}}}\sqrt{K}\big). \tag{4.27}$$

□

Fig. 4.5 shows the performance of the diversity case for a network of 20 nodes. It can be seen that the approximation of Eq. 4.27 matches the true probability of accurate consensus considerably well.

Fig. 4.6 compares the performance of fusion and diversity strategies for $p = 0.5$ and $p = 0.8$. It can be seen that the fusion scheme provides a faster convergence rate through

[4] similar expressions can be derived for the case that the majority of initial votes is zero.

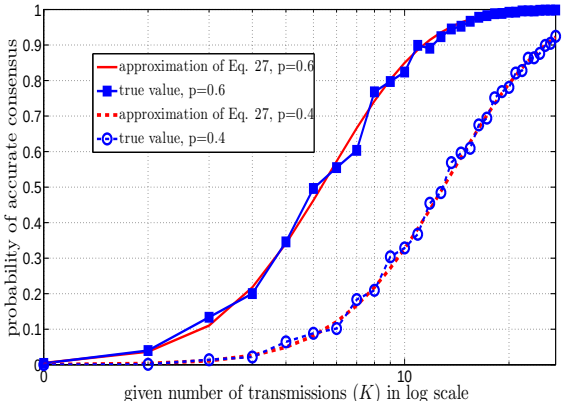

Figure 4.5: Binary consensus over a rapidly-changing network topology with ideal communication links for the case of diversity with $M = 20$.

fusing the received information, which helps propagating the information over the network. As p decreases, the fusion strategy outperforms the diversity one more drastically, as expected. However, it lacks asymptotic accurate convergence guarantees, as was shown in Theorem 4.

4.3.2 Binary consensus over a not fully-connected and time-varying topology with fading channels

In this section we extend our analysis to the case where links experience fading and noise. We consider the case where $\text{CNR}_{j,i}(k)$s can be modeled as i.i.d. random variables with average of μ. Due to the presence of fading, the graph will be directed in this case, i.e. there could be a link from node j to node i with no reverse link. More specifically, there exists a link from node j to node i, at time k, if $\text{CNR}_{j,i}(k) > \text{CNR}_{\text{Th}}$, for a given threshold of $\text{CNR}_{\text{Th}} \geq 0$. We only consider the case where knowledge of CNR is available in the receiver. Let $\text{CNR}_i^{\text{set}}(k)$ represent the set of $\text{CNR}_{j,i}(k)$s where $j \in \mathcal{N}_i(k)$. Then node i has access to this set. Similar to Section 4.2.4, the results of this section can be easily extended to the case where such knowledge is not fully available.

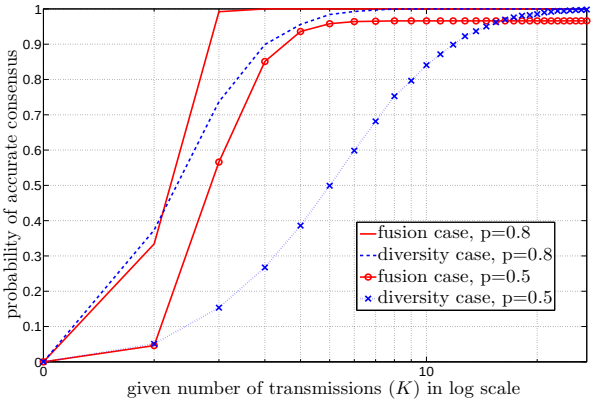

Figure 4.6: Binary consensus over a rapidly-changing network topology with ideal communication links for $M = 10$, comparison of diversity and fusion strategies.

Fusion case

For the fusion case, we will have the following decision-making function by extending Eq. 4.13 to a not fully-connected time-varying topology:

$$b_i(k+1) = \text{Dec}\Big(\frac{1}{1+|\mathcal{N}_i(k)|}\Big[b_i(k) + \sum_{j\in\mathcal{N}_i(k)} \widetilde{\alpha}_{j,i}(k)\frac{b_{j,i}(k)}{r_{j,i}(k)} + \widetilde{\beta}_i(k)\Big]\Big)$$
$$= \text{Dec}\Big(\frac{1}{1+|\mathcal{N}_i(k)|}\Big[b_i(k) + \sum_{j\in\mathcal{N}_i(k)} \widetilde{\alpha}_{j,i}(k)b_j(k) + \widetilde{\beta}_i(k) + \widetilde{n}_i(k)\Big]\Big), \quad (4.28)$$

where $\widetilde{\alpha}_{j,i}(k) = \frac{1}{1+\frac{4}{\text{CNR}_{j,i}(k)}}$ and $\widetilde{\beta}_i(k) = \sum_{j\in\mathcal{N}_i(k)} \frac{1-\widetilde{\alpha}_{j,i}(k)}{2}$. Furthermore, $\widetilde{n}_i(k)$ is a zero-mean Gaussian random variable with the variance of $\widetilde{\sigma}_i^2(k) = \frac{1}{(1+|\mathcal{N}_i(k)|)^2}\sum_{j\in\mathcal{N}_i(k)} \frac{\frac{1}{\text{CNR}_{j,i}(k)}}{\Big(1+\frac{4}{\text{CNR}_{j,i}(k)}\Big)^2}$.

Let $\iota_1, \iota_2, \cdots, \iota_m$ be i.i.d. random variables. For a given $\mu_{\text{Th}} \geq 0$, we define the following functions:

$$f(m, \mu_{\text{Th}}) = E\Big\{\frac{1}{\sqrt{\sum_{i=1}^{m} \frac{\frac{1}{\iota_i}}{(1+\frac{4}{\iota_i})^2}}} \mid \iota_1 > \mu_{\text{Th}}, \cdots, \iota_m > \mu_{\text{Th}}\Big\},$$

$$g(m,\mu_{\text{Th}}) = E\Big\{ \frac{\frac{1}{1+\frac{4}{\iota_1}}}{\sqrt{\sum_{i=1}^{m} \frac{\frac{1}{\iota_i}}{(1+\frac{4}{\iota_i})^2}}} \mid \iota_1 > \mu_{\text{Th}}, \cdots, \iota_m > \mu_{\text{Th}} \Big\}.$$

Theorem 6. *Consider binary consensus over a rapidly-changing network topology with i.i.d. fading channels and additive Gaussian noise, where* $p = prob\Big\{CNR_{j,i}(k) > CNR_{Th}\Big\}$ *represents the probability that a link exists at a given time step. Then the average dynamical system, based on the decision-making function of Eq. 4.28, will have asymptotic memoryless behavior if* $p \neq 0$ *and* $\forall i,k$ *that* $|\mathcal{N}_i(k)| \neq 0 \rightarrow \widetilde{\sigma}_i(k) \neq 0$*. Furthermore, we have the following approximation for the second largest eigenvalue of the average of the underlying dynamical system:*

$$\lambda_{1,\overline{P}_{approx,TV,fading}} = 1 - 2\big(1-(1-p)^{M-1}\big)Q\Big(\frac{\rho_1 + (M-1)\rho_2}{2(1-(1-p)^{M-1})}\Big), \tag{4.29}$$

where $\rho_1 = \sum_{m=1}^{M-1} f(m, CNR_{Th}) \binom{M-1}{m} p^m (1-p)^{M-1-m}$, $\rho_2 = \sum_{m=1}^{M-1} g(m, CNR_{Th}) \binom{M-2}{m-1} p^m (1-p)^{M-1-m}$ *and subscript TV indicates the case of time-varying network topology.*

Proof. Let $D(k)$, $\phi^{(m)}$ and b_j^m be as defined for Eq. 4.6. Let $\nu^{(i)}_{\phi^{(m)},\mathcal{N}_i(k),\text{CNR}_i^{\text{set}}(k)}(k) = \text{prob}[b_i(k+1) = 1 | D(k) = \phi^{(m)}, \mathcal{N}_i(k) \text{ and } \text{CNR}_i^{\text{set}}(k)]$ represent the probability that node i votes one in the next time step, given the current state of $D(k) = \phi^{(m)}$, the current connectivity set of $\mathcal{N}_i(k)$ and the current fading coefficients of $\text{CNR}_i^{\text{set}}(k)$. We have,

$$\nu^{(i)}_{\phi^{(m)},\mathcal{N}_i(k),\text{CNR}_i^{\text{set}}(k)}(k) = \begin{cases} Q\Big(\frac{\frac{1}{2} - \frac{b_i^m + \sum_{j\in\mathcal{N}_i(k)} \widetilde{\alpha}_{j,i}(k) b_j^m + \sum_{j\in\mathcal{N}_i(k)} \frac{1-\widetilde{\alpha}_{j,i}(k)}{2}}{1+|\mathcal{N}_i(k)|}}{\widetilde{\sigma}_i(k)}\Big) & |\mathcal{N}_i(k)| \neq 0 \\ b_i^m & |\mathcal{N}_i(k)| = 0 \end{cases} \tag{4.30}$$

Then we have the following for the average of the state transition matrix (averaged over different graph possibilities and fading coefficients):

$$\overline{P}_{m,n} = \prod_{i=1}^{M} \Big[b_i^n \overline{\nu^{(i)}_{\phi^{(m)}}(k)} + (1-b_i^n)\big(1 - \overline{\nu^{(i)}_{\phi^{(m)}}(k)}\big)\Big], \tag{4.31}$$

where $\overline{\nu^{(i)}_{\phi^{(m)}}(k)} = E_{\text{graph,fading}}\{\nu^{(i)}_{\phi^{(m)},\mathcal{N}_i(k),\text{CNR}_i^{\text{set}}(k)}(k)\}$ represents the average of $\nu^{(i)}_{\phi^{(m)},\mathcal{N}_i(k),\text{CNR}_i^{\text{set}}(k)}(k)$

over graph possibilities and fading:

$$\overline{\nu^{(i)}_{\phi^{(m)}}(k)} = \text{prob}\Big\{|\mathcal{N}_i(k)|=0\Big\}b_i^m + \text{prob}\Big\{|\mathcal{N}_i(k)|\neq 0\Big\}$$
$$\times E_{\text{graph,fading}}\Big\{Q\Big(\frac{\frac{1}{2} - \frac{b_i^m + \sum_{j\in\mathcal{N}_i(k)}\widetilde{\alpha}_{j,i}(k)b_j^m + \sum_{j\in\mathcal{N}_i(k)}\frac{1-\widetilde{\alpha}_{j,i}(k)}{2}}{1+|\mathcal{N}_i(k)|}}{\widetilde{\sigma}_i(k)}\Big)\Big||\mathcal{N}_i(k)|\neq 0\Big\}. \quad (4.32)$$

If $p\neq 0$ and $\forall i,k$ that $|\mathcal{N}_i(k)|\neq 0$, then $\widetilde{\sigma}_i(k)\neq 0$, it can be seen that $0<\overline{\nu^{(i)}_{\phi^{(m)}}(k)}<1$ and therefore $\overline{P}_{m,n}>0$. Furthermore $\overline{P}$ is stochastic. Consider the average dynamical system. Then similar to Section 4.2.2, it can be easily confirmed that, asymptotically, the memory of the initial state will be lost by evoking Perron's Theorem. Next, we characterize an approximation for the second largest eigenvalue in order to characterize the transient behavior. Under Assumption 1, we have the following approximation:

$$Q(\frac{\frac{1}{2} - \frac{b_i^m + \sum_{j\in\mathcal{N}_i(k)}\widetilde{\alpha}_{j,i}(k)b_j^m + \widetilde{\beta}_i(k)}{1+|\mathcal{N}_i(k)|}}{\widetilde{\sigma}_i(k)}) \approx \frac{1}{2} - c'\Big(\frac{1}{2\widetilde{\sigma}_i(k)} - (\frac{b_i^m}{(1+|\mathcal{N}_i(k)|)\widetilde{\sigma}_i(k)}$$
$$+ \frac{\sum_{j\in\mathcal{N}_i(k)}\widetilde{\alpha}_{j,i}(k)b_j^m}{(1+|\mathcal{N}_i(k)|)\widetilde{\sigma}_i(k)}) - \frac{|\mathcal{N}_i(k)| - \sum_{j\in\mathcal{N}_i(k)}\widetilde{\alpha}_{j,i}(k)}{2(1+|\mathcal{N}_i(k)|)\widetilde{\sigma}_i(k)}\Big), \quad (4.33)$$

where $c'=\frac{1}{\sqrt{2\pi}}$. We will then have the following equations where ρ_1 and ρ_2 are as defined in Theorem 6 and $\rho_3=\sum_{m=1}^{M-1} f(m,\text{CNR}_{\text{Th}})\, m\binom{M-1}{m}p^m(1-p)^{M-1-m}$:

$$E_{\text{graph,fading}}\Big\{\frac{1}{(1+|\mathcal{N}_i(k)|)\widetilde{\sigma}_i(k)} \mid |\mathcal{N}_i(k)|\neq 0\Big\} = \frac{\rho_1}{1-(1-p)^{M-1}},$$
$$E_{\text{graph,fading}}\Big\{\frac{|\mathcal{N}_i(k)|}{(1+|\mathcal{N}_i(k)|)\widetilde{\sigma}_i(k)} \mid |\mathcal{N}_i(k)|\neq 0\Big\} = \frac{\rho_3}{1-(1-p)^{M-1}},$$
$$E_{\text{graph,fading}}\Big\{\frac{\sum_{j\in\mathcal{N}_i(k)}\widetilde{\alpha}_{j,i}(k)b_j^m}{(1+|\mathcal{N}_i(k)|)\widetilde{\sigma}_i(k)} \mid |\mathcal{N}_i(k)|\neq 0\Big\} = \frac{\rho_2}{1-(1-p)^{M-1}}\sum_{j\neq i} b_j^m,$$
$$E_{\text{graph,fading}}\Big\{\frac{\sum_{j\in\mathcal{N}_i(k)}\widetilde{\alpha}_{j,i}(k)}{(1+|\mathcal{N}_i(k)|)\widetilde{\sigma}_i(k)} \mid |\mathcal{N}_i(k)|\neq 0\Big\} = \frac{(M-1)\rho_2}{1-(1-p)^{M-1}}. \quad (4.34)$$

Next we provide the proofs of the above equations. Assume $\text{CNR}_{j,i}(k)$s are i.i.d. ran-

dom variables, we have

$$E_{\text{graph,fading}}\Big\{\frac{1}{(1+|\mathcal{N}_i(k)|)\widetilde{\sigma}_i(k)} \mid |\mathcal{N}_i(k)|\neq 0\Big\} \tag{4.35}$$
$$= E_{\text{fading}}\Big\{\sum_{n=1}^{M-1}\sum_{\substack{\mathcal{N}_i(k)\\ \text{s.t. } |\mathcal{N}_i(k)|=n}} \frac{\frac{p^n(1-p)^{M-1-n}}{1-(1-p)^{M-1}}}{\sqrt{\sum_{j\in\mathcal{N}_i(k)}\frac{\frac{1}{\text{CNR}_{j,i}(k)}}{(1+\frac{4}{\text{CNR}_{j,i}(k)})^2}}}\Big\}$$
$$= \sum_{n=1}^{M-1}\binom{M-1}{n}\frac{p^n(1-p)^{M-1-n}}{1-(1-p)^{M-1}} \times E_{\text{fading}}\Big\{\frac{1}{\sqrt{\sum_{\substack{j\in\mathcal{N}_i(k)\\ \text{s.t. } |\mathcal{N}_i(k)|=n}}\frac{\frac{1}{\text{CNR}_{j,i}(k)}}{(1+\frac{4}{\text{CNR}_{j,i}(k)})^2}}}\Big\}$$
$$= \frac{\rho_1}{1-(1-p)^{M-1}}$$

and

$$E_{\text{graph,fading}}\Big\{\frac{|\mathcal{N}_i(k)|}{(1+|\mathcal{N}_i(k)|)\widetilde{\sigma}_i(k)} \mid |\mathcal{N}_i(k)|\neq 0\Big\} \tag{4.36}$$
$$= E_{\text{fading}}\Big\{\sum_{n=1}^{M-1}\sum_{\substack{\mathcal{N}_i(k)\\ \text{s.t. } |\mathcal{N}_i(k)|=n}} \frac{\frac{p^n(1-p)^{M-1-n}}{1-(1-p)^{M-1}}\times n}{\sqrt{\sum_{j\in\mathcal{N}_i(k)}\frac{\frac{1}{\text{CNR}_{j,i}(k)}}{(1+\frac{4}{\text{CNR}_{j,i}(k)})^2}}}\Big\}$$
$$= \sum_{n=1}^{M-1} n\binom{M-1}{n}\frac{p^n(1-p)^{M-1-n}}{1-(1-p)^{M-1}} E_{\text{fading}}\Big\{\frac{1}{\sqrt{\sum_{\substack{j\in\mathcal{N}_i(k)\\ \text{s.t. } |\mathcal{N}_i(k)|=n}}\frac{\frac{1}{\text{CNR}_{j,i}(k)}}{(1+\frac{4}{\text{CNR}_{j,i}(k)})^2}}}\Big\}$$
$$= \frac{\rho_3}{1-(1-p)^{M-1}}.$$

Let b_j^m denote the vote of the j^{th} node given that the current state is $\phi^{(m)}$, as defined in Section 4.1. We have $\phi^{(m)} = [b_1^m\ b_2^m \cdots b_M^m]$, we have:

$$
\begin{aligned}
&E_{\text{graph,fading}}\Big\{\frac{\sum_{j\in\mathcal{N}_i(k)}\widetilde{\alpha}_{j,i}(k)b_j^m}{(1+|\mathcal{N}_i(k)|)\widetilde{\sigma}_i(k)} \mid |\mathcal{N}_i(k)|\neq 0\Big\}\\
&=\sum_{n=1}^{M-1}\frac{p^n(1-p)^{M-1-n}}{1-(1-p)^{M-1}}\times\sum_{\substack{\mathcal{N}_i(k)\\ \text{s.t. }|\mathcal{N}_i(k)|=n}}\sum_{j\in\mathcal{N}_i(k)}b_j^m E_{\text{fading}}\Big\{\frac{\widetilde{\alpha}_{j,i}(k)}{(1+n)\widetilde{\sigma}_i(k)}\Big\}\\
&=\sum_{n=1}^{M-1}\frac{p^n(1-p)^{M-1-n}}{1-(1-p)^{M-1}}\times\sum_{\substack{\mathcal{N}_i(k)\\ \text{s.t. }|\mathcal{N}_i(k)|=n}}\sum_{j\in\mathcal{N}_i(k)}b_j^m E_{\text{fading}}\Bigg\{\frac{\frac{1}{1+\frac{4}{\text{CNR}_{j,i}(k)}}}{\sqrt{\sum_{\substack{l\in\mathcal{N}_i(k)\\ \text{s.t. }|\mathcal{N}_i(k)|=n}}\frac{\frac{1}{\text{CNR}_{l,i}(k)}}{(1+\frac{4}{\text{CNR}_{l,i}(k)})^2}}}\Bigg\}\\
&=\sum_{n=1}^{M-1}\frac{p^n(1-p)^{M-1-n}}{1-(1-p)^{M-1}}g(n,\text{CNR}_{\text{Th}})\binom{M-2}{n-1}\sum_{j\neq i}b_j^m\\
&=\frac{\rho_2}{1-(1-p)^{M-1}}\sum_{j\neq i}b_j^m. \qquad (4.37)
\end{aligned}
$$

The last equation of 4.34 can then be easily confirmed by inserting $\phi^{(m)}=[1,1\cdots 1]$ in Eq. 4.37. Then we have the following by using the approximation of Eq. 4.33

$$
\begin{aligned}
\overline{\nu_{\phi^{(m)}}^{(i)}} &\approx (1-p)^{M-1}b_i^m+\frac{1}{2}(1-(1-p)^{M-1})\\
&-\frac{1}{2}c'(1-(1-p)^{M-1})E_{\text{graph,fading}}\Big\{\frac{1}{\widetilde{\sigma}_i(k)}\mid|\mathcal{N}_i(k)|\neq 0\Big\}\\
&+c'(1-(1-p)^{M-1})\\
&\times E_{\text{graph,fading}}\Big\{\frac{b_i^m+\sum_{j\in\mathcal{N}_i(k)}\widetilde{\alpha}_{j,i}(k)b_j^m}{(1+|\mathcal{N}_i(k)|)\widetilde{\sigma}_i(k)}\mid|\mathcal{N}_i(k)|\neq 0\Big\}\\
&+\frac{1}{2}c'(1-(1-p)^{M-1})\\
&\times E_{\text{graph,fading}}\Big\{\frac{|\mathcal{N}_i(k)|}{(1+|\mathcal{N}_i(k)|)\widetilde{\sigma}_i(k)}\mid|\mathcal{N}_i(k)|\neq 0\Big\}\\
&-\frac{1}{2}c'(1-(1-p)^{M-1})\\
&\times E_{\text{graph,fading}}\Big\{\frac{\sum_{j\in\mathcal{N}_i(k)}\widetilde{\alpha}_{j,i}(k)}{(1+|\mathcal{N}_i(k)|)\widetilde{\sigma}_i(k)}\mid|\mathcal{N}_i(k)|\neq 0\Big\}, \qquad (4.38)
\end{aligned}
$$

which results in

$$\begin{aligned}\overline{\nu^{(i)}_{\phi^{(m)}}} &\approx (1-p)^{M-1}b_i^m + \frac{1}{2}(1-(1-p)^{M-1}) \\ &- \frac{1}{2}c'(\rho_1+\rho_3) + c'\rho_1 b_i^m + c'\rho_2\sum_{j\neq i} b_j^m + \frac{1}{2}c'\rho_3 \\ &- \frac{1}{2}c'(M-1)\rho_2 = (1-p)^{M-1}b_i^m \\ &+ \frac{1}{2}(1-(1-p)^{M-1}) - \frac{1}{2}c'(\rho_1+(M-1)\rho_2) \\ &+ c'\rho_1 b_i^m + c'\rho_2\sum_{j\neq i} b_j^m. \end{aligned} \tag{4.39}$$

After a straightforward derivation, it can be confirmed that:

$$\overline{\nu^{(i)}_{\phi^{(m)}}} \approx \frac{S_{i,\phi^{(m)}}}{M} + (1-\frac{2S'_{i,\phi^{(m)}}}{M})\overline{\nu^{(i)}_{\phi^{(0)}}}, \tag{4.40}$$

where $S'_{i,\phi^{(m)}} = M\frac{\rho_1 b_i^m+\rho_2\sum_{j\neq i}b_j^m}{\rho_1+(M-1)\rho_2}$, $S_{i,\phi^{(m)}} = (1-p)^{M-1}Mb_i^m+(1-(1-p)^{M-1})S'_{i,\phi^{(m)}}$ and $\overline{\nu^{(i)}_{\phi^{(0)}}} \approx (1-(1-p)^{M-1})(\frac{1}{2}-\frac{c'}{2}\frac{\rho_1+(M-1)\rho_2}{(1-(1-p)^{M-1})}) \approx (1-(1-p)^{M-1})Q\Big(\frac{\rho_1+(M-1)\rho_2}{2(1-(1-p)^{M-1})}\Big)$. It can be easily confirmed that $\sum_{i=1}^M S_{i,\phi^{(m)}} = \sum_{i=1}^M S'_{i,\phi^{(m)}} = M\text{sum}(\phi^{(m)})$. Then by applying Theorem 1, we have $1-2\big(1-(1-p)^{M-1}\big)Q\Big(\frac{\rho_1+(M-1)\rho_2}{2(1-(1-p)^{M-1})}\Big)$ as one of the eigenvalues of the average underlying dynamical system. As p goes to one and $\forall i, j\neq i$ $\text{CNR}_{j,i}$ goes to infinity, this eigenvalue goes to one. Therefore, it is the second largest one. □

If Theorem 6 is deployed with $\text{CNR}_{\text{Th}} = 0$, i.e. for a fully-connected time-invariant topology, we will have $\rho_1 = \frac{1}{M\delta}$ and $\rho_2 = \frac{\gamma}{M}$, where δ and γ are as defined in Theorem 2. This results in the second largest eigenvalue that is derived in Theorem 2, as expected.

Diversity case

The diversity decision-making strategy of Eq. 4.23 can be extended to the case of fading channels as follows:

$$\begin{aligned} b_i(k) &= b_i(k-1) \text{ for } 1 \leq k \leq K-1, \\ b_i(K) &= \text{Dec}\Big(\frac{1}{M}\big(b_i(0) + \frac{1}{Kp}\sum_{t=0}^{K-1}\sum_{j\in\mathcal{N}_i(t)}\frac{b_{j,i}(t)}{r_{j,i}(t)}\big)\Big), \end{aligned}$$

where $b_{j,i}(t) = r_{j,i}(t)b_j(0) + n_{j,i}(t)$, $E\{n_{j,i}^2(t)\} = \sigma_{j,i}^2$ and $r_{j,i}(t)$ represents the fading coefficient of the link from node j to node i for $0 \leq t \leq K-1$. As expected, this strategy will help when links experience poor quality due to fading or noise. The asymptotic accurate consensus behavior of this approach can be established in a similar manner, as follows:

Theorem 7. *Consider binary consensus over a rapidly-changing network topology with i.i.d. fading channels where* $p = prob\Big\{CNR_{j,i}(k) > CNR_{Th}\Big\}$ *represents the probability that a link exists. Then the decision-making function of Eq. 4.41 results in asymptotic accurate consensus almost surely iff* $S(0) \neq \frac{M}{2}$ *and* $CNR_{Th} > 0$.

Proof. We have: $\frac{1}{Kp}\sum_{t=0}^{K-1}\sum_{j\in\mathcal{N}_i(t)}\frac{b_{j,i}(t)}{r_{j,i}(t)} = \frac{1}{Kp}\sum_{t=0}^{K-1}\sum_{j\in\mathcal{N}_i(t)} b_j(0) + \frac{1}{Kp}\sum_{t=0}^{K-1}\sum_{j\in\mathcal{N}_i(t)}\frac{n_{j,i}(t)}{r_{j,i}(t)}$. The first term on the right-hand side almost surely converges to $\sum_{j=1,j\neq i}^{M} b_j(0)$. For the second term, which is the noise term, we have:

$$E\Big\{\big(\frac{1}{Kp}\sum_{t=0}^{K-1}\sum_{j\in\mathcal{N}_i(t)}\frac{n_{j,i}(t)}{r_{j,i}(t)}\big)^2\Big\} = E_{\text{graph}}\Big\{\frac{1}{K^2p^2}\sum_{t=0}^{K-1}\sum_{j\in\mathcal{N}_i(t)} E_{\text{fading}}\big\{\frac{1}{\text{CNR}_{j,i}(t)}|\text{CNR}_{j,i}(t) > \text{CNR}_{\text{Th}}\big\}\Big\}$$

$$= E_{\text{graph}}\Big\{\frac{\overline{\text{CNR}}_{\text{inv,cond}}}{K^2p^2}\sum_{t=0}^{K-1}|\mathcal{N}_i(t)|\Big\} = \frac{(M-1)\overline{\text{CNR}}_{\text{inv,cond}}}{Kp}, \tag{4.41}$$

where $\overline{\text{CNR}}_{\text{inv,cond}} = E_{\text{fading}}\big\{\frac{1}{\text{CNR}_{j,i}(t)}|\text{CNR}_{j,i}(t) > \text{CNR}_{\text{Th}}\big\}$ is finite for any $\text{CNR}_{\text{Th}} > 0$. For instance, for exponentially distributed i.i.d. fading coefficients, we have $\overline{\text{CNR}}_{\text{inv,cond}} = \frac{1}{\mu}\exp(\frac{\text{CNR}_{\text{Th}}}{\mu})\text{E}_1(\frac{\text{CNR}_{\text{Th}}}{\mu})$, where $\mu = E\{\text{CNR}_{j,i}(t)\}$ and $\text{E}_1(z_{\text{Th}}) = \int_{z_{\text{Th}}}^{\infty}\frac{e^{-u}}{u}du$ for an arbitrary u and $z_{\text{Th}} > 0$. Therefore, as $K \to \infty$, variance of this noise goes to zero. Hence, the

argument of the decision function of Eq. 4.41 asymptotically and almost surely converges to the average of the initial votes, which results in asymptotic accurate consensus with the probability of 1 as long as $S(0) \neq \frac{M}{2}$. □

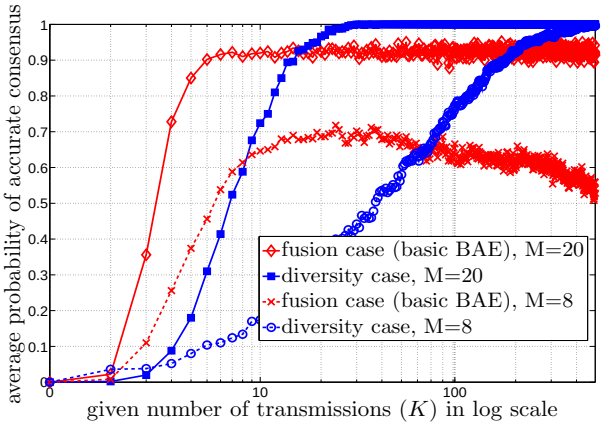

Figure 4.7: Comparison of fusion and diversity strategies for binary consensus over a rapidly-changing graph with fading links, $\text{CNR}_{\text{Th}} = -7$dB and $p = 0.82$.

Similar to Theorem 5, we can find an approximation for the average probability of accurate consensus (averaged over noise, fading and graph) in this case. Fig. 4.7 shows a comparison of the fusion and diversity strategies for fading case and for $\overline{\text{CNR}}$ =0dB, $\text{CNR}_{\text{Th}} = -7$dB (which results in $p = 0.82$) and $M = 8, 20$. As can be seen, the diversity strategy reaches asymptotic accurate consensus with probability of one, as expected. The fusion strategy, on the other hand, has a better transient behavior. This is due to the fact that, by fusing the received information, the fusion strategy propagates the information faster in the network. It is as if virtual links have been introduced between nodes through fusion. On the other hand, it lacks asymptotic accurate consensus and suffers from asymptotic memoryless behavior. As M increases, the overall impact of link errors is reduced, resulting in the better performance of the fusion case. The analysis and simulation of this part shows the underlying tradeoffs between fusion and diversity strategies. In practice, the network may only have a limited time (for instance on the order of 10s of iterations)

for reaching consensus. In such a case, fusion strategy may be more suitable. On the other hand, if the network can wait long enough, a diversity strategy can provide a better asymptotic performance. In practice, a combination of both strategies may provide the best overall performance.

4.4 Summary

In this chapter we studied a cooperative network that is trying to reach consensus on the occurrence of an event, by communicating over fading channels. We first considered a fully-connected and time-invariant network topology. We proposed a novel consensus-seeking strategy based on the Best Affine Estimation (BAE) of network state and characterized the asymptotic and transient behaviors of the network probabilistically. We showed that the network converges to a memoryless state asymptotically, which is undesirable. To see the transient behavior, we then characterized the rate of convergence by deriving an approximation for the second largest eigenvalue of the underlying average dynamical system.

We then extended the binary consensus scenario to the case of not fully-connected and rapidly-changing network topologies. We mathematically characterized the impact of fading, noise, network connectivity and time-varying topology on consensus performance, which becomes challenging due to all these uncertainties. Furthermore, we considered two different decision-making strategies, in term of using the available transmissions: fusion and diversity. We showed that the diversity strategy reaches asymptotic accurate consensus with probability of one. On the other hand, the fusion strategy provides a better transient behavior by propagating the information faster in the network. However, it lacks asymptotic accurate consensus.

Chapter 5

An integrated framework for binary consensus over time-invariant network topologies

In this chapter we utilize the aforementioned fusion and diversity decision-making strategies of Chapter 4 for binary consensus over the general *time-invariant network topologies* (not necessarily fully connected) with fading channels. We then propose an integrated framework that keeps the benefits of both fusion and diversity strategies, in terms of network information flow and link error robustness, for binary consensus over time-invariant network topologies with fading channels.

The chapter is organized as follows. In Section 5.1 we study the binary consensus over time-invariant network topologies with ideal communication links. In Section 5.2 we discuss the fusion and diversity decision-making strategies of Chapter 4 over time-invariant network topologies. We then propose an integrated diversity and fusion framework in Section 5.3 and mathematically analyze the proposed framework and show how the network achieves accurate consensus asymptotically. In Section 5.4 we then utilize the proposed

framework over regular ring lattice networks. We show the simulation results in Section 5.5. A summary of the results of the chapter is provided in Section 5.6.

5.1 Binary consensus over time-invariant topologies with ideal communication links

In this section, we revisit the binary consensus over not fully-connected and time-invariant network topologies. Furthermore, we assume ideal communication links. Studying this case allows us to focus solely on the impact of information flow in the network and characterize a benchmark for the performance of binary consensus algorithms over non-ideal communication links.

Consider the binary consensus scenario of Chapter 4. In this scenario, each node has its own initial decision, based on its one-time sensing. The goal of the network is for each node to reach a decision that is equal to the majority of the initial votes. In order to achieve this, each node will transmit its current decision to other nodes over a time invariant network topology. We model the underlying network as an undirected graph $G(\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1, \cdots, M\}$ represents the vertex set and $\mathcal{E}$ is the link set (the set of available communication links among the nodes).

Let $\tilde{b}_i^0 \in \{0, 1\}$ represent the initial vote of the ith node, where $\tilde{b}_i^0 = 1$ indicates that the ith agent initially decides that the event occurred whereas $\tilde{b}_i^0 = 0$ denotes otherwise. Consider the binary consensus over ideal communication links. In this scenario, each node receives the votes of its neighbors over the ideal communication links. It then fuses all the received information and updates its vote based on the majority of its neighbors' votes.

Let $\tilde{b}_i^u$ denote the ith node's vote after u fusion steps over ideal communication links. We have $\tilde{b}_i^u = \text{Dec}\Big(\frac{1}{1+|\mathcal{N}_i|}\big[\tilde{b}_i^{u-1} + \sum_{j\in\mathcal{N}_i} \tilde{b}_j^{u-1}\big]\Big)$ if $1/(1+|\mathcal{N}_i|)\big[\tilde{b}_i^{u-1} + \sum_{j\in\mathcal{N}_i} \tilde{b}_j^{u-1}\big] \neq$

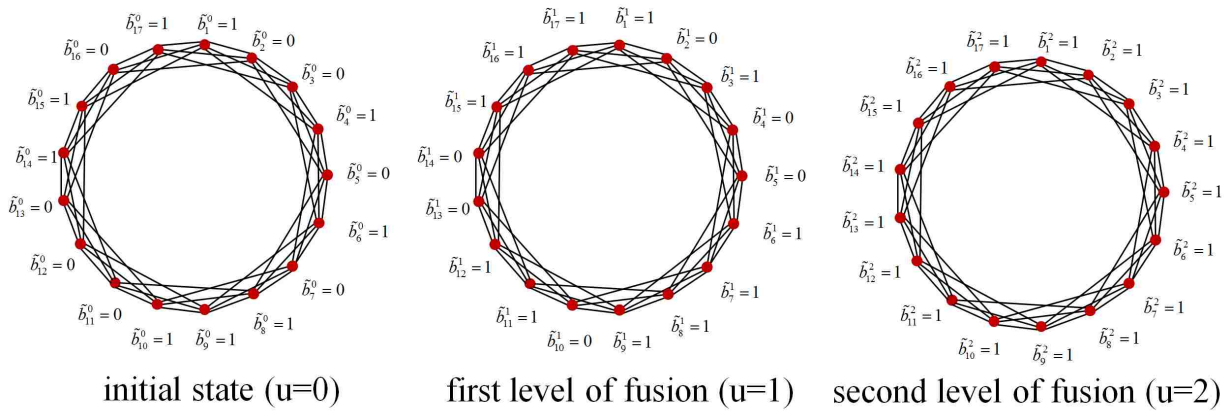

Figure 5.1: Demonstration of different fusion levels for $M = 17$ nodes trying to reach consensus over a 6-regular ring lattice.

0.5, where $\mathcal{N}_i$ denotes the neighbor set of the ith node and $\text{Dec}(x) = \begin{cases} 1 & x > 0.5 \\ 0 & x < 0.5 \end{cases}$. Otherwise $\tilde{b}_i^u = \tilde{b}_i^{u-1}$. Fig. 5.1 shows a network of $M = 17$ nodes trying to reach consensus over a 6-regular ring lattice.[1] In this topology each node receives information from its neighbors over ideal communication links. As can be seen, all the nodes reach the majority of initial votes at the second level of fusion. Let $D^u = \{\tilde{b}_1^u, \tilde{b}_2^u, \cdots, \tilde{b}_M^u\}$ denote the network state after u fusion steps. Define $u_G(D^0) \triangleq \min\{u | D^u \text{ is an accurate consensus state}\}$. For instance for the example of Fig. 5.1, we have $D^0 = \{1, 0, 0, 1, 0, 1, 0, 1, 1, 1, 0, 0, 0, 1, 1, 0, 1\}$ and $u_G(D^0) = 2$. We then define $\mathcal{D}_G \triangleq \{D^0 |\ u_G(D^0) < \infty\}$ and $u_G = \max_{D^0 \in \mathcal{D}_G} u_G(D^0)$. u_G is only a function of graph connectivity and represents the maximum number of required fusion levels in order to achieve accurate consensus if $D^0 \in \mathcal{D}_G$. For instance, for the fully connected graph, i.e., each node is connected to all the nodes in the network, we have $u_G = 1$. In the next section we consider the binary consensus over time-invariant network topologies with non-ideal communication links.

[1] An L-regular ring lattice is an L-regular graph with M vertices in a ring in which each vertex is connected to its L neighbors ($\frac{L}{2}$ on each side for an even value of L).

5.2 The underlying tradeoffs between information flow and robustness

Let $b_{j,i}(k)$ represent the reception of the ith node from the transmission of the jth one at the kth time step. We have the following if there exists a link from the jth node to the ith one: $b_{j,i}(k) = r_{j,i}(k)b_j(k) + n_{j,i}(k)$, where $b_j(k)$ denotes the jth node vote at time k and $r_{j,i}(k)$ and $n_{j,i}(k)$ are the fading coefficient and channel noise of the link from jth node to the ith one at time k respectively. Each agent will then update its assessment based on the received information.

We next discuss the fusion and diversity decision-making strategies of Chapter 4 for binary consensus over *not fully connected* and *time-invariant* network topologies before we propose our integrated approach. By not fully-connected and time-invariant network typologies with fading links, we mean networks where some links may not be connected. We assume that the set of connected links is fixed and therefore the underlying topology is time invariant. Similar to Chapter 4, we also assume that the nodes move fast enough such that the fading coefficients of the established links change and become uncorrelated from one time step to the next. However, their movements are limited to a small enough area such that the underlying network topology is not changing (thus a time-invariant network topology).

Fusion case: In Chapter 4, we proposed a fusion-based approach that utilizes information of link qualities and noise variance. Eq. 4.13 of Chapter 4 will be as follows for the case of time-invariant graphs:

$$b_i(k+1) = \text{Dec}\left(\frac{1}{1+|\mathcal{N}_i|}\Big[b_i(k) + \sum_{j\in\mathcal{N}_i}\Big(\alpha_{j,i}(k)b_{j,i}(k) + \beta_{j,i}(k)\Big)\Big]\right), \tag{5.1}$$

where $b_i(0) = \tilde{b}_i^0$ for $1 \leq i \leq M$. Similarly, the corresponding coefficients can be

represented as:

$$\alpha_{j,i}(k) = \frac{r_{j,i}(k)}{r_{j,i}^2(k) + \frac{\sigma_{j,i}^2}{q_j(k)(1-q_j(k))}} \text{ and } \beta_{j,i}(k) = \big(1 - \alpha_{j,i}(k) r_{j,i}(k)\big) q_j(k), \tag{5.2}$$

where $q_j(k) = \mathbf{E}[b_j(k)]$.

Diversity case: In Section 4.3.2, we proposed a diversity-based approach, where each node uses all of its transmissions to repeat its initial vote and only fuses the received information afterward. This strategy can, in particular, be useful in reducing the impact of link errors. Let $X_{j,i}(K) = [b_{j,i}(0), \cdots, b_{j,i}(K)]^T$ denote the receptions of ith node from the jth node in $K+1$ transmissions. Define $R_{j,i}(K) = [r_{j,i}(0), \cdots, r_{j,i}(K)]^T$ and $N_{j,i}(K) = [n_{j,i}(0), \cdots, n_{j,i}(K)]^T$. For the diversity case, all the nodes use their transmissions to repeat their initial votes, i.e., $b_j(k) = \tilde{b}_j^0$ for $0 \leq k \leq K$. Then we have the following, considering all the receptions of the ith node from the jth node: $X_{j,i}(K) = \tilde{b}_j^0 R_{j,i}(K) + N_{j,i}(K)$. The ith node can then estimate $\tilde{b}_j^0$ by using the best affine unbiased function of the received information: $\hat{\tilde{b}}_j^0(K) = \rho_{j,i}^T(K) X_{j,i}(K) + \eta_{j,i}(K)$. To ensure an unbiased estimator, we should have, $\mathbf{E}\big[\hat{\tilde{b}}_j^0(K)\big] = \mathbf{E}\big[\tilde{b}_j^0\big] \Rightarrow \eta_{j,i}(K) = \tilde{q}_j^0\big(1 - \rho_{j,i}^T(K) R_{j,i}(K)\big)$, where $\tilde{q}_j^0 = \mathbf{E}\big[\tilde{b}_i^0\big]$. Let $\zeta_{j,i}(K) = \mathbf{E}[\big(\hat{\tilde{b}}_j^0(K) - \tilde{b}_j^0\big)^2]$ denote the corresponding estimation error variance. We then have the following optimization problem:

$$\begin{aligned} \rho_{j,i}(K) &= \arg\ \min\ \zeta_{j,i}(K), \\ \text{subject to } \eta_{j,i}(K) &= \tilde{q}_j^0\big(1 - \rho_{j,i}^T(K) R_{j,i}(K)\big). \end{aligned} \tag{5.3}$$

We have

$$\begin{aligned} \zeta_{j,i}(K) &= E[\big(\hat{\tilde{b}}_j^0(K) - \tilde{b}_j^0\big)^2] \\ &= E\Big[\Big(\big(\tilde{b}_j^0 - \tilde{q}_j^0\big)\rho_{j,i}^T(K) R_{j,i}(K) - \big(b_j(0) - \tilde{q}_j^0\big) + \rho_{j,i}^T N_{j,i}(K)\Big)^2\Big] \\ &= \rho_{j,i}^T(K)\Big(\tilde{q}_j^0(1-\tilde{q}_j^0) R_{j,i}(K) R_{j,i}^T(K) + \sigma_{j,i}^2 I_{K\times K}\Big)\rho_{j,i}(K) \\ &\quad - 2\tilde{q}_j^0(1-\tilde{q}_j^0)\rho_{j,i}^T(K) R_{j,i}(K) + \tilde{q}_j^0(1-\tilde{q}_j^0). \end{aligned} \tag{5.4}$$

By noting that 5.4 is a convex function of $\rho_{j,i}(K)$, we have

$$\begin{aligned}\rho_{j,i}(K) &= \frac{\tilde{q}_j^0(1-\tilde{q}_j^0)}{\sigma_{j,i}^2}\Big(\frac{\tilde{q}_j^0(1-\tilde{q}_j^0)}{\sigma_{j,i}^2}R_{j,i}(K)R_{j,i}^T(K)+I_{K\times K}\Big)^{-1}R_{j,i}(K)\\ &= \frac{\tilde{q}_j^0(1-\tilde{q}_j^0)}{\sigma_{j,i}^2}\Big(I_{K\times K}-\frac{\frac{\tilde{q}_j^0(1-\tilde{q}_j^0)}{\sigma_{j,i}^2}}{1+\frac{\tilde{q}_j^0(1-\tilde{q}_j^0)}{\sigma_{j,i}^2}\|R_{j,i}(K)\|^2}R_{j,i}(K)R_{j,i}^T(K)\Big)R_{j,i}(K)\\ &= \frac{1}{\frac{\sigma_{j,i}^2}{\tilde{q}_j^0(1-\tilde{q}_j^0)}+\|R_{j,i}(K)\|^2}R_{j,i}(K),\end{aligned}$$

and $\eta_{j,i}(K)=\tilde{q}_j^0\big(1-\rho_{j,i}^T(K)R_{j,i}(K)\big)$. We then have the following for the estimation error variance after $K+1$ transmissions: $\zeta_{j,i}(K)=\frac{\sigma_{j,i}^2}{\frac{\sigma_{j,i}^2}{\tilde{q}_j^0(1-\tilde{q}_j^0)}+\|R_{j,i}(K)\|^2}=\frac{1}{\frac{1}{\tilde{q}_j^0(1-\tilde{q}_j^0)}+\sum_{k=0}^{K}\mathrm{CNR}_{j,i}(k)}$, where $\mathrm{CNR}_{j,i}(k)=\frac{r_{j,i}^2(k)}{\sigma_{j,i}^2}$ denotes the channel to noise ratio for the link, from node j to node i at time k. We then have the following decision making function:

$$\begin{aligned}b_i(K+1) &= \mathrm{Dec}\left(\frac{1}{1+|\mathcal{N}_i|}\Big(b_i(0)+\sum_{j\in\mathcal{N}_i}\hat{\tilde{b}}_j^0(K)\Big)\right)\\ &= \mathrm{Dec}\left(\frac{1}{1+|\mathcal{N}_i|}\Big(b_i(0)+\sum_{j\in\mathcal{N}_i}\rho_{j,i}^T(K)X_{j,i}(K)+\eta_{j,i}(K)\Big)\right),\end{aligned}\tag{5.5}$$

where $b_i(k)=\tilde{b}_i^0$ for $0\le k\le K$.

Lemma 1. *Consider binary consensus over a not fully connected network topology. Under the assumption of i.i.d. Rayleigh fading channels, $CNR_{j,i}(k)$ are i.i.d. exponential random variables with $\overline{CNR}=\mathbf{E}\{CNR_{j,i}(k)\}$. Then the dynamical system, based on the decision-making function of Eq. 5.5, will asymptotically reach $D^1=\{\tilde{b}_1^1,\tilde{b}_2^1,\cdots,\tilde{b}_M^1\}$, where D^1 represents the network state after one level of fusion over ideal communication links.*

Proof. We have,

$$\begin{aligned}\text{prob}\Big(b_i(K+1) \neq \tilde{b}_i^1\Big) &\leq \text{prob}\Big(\Big|b_i(0) + \sum_{j\in\mathcal{N}_i} \hat{\tilde{b}}_j^0(K) - \sum_{j\in\mathcal{N}_i \bigcup\{i\}} \tilde{b}_j^0\Big| \geq \frac{1}{2}\Big) \\ &= \text{prob}\Big(\Big|\sum_{j\in\mathcal{N}_i} \big[\hat{\tilde{b}}_j^0(K) - \tilde{b}_j^0\big]\Big| \geq \frac{1}{2}\Big) \\ &\leq 4\sum_{j\in\mathcal{N}_i} E\Big[\big(\hat{\tilde{b}}_j^0(K) - \tilde{b}_j^0\big)^2\Big] = 4\sum_{j\in\mathcal{N}_i} \zeta_{j,i}(K),\end{aligned}$$

where in the last line we are using Chebyshev's inequality. Furthermore, we have,

$$\zeta_{j,i}(K) = \frac{1}{\frac{1}{\tilde{q}_j^0(1-\tilde{q}_j^0)} + \sum_{k=0}^{K} \text{CNR}_{j,i}(k)} < \gamma_{j,i}^{-1}(K) \tag{5.6}$$

where $\gamma_{j,i}(K) \triangleq \sum_{k=0}^{K} \text{CNR}_{j,i}(k)$. Since $\text{CNR}_{j,i}(k)$s are i.i.d. exponential random variables with the mean of $\overline{\text{CNR}}$, $\gamma_{j,i}(K)$ has Gamma distribution with the following parameters: $\gamma_{j,i}(K) \sim \textit{Gamma}(K+1, \overline{\text{CNR}})$. As a result, $\gamma_{j,i}^{-1}(K)$ is an Inverse Gamma random variable. We then have, $\lim_{K\to\infty} \mathbf{E}\big\{\gamma_{j,i}^{-1}(K)\big\} = \lim_{K\to\infty} \frac{1}{(K+1)\times\overline{\text{CNR}}} = 0$. Since $\gamma_{j,i}^{-1}(K)$ is a non-negative random variable, we have

$$\lim_{K\to\infty} \frac{1}{\gamma_{j,i}(K)} = 0. \tag{5.7}$$

Therefore, $\lim_{K\to\infty} \zeta_{j,i}(K) = 0$, which results in $\lim_{K\to\infty} \text{prob}\Big(b_i(K+1) = \tilde{b}_i^1\Big) = 1$ for $1 \leq i \leq M$. □

In Fig. 4.7 of Chapter 4, we compared the performance of both fusion and diversity strategies for binary consensus over rapidly-changing network topologies with fading links. We next confirm that similar tradeoffs exist for the case of not fully-connected but time-invariant network topologies. Similar behaviors are observed. For instance, Fig. 5.2 shows the comparison for a 16-regular ring topology, i.e. a fully connected graph, where the transmissions occur over Rayleigh channels with $\overline{\text{CNR}} = -6$dB. Since the main bottleneck is the link quality, the diversity approach outperforms the fusion approach and reaches accurate consensus asymptotically, as expected. Fig. 5.3 compares the performance of both approaches for 6-regular ring topology and $\overline{\text{CNR}} = 6$dB. Similar to the results of Chapter 4, since the main bottleneck is the network connectivity, the fusion

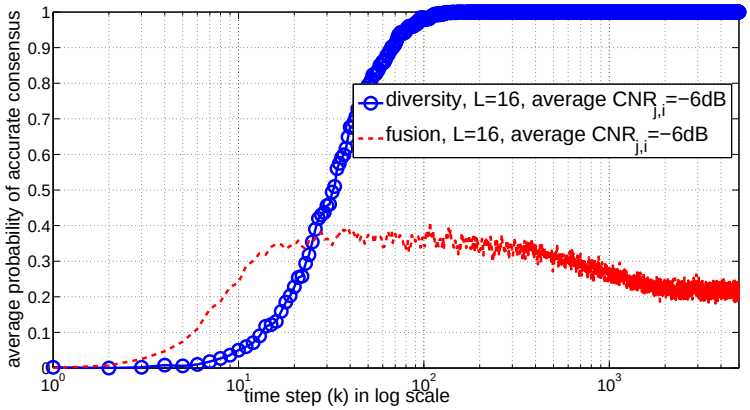

Figure 5.2: Performance comparison of fusion and diversity approaches for binary consensus over 16-regular ring lattice topology of $M = 17$ nodes with fading channels of $\overline{\text{CNR}} = -6$dB(averaging is done over several runs).

approach outperforms the diversity one drastically. In the next section we propose an integrated framework that keeps the benefits of both fusion and diversity approaches.

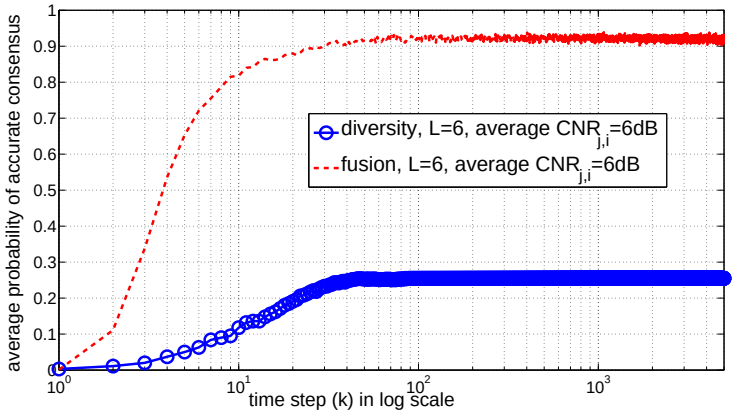

Figure 5.3: Performance comparison of fusion and diversity approaches for binary consensus over 6-regular ring lattice topology of $M = 17$ nodes with fading channels of $\overline{\text{CNR}} = 6$dB(averaging is done over several runs).

5.3 An integrated diversity-fusion framework for binary consensus over fading channels

In this section we propose our integrated diversity-fusion framework for binary consensus over fading channels. In this strategy, each agent sends a vector to its neighbors. This vector consists of the estimations of the votes corresponding to the different fusion levels. Throughout the repeated communications (diversity), each node tries to refine its assessments of different fusion levels in order to reach consensus.

Let $B_i(k) = [b_i^0(k), b_i^1(k), \cdots, b_i^{l_G(k)}(k)]^T$ represent the vector that node i will send to all its neighbors over fading channels at time k, where $b_i^u(k)$ represents the ith node estimate of $\tilde{b}_i^u$ at time k and $l_G(k) = \min\{k, u_G - 1\}$. Table 5.1 shows the time progression of the transmitted vector by node i to all its neighbors. As can be seen, at time $t = 0$, node i only transmits its initial vote to its neighbors. It also receives its neighbors' initial votes over fading channels. Next, in time step $t = 1$, node i fuses all its receptions in order to come up with an estimate of $\tilde{b}_i^1$. It then transmits its initial vote together with the estimate of $\tilde{b}_i^1$ to all its neighbors. This process will go on for a while and each agent sends its estimate of different fusion levels to it neighbors. As can be seen, the ith node's estimate of $\tilde{b}_i^u$ will not be available till $k = u$. Therefore, for $k < u_G - 1$, $B_i(k) = [b_i^0(k), b_i^1(k), \cdots, b_i^k(k)]^T$ and $l_G(k) = k$, however, for $t \geq u_G - 1$ the length of the transmitted vector becomes fixed and equal to u_G.

Table 5.1: Time progression of transmitted vector by node i

$t=0$	$t=1$	$\cdots$	$t=u_G-1$	$\cdots$	$t=k$
$b_i^0(0)$	$b_i^0(1)$	$\cdots$	$b_i^0(u_G-1)$	$\cdots$	$b_i^0(k)$
	$b_i^1(1)$	$\cdots$	$b_i^1(u_G-1)$	$\cdots$	$b_i^1(k)$
		$\cdots$	$\vdots$	$\cdots$	$\vdots$
		$\cdots$	$b_i^{u_G-1}(u_G-1)$	$\cdots$	$b_i^{u_G-1}(k)$

Let $r_{j,i}^u(k)$ and $n_{j,i}^u(k)$ represent the fading coefficient and the receiver noise of the link,

which transmits $b_j^u(k)$ from node j to node i respectively. $n_{j,i}^u(k)$ is zero-mean Gaussian with the variance of $\sigma_{j,i,u}^2$. We take the fading coefficient and the receiver noise of all the receptions to be uncorrelated. Let $b_{j,i}^u(k)$ represent the reception of the ith node from the transmission of $b_j^u(k)$ at time k. We have the following, $b_{j,i}^u(k) = r_{j,i}^u(k)b_j^u(k) + n_{j,i}^u(k)$ for $1 \leq i \leq M$ and $j \in \mathcal{N}_i$. Let $X_{j,i}^u(k) = [b_{j,i}^u(u), \cdots, b_{j,i}^u(k)]^T$ represent all the receptions of node i from node j corresponding to the uth level of fusion till time k. Node i will then update its estimate of b_i^{u+1} at time $k+1$ based on $b_i^u(k)$ and all the uth level votes of its neighbors until time k.

$$b_i^{u+1}(k+1) = \mathcal{F}\Big(b_i^u(k), X_{j,i}^u(k)\ \forall j \in \mathcal{N}_i\Big), \tag{5.8}$$

where $\mathcal{F}$ represents a decision-making function. The redundancy in vector $X_{j,i}^u(k)$ tries to improve the link qualities. We next show how to design this function. Define the following variables, $\tilde{B}_j^u(k) = [b_j^u(u), \cdots, b_j^u(k)]^T$, $R_{j,i}^u(k) = [r_{j,i}^u(u), \cdots, r_{j,i}^u(k)]^T$, $H_{j,i}^u(k) = \text{diag}\big(R_{j,i}^u(k)\big)$, $N_{j,i}^u(k) = [n_{j,i}^u(u), \cdots, n_{j,i}^u(k)]^T$, where $N_{j,i}^u(k) \sim \mathcal{N}\big(0, \sigma_{j,i,u}^2 I_{(k-u+1)\times(k-u+1)}\big)$ and diag$(.)$ is a diagonal matrix with the elements of the argument on its main diagonal. Therefore, we have $X_{j,i}^u(k) = H_{j,i}^u(k)\tilde{B}_j^u(k) + N_{j,i}^u(k)$ for $k \geq u$. The ith node will then try to estimate $\tilde{b}_j^u$, based on all receptions from jth node, i.e. $X_{j,i}^u(k)$. In order to characterize the best affine estimation of $\tilde{b}_j^u$, the second-order statistic of $\tilde{B}_j^u(k)$ is required. However, finding a closed-form expression for the second-order statistics of this variable is challenging. The vector $\tilde{B}_j^u(k)$ contains different temporal assessments of the jth node about $\tilde{b}_j^u$. To simplify the mathematical derivation, we consider the case, where the uth level of fusion is in its steady state, i.e. $\tilde{B}_j^u(k) = \tilde{b}_j^u \vec{1}_{k-u+1}\ \ \forall j$. Under this assumption the BAE estimation can be derived similar to Eq. 5.3 as follows: $\hat{\tilde{b}}_j^u(k) = {\rho_{j,i}^u}^T(k)X_{j,i}^u(k) + \eta_{j,i}^u(k)$, where $\rho_{j,i}^u(k) = \frac{1}{\frac{\sigma_{j,i,u}^2}{\tilde{q}_j^u(1-\tilde{q}_j^u)} + \|R_{j,i}^u(k)\|^2} R_{j,i}^u(k)$, $\eta_{j,i}^u(k) = \tilde{q}_j^u\big(1 - {\rho_{j,i}^u}^T(k)R_{j,i}^u(K)\big)$ and $\tilde{q}_j^u = \mathbf{E}[\tilde{b}_j^u]$. Since $\tilde{B}_j^u(k) \neq \tilde{b}_j^u \vec{1}_{k-u+1}$, the estimator is suboptimal. However, in Theorem 1 we will show that $\lim_{k\to\infty} b_j^u(k) = \tilde{b}_j^u$ for $1 \leq j \leq M$. Therefore, the proposed suboptimal estimator is asymptotically optimal. The ith node will then update its $(u+1)$th level decision

as:

$$b_i^{u+1}(k+1) = \text{Dec}\Big(\frac{1}{|\mathcal{N}_i|+1}\big[b_i^u(k) + \sum_{j\in\mathcal{N}_i}\hat{\tilde{b}}_j^u(k)\big]\Big)$$
$$= \text{Dec}\Big(\frac{1}{|\mathcal{N}_i|+1}\Big[b_i^u(k) + \sum_{j\in\mathcal{N}_i}{\rho_{j,i}^u}^T(k)X_{j,i}^u(k) + \eta_{j,i}^u(k)\Big]\Big), \qquad (5.9)$$

for $k \geq u$, $0 \leq u \leq u_G - 1$ and $b_i^0(k) = \tilde{b}_i^0\ \forall i, k$. Next we show that the decision-making function of Eq. 5.9 achieves accurate consensus asymptotically.

Theorem 1. *Consider binary consensus over a time-invariant network topology with i.i.d. Rayleigh fading channels. Assume $|\mathcal{N}_i|$ is even for $1 \leq i \leq M$. Then, the decision-making function of Eq. 5.9 asymptotically converges (in probability) to accurate consensus iff $D^0 \in \mathcal{D}_G$.*

Proof. We prove the theorem by induction. Define $\omega_i^u(k) \triangleq \text{prob}\Big(b_i^u(k) \neq \tilde{b}_i^u\Big)$. For $u = 0$, we have $b_i^0(k) = \tilde{b}_i^0 \quad \forall k, i$. From Lemma 1, it can be easily confirmed that $\lim_{k\to\infty}\omega_i^1(k) = 0\ \ \forall i$. Assume $\lim_{k\to\infty}\omega_i^u(k) = 0$. We next prove that $\lim_{k\to\infty}\omega_i^{u+1}(k+1) = 0$. We have,

$$\omega_i^{u+1}(k+1) = \text{prob}\Big(b_i^{u+1}(k+1) \neq \tilde{b}_i^{u+1}\Big)$$
$$\leq \text{prob}\Big(\Big|b_i^u(k) + \sum_{j\in\mathcal{N}_i}{\rho_{j,i}^u}^T(k)X_{j,i}^u(k) + \eta_{j,i}^u(k) - \sum_{j\in\mathcal{N}_i\bigcup\{i\}}\tilde{b}_j^u\Big| \geq \frac{1}{2}\Big)$$
$$= \text{prob}\Bigg(\Bigg|b_i^u(k) - \tilde{b}_i^u + \sum_{j\in\mathcal{N}_i}\frac{1}{\frac{\sigma_{j,i,u}^2}{\tilde{q}_j^u(1-\tilde{q}_j^u)} + \|R_{j,i}^u(k)\|^2}$$
$$\times\Bigg(\sum_{t=u}^{k}{r_{j,i}^u}^2(t)\Big(b_j^u(t) - \tilde{b}_j^u\Big) + \frac{\sigma_{j,i,u}^2}{1-\tilde{q}_j^u}(1 - \frac{\tilde{b}_j^u}{\tilde{q}_j^u}) + {R_{j,i}^u}^T(k)N_{j,i}^u(k)\Bigg)\Bigg| \geq \frac{1}{2}\Bigg)$$
$$\leq \text{prob}\Bigg(\Big|b_i^u(k) - \tilde{b}_i^u\Big| + \sum_{j\in\mathcal{N}_i}\frac{1}{\|R_{j,i}^u(k)\|^2} \times \Bigg(\sum_{t=u}^{k}{r_{j,i}^u}^2(t)\Big|b_j^u(t) - \tilde{b}_j^u\Big|$$
$$+ \frac{\sigma_{j,i,u}^2}{1-\tilde{q}_j^u}\Big|1 - \frac{\tilde{b}_j^u}{\tilde{q}_j^u}\Big| + \Big|{R_{j,i}^u}^T(k)N_{j,i}^u(k)\Big|\Bigg) \geq \frac{1}{2}\Bigg),$$

where in the last inequality, we are using the property that for any random variables Ξ_i,

we have $\text{prob}\Big(\big|\sum_i \Xi_i\big| \geq \epsilon\Big) \leq \text{prob}\Big(\sum_i |\Xi_i| \geq \epsilon\Big)$. Furthermore, we have $\mathbf{E}\big\{\big|b_i^u(k) - \tilde{b}_i^u\big|\big\} = \omega_i^u(k)$. Applying Markov's inequality results in:

$$
\begin{aligned}
&\omega_i^{u+1}(k+1) \\
&\leq 2\Bigg[\omega_i^u(k) + \sum_{j\in\mathcal{N}_i} \frac{1}{\|R_{j,i}^u(k)\|^2} \times \left(\sum_{t=u}^{k} {r_{j,i}^u}^2(t)\omega_j^u(k) + \frac{\sigma_{j,i,u}^2}{1-\tilde{q}_j^u}\Big|1 - \frac{\tilde{b}_j^u}{\tilde{q}_j^u}\Big| + \mathbf{E}\Big\{\Big|{R_{j,i}^u}^T(k)N_{j,i}^u(k)\Big|\Big\}\right)\Bigg] \\
&= 2\Bigg[\omega_i^u(k) + \sum_{j\in\mathcal{N}_i} \frac{1}{\|R_{j,i}^u(k)\|^2} \left(\sum_{t=u}^{k} {r_{j,i}^u}^2(t)\omega_j^u(k) + \frac{\sigma_{j,i,u}^2}{1-\tilde{q}_j^u}\Big|1 - \frac{\tilde{b}_j^u}{\tilde{q}_j^u}\Big|\right) + \sqrt{\frac{2}{\pi}} \sum_{j\in\mathcal{N}_i} \frac{\sigma_{j,i,u}}{\|R_{j,i}^u(k)\|}\Bigg] \\
&= 2\Bigg[\omega_i^u(k) + \sum_{j\in\mathcal{N}_i} \frac{1}{\|R_{j,i}^u(k)\|^2} \sum_{t=u}^{k} {r_{j,i}^u}^2(t)\omega_j^u(k) + \sum_{j\in\mathcal{N}_i} \frac{\Big|1 - \frac{\tilde{b}_j^u}{\tilde{q}_j^u}\Big|}{1-\tilde{q}_j^u} \frac{1}{\gamma_{j,i}^u(k)} + \sqrt{\frac{2}{\pi}} \sum_{j\in\mathcal{N}_i} \frac{1}{\sqrt{\gamma_{j,i}^u(k)}}\Bigg],
\end{aligned}
\tag{5.10}
$$

where $\gamma_{j,i}^u(k) = \sum_{t=u}^{k} \text{CNR}_{j,i}^u(t) = \frac{\|R_{j,i}^u(k)\|^2}{\sigma_{j,i,u}^2}$. Let $\epsilon > 0$. Since $\lim_{k\to\infty} \omega_j^u(k) = 0$, there exists a number $T_j^u(\epsilon)$ such that $\omega_j^u(k) < \epsilon$ for all $k \geq T_j^u(\epsilon)$. Therefore,

$$
\begin{aligned}
\frac{1}{\|R_{j,i}^u(k)\|^2} \sum_{t=u}^{k} {r_{j,i}^u}^2(t)\omega_j^u(t) &\leq \frac{1}{\|R_{j,i}^u(k)\|^2} \sum_{t=u}^{T_j^u(\epsilon)} {r_{j,i}^u}^2(t)\omega_j^u(t) + \frac{\|R_{j,i}^u(k)\|^2 - \|R_{j,i}^u(T_j^u(\epsilon))\|^2}{\|R_{j,i}^u(k)\|^2}\epsilon \\
&\leq \frac{1}{\|R_{j,i}^u(k)\|^2} \sum_{t=u}^{T_j^u(\epsilon)} {r_{j,i}^u}^2(t)\omega_j^u(t) + \epsilon.
\end{aligned}
\tag{5.11}
$$

Since $r_{j,i}^u(t)$'s are i.i.d. random variables, we have

$$
\lim_{k\to\infty} \mathbf{E}\left\{\frac{\sum_{t=u}^{T_j^u(\epsilon)} {r_{j,i}^u}^2(t)}{\|R_{j,i}^u(k)\|^2}\right\} = \lim_{k\to\infty} \frac{T_j^u(\epsilon) - u + 1}{k - u + 1} = 0. \tag{5.12}
$$

Since $\frac{1}{\|R_{j,i}^u(k)\|^2} \sum_{t=u}^{T_j^u(\epsilon)} {r_{j,i}^u}^2(t)$ is a non-negative random variable, it goes to 0 as $k \to \infty$. Therefore, we have, $\lim_{k\to\infty} \frac{1}{\|R_{j,i}^u(k)\|^2} \sum_{t=u}^{k} {r_{j,i}^u}^2(t)\omega_j^u(t) = 0$. Furthermore, similar to Eq. 5.7, we can show that $\lim_{k\to\infty} \frac{1}{\gamma_{j,i}^u(k)} = 0$. By substituting these values in Eq. 5.10, it can be easily confirmed that $\lim_{k\to\infty} \omega_i^{u+1}(k+1) = 0$. Using induction, we have $\lim_{k\to\infty} \omega_i^{u_G}(k) = 0$ and as a result $\lim_{k\to\infty} \text{prob}\Big(b_i^{u_G}(k) = \tilde{b}_i^{u_G}\Big) = 1$ for $1 \leq i \leq M$. Therefore, if $D^0 \in \mathcal{D}_G$, then $\lim_{k\to\infty}[b_1^{u_G}(k), \cdots, b_M^{u_G}(k)]$ is an accurate consensus state with the probability of one, which proves the theorem. □

Remark 1. *The decision-making function of Eq. 5.9 uses information of $\sigma_{j,i,u}$s. If such*

information is not available, then the ith node can estimate $\tilde{b}_j^u$ using Best Linear Unbiased Estimation (BLUE): $\hat{\tilde{b}}_j^u(k) = \frac{1}{\|R_{j,i}^u(k)\|^2} {R_{j,i}^u}^T(k) X_{j,i}^u(k)$, assuming that the uth level of fusion is in its steady state. Each node will then update its $(u+1)$th level decision as follows:

$$b_i^{u+1}(k+1) = Dec\Big(\frac{1}{|\mathcal{N}_i|+1}\big[b_i^u(k) + \sum_{j\in\mathcal{N}_i} \hat{\tilde{b}}_j^u(k)\big]\Big)$$
$$= Dec\Big(\frac{1}{|\mathcal{N}_i|+1}\Big[b_i^u(k) + \sum_{j\in\mathcal{N}_i} \frac{1}{\|R_{j,i}^u(k)\|^2} {R_{j,i}^u}^T(k) X_{j,i}^u(k)\Big]\Big), \qquad (5.13)$$

for $k \geq u$, $0 \leq u \leq u_G - 1$ and $b_i^0(k) = \tilde{b}_i^0 \ \forall i, k$. This receiver can be considered as a special case of the decision-making function of Eq. 5.9, where $\sigma_{j,i,u} = 0$. Therefore, it achieves accurate consensus asymptotically. Furthermore, we have $\lim_{k\to\infty} \rho_{j,i}^u(k) \approx \frac{1}{\|R_{j,i}^u(k)\|^2} R_{j,i}^u(k)$ and $\lim_{k\to\infty} \eta_{j,i}^u(k) \approx 0$. Therefore, the decision-making function of Eq. 5.9 will be boiled down to the decision-making function of Eq. 5.13 for enough large k.

In the subsequent sections, we utilize the integrated framework for a special class of undirected graphs and show the performance of the proposed framework.

5.4 Integrated framework over regular ring lattice topologies

In Section 5.3, we introduced our proposed framework, which asymptotically achieves accurate consensus. In this approach, node i will send a vector $B_i(k)$ to all its neighbors over fading channels. The length of this vector is u_G for $t \geq u_G - 1$, which is a function of graph connectivity. Intuitively, networks with higher connectivity require smaller values of u_G. For instance, for fully connected networks, we have $u_G = 1$. In general, u_G is a function of network topology and independent of communication quality. This parameter needs to be determined before running the algorithm. In this section, we mainly focus on L-regular ring lattice topologies with ideal communication links in order to characterize u_G. Let

$\mathcal{V} = \{1, 2, \cdots, M\}$ denote the vertex set. Without loss of generality, we assume that the vertices are ordered clockwise on the ring (see Fig. 5.1). Furthermore, we assume that M is odd. Under the above assumptions, the adjacency matrix of an L-regular ring lattice, i.e., A_L, can be represented by a circulant matrix with the first row of $[0, \vec{1}^T_{\frac{L}{2}}, \vec{0}^T_{M-L-1}, \vec{1}^T_{\frac{L}{2}}]$. Therefore, the dynamic of the network state at different fusion levels can be expressed as follows: $D^u = \text{Dec}\Big(\frac{1}{L+1}(A_L + I_{M\times M}) \times D^{u-1}\Big)$, where A_L denotes the adjacency matrix and Dec acts entry-wise on its argument vector. For this class of graphs, L is a notion of connectivity. Therefore, we try to show how u_G changes as a function of L. We then have the following lemma, which will be used in Theorem 2.

Lemma 2. *For an L-regular ring lattice, if $L = M - 1$, then $\Big|\cap_{i=1}^{M}\big\{\mathcal{N}_i \cup \{i\}\big\}\Big| = M$. Moreover, if $L \leq M-3$, then $\Big|\cap_{n=1}^{p}\big\{\mathcal{N}_{i_n} \cup \{i_n\}\big\}\Big| \leq L+2-p$. The equality is achieved if and only if $\{i_1, \cdots, i_p\} \subset \mathcal{V}$ denotes a set of consecutive nodes on the corresponding ring.*

Proof. The proof is straightforward and we skip it. □

Let $m^u = \min\Big(\sum_{j=1}^{M}\tilde{b}^u_j, M - \sum_{j=1}^{M}\tilde{b}^u_j\Big)$ represent the number of nodes, with the minority of the votes at uth level of fusion. We then have the following theorem.

Theorem 2. *Consider binary consensus over an L-regular ring lattice, we then have the following properties:*

1- Assume $\mathcal{V} = \bigcup_{i=1}^{c}\mathcal{V}_i$, such that $\mathcal{V}_i$ represents a set of consecutive nodes on the ring that vote the same and $|\mathcal{V}_i| \geq \frac{L}{2}+1$. Then the corresponding network state is an absorbing state.

2- For $m^0 \leq \frac{L}{2}$, the accurate consensus is achievable after one level of fusion.

3- For $m^0 = \frac{L}{2}+1$ and $D^0 \in \mathcal{D}_G$, accurate consensus is achievable at most after two fusion steps.

Proof. Let $\mathcal{V}_i = \{o_1^i, \cdots, o_{|\mathcal{V}_i|}^i\}$ denote the ith partition, where o_j^i represents the index of jth node in the ith partition. Therefore, we have, $\left|\left\{\mathcal{N}_{o_j^i} \cup \{o_j^i\}\right\} \bigcap \mathcal{V}_i\right| \geq \frac{L}{2} + 1$ for all $1 \leq j \leq |\mathcal{V}_i|$. Since all the nodes in $\mathcal{V}_i$ vote the same, we have $\tilde{b}_{o_j^i}^{u+1} = \tilde{b}_{o_j^i}^{u}$ for $1 \leq i \leq c$ and $1 \leq j \leq |\mathcal{V}_i|$.

We next prove the second part. Consider the case where the majority of the initial votes is 1. If $m^0 \leq \frac{L}{2}$, we then have $\frac{\tilde{b}_i^0 + \sum_{j \in \mathcal{N}_i} \tilde{b}_j^0}{L+1} \geq \frac{L+1-m^0}{L+1} > \frac{1}{2} \; \forall \; i$, which results in $\tilde{b}_i^1 = 1$. For the case, where the majority of the initial votes is 0, we have $\frac{\tilde{b}_i^0 + \sum_{j \in \mathcal{N}_i} \tilde{b}_j^0}{L+1} \leq \frac{m^0}{L+1} < \frac{1}{2} \; \forall \; i$, which results in $\tilde{b}_i^1 = 0$. Therefore, for $m^0 \leq \frac{L}{2}$ accurate consensus is achievable in one iteration.

Next, we show the third statement. First we show that for $m^0 = \frac{L}{2} + 1$ and $D^0 \in \mathcal{D}_G$, we have $m^1 \leq \frac{L}{2}$. Let $\left\{i_1^0, i_2^0, \cdots, i_{m^0}^0\right\}$ denote an ordered set of the nodes, which vote to the minority of the initial votes. Lemma 2 says that at most $L + 2 - m^0 = \frac{L}{2} + 1$ nodes can have $\{i_1^0, \cdots, i_{m^0}^0\}$ in their neighbor set if and only if $\left\{i_1^0, i_2^0, \cdots, i_{m^0}^0\right\}$ is a set of consecutive vertices. Therefore, $m^1 = \frac{L}{2} + 1$ is achievable if and only if the initial state is an absorbing state (see Theorem 2-1). Therefore, if $D^0 \in \mathcal{D}_G$, then $m^1 \leq \frac{L}{2}$, which reaches accurate consensus in $u = 2$ (see Theorem 2-2). □

In Theorem 2, we showed that the number of fusion levels, required to achieve accurate consensus is a function of graph connectivity (L) and initial state (D^0). For instance, for a fixed connectivity, if $m^0 \leq \frac{L}{2}$, accurate consensus is achievable after one level of fusion. However, if m^0 increases to $\frac{L}{2} + 1$, then the network may require two fusion steps to reach consensus. In order to understand the impact of the connectivity on the required fusion steps, Fig. 5.4 characterizes u_G as a function of L over a regular ring lattice topology of $M = 17$ nodes. As can be seen, as L increases, u_G decreases. For instance, for $L = 16$, i.e., fully connected graph, the network requires only one level of fusion to reach accurate consensus. It can also be seen that u_G is only a function of the graph topology. In practical applications, this parameter needs to be determined before running the integrated algorithm.

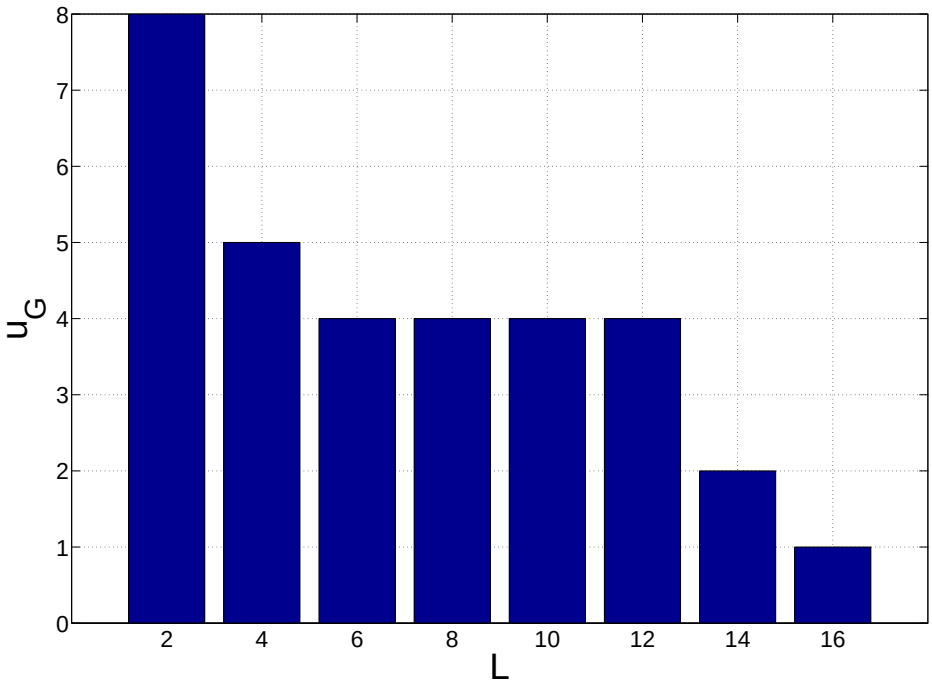

Figure 5.4: u_G as a function of L for the regular ring lattice of $M = 17$ nodes

5.5 Simulation and comparison

Fig. 5.5 shows the performance of the proposed framework over a regular ring lattice topology of $M = 17$ nodes. In order to show the impact of network connectivity on the performance of the proposed algorithm, we consider a regular graph topology with different values of L. For each L, the number of required fusion levels, i.e., u_G, can be found from Fig. 5.4. Each node will then send a binary vector of length u_G to all its neighbors over communication channels. Moreover, we assume all channels experience the same noise variance ($\sigma_{j,i,u} = 1.5$) and the average power of fading coefficients is equal to one ($E\left[{r_{j,i}^u}^2(k)\right] = 1$). In order to make a fair comparison, we keep the number of transmitted bits fixed. Therefore, in Fig. 5.5, we show the probability of accurate consensus as a function of number of transmitted bits per node.

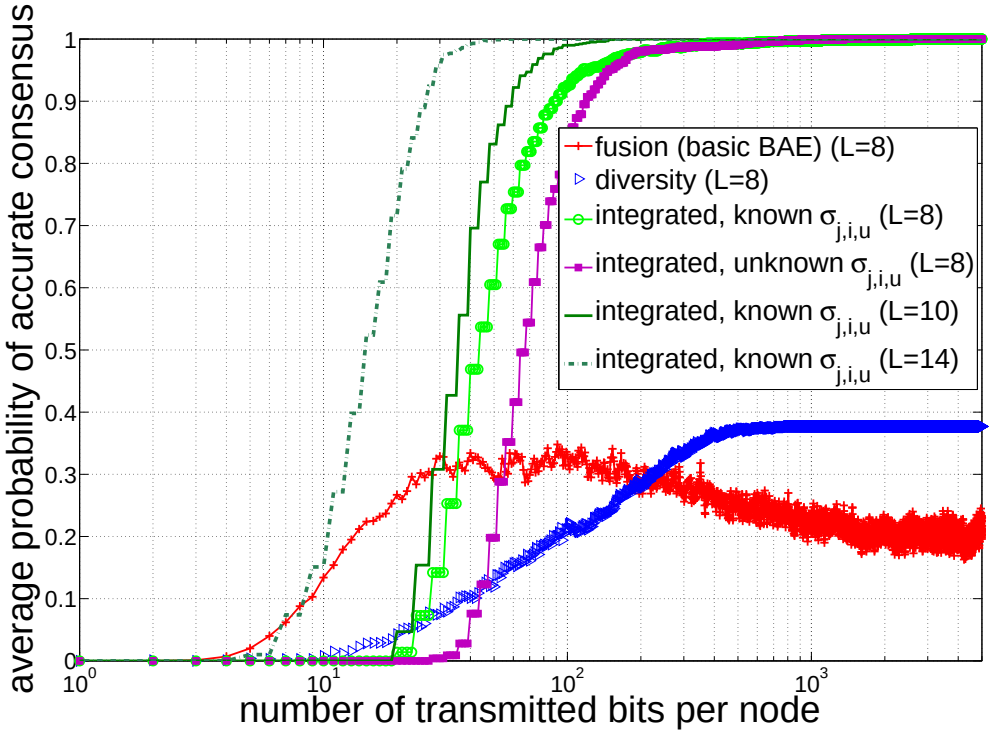

Figure 5.5: average probability of accurate consensus over regular ring lattice of 17 nodes, where $\overline{\text{CNR}} = -3.5$dB

Fig. 5.5 shows the performance of the proposed framework for $L = 8, 10$ and 14. For these simulations, we assume that node i does not have the knowledge of q_j^u for $j \in \mathcal{N}_i$, so it simply assumes $q_j^u = 0.5$ in the decision making function of Eq. 5.9. As can be seen, the integrated approach, independent of network connectivity, achieves accurate consensus asymptotically. Furthermore, networks with higher connectivity, i.e., larger values of L, reach their steady state in fewer transmissions. For the case of $L = 8$, the performance of both fusion and diversity approaches of Section 5.2 are also shown for comparison. As can be seen, the proposed approach keeps the benefits of both fusion and diversity in terms of the transient and asymptotic behaviors respectively. Furthermore, the performance of the integrated approach for the case where $L = 8$ and knowledge of $\sigma_{j,i,u}$ is not available (Remark 1) is also shown in Fig. 5.5. As can be seen, the integrated approach with known

link quality slightly outperforms the case of unknown link quality. However, as mentioned earlier in Remark 1, as time goes by, both cases provide similar performance.

5.6 Summary

In this section we considered a cooperative network that is trying to reach consensus over not fully-connected and time-invariant network topologies with fading channels. We utilized the aforementioned fusion and diversity decision-making strategies of Chapter 4 for binary consensus over the general time-invariant network topologies (not necessarily fully connected) with fading channels. We then proposed an integrated framework that keeps the benefits of both fusion and diversity strategies, in terms of the network information flow and link error robustness, for binary consensus over time-invariant network topologies with fading channels. Our results indicate that the proposed technique can improve the consensus performance considerably.

Chapter 6

Conclusions and further extensions

In cooperative network applications, communication plays a key role in the overall performance of the network as each mobile agent improves its knowledge by processing the information received from others. In order to realize the full potentials of these networks, an integrative approach to communication and motion planning issues is essential, i.e., each robot should have an awareness of the impact of its motion decisions on link qualities, when planning its trajectory. This requires each robot to assess the quality of the communication link in the locations that it has not yet visited. As a result, proper prediction of the communication signal strength and fundamentally understanding the spatial predictability of a wireless channel, based on only a few measurements, become considerably important. We addressed this problem in this dissertation.

More specifically, in Chapter 2 we proposed a probabilistic channel prediction framework for predicting the spatial variations of a wireless channel, based on a small number of measurements. By using this framework, we then developed a mathematical foundation for understanding the spatial predictability of wireless channels. More specifically, we characterized the impact of different environments, in terms of their underlying parameters, on wireless channel predictability. We furthermore showed how sampling positions

can be optimized to improve the prediction quality. Finally, we showed the performance of the proposed framework in predicting (and justifying the predictability of) the spatial variations of real channels, using several measurements in our building. In Chapter 3, we showed how to utilize the sparsity of the channel in the frequency domain in order to estimate the spatial variations of a wireless channel, based on a small number of measurements. We also characterized the underlying tradeoffs between the probabilistic approach of Chapter 2 and our sparsity-based estimator, which motivated developing an integrated framework. We then proposed an integrated sparsity and model-based channel prediction framework. Our approach properly takes advantage of both channel compressibility in the frequency domain and channel probabilistic characterization in the spatial domain. We tested our framework using outdoor and indoor channel measurements. The results confirmed the superior performance of the proposed integrated approach.

In Chapters 4 and 5, we studied different cooperative network operations with limited local sensing and realistic modeling of communication links. In Chapter 4, we considered the group agreement problem over fading channels. More specifically, a cooperative network is trying to reach consensus on the occurrence of an event, by communicating over *time-varying network topologies* with fading channels. We characterized the impact of fading and rapidly-changing topologies on both the asymptotic and transient behaviors of the network. We showed that the network can converge to a memoryless state asymptotically, depending on the utilized decision-making function. To see the transient behavior, we then characterized the rate of convergence by deriving an approximation for the second largest eigenvalue of the underlying average dynamical system for different decision-making strategies. We then showed how to significantly improve both the asymptotic and transient consensus performance by incorporating sensing and communication trust factors in the local decision functions. We furthermore considered two cases of diversity and fusion for local decision making. We mathematically characterized the performance of our proposed framework. Our derivations, for instance, showed how noise, fading and connectivity impact the performance. They furthermore highlighted the underlying tradeoffs

between diversity and fusion approaches in terms of speed of convergence and memoryless asymptotic value. In Chapter 5, we then proposed a framework that keeps the benefits of both fusion and diversity strategies. We mathematically analyzed the proposed algorithm and showed how it can achieve accurate consensus asymptotically. Our results indicated that the proposed technique improves the consensus performance considerably.

There are several possible extensions of the results of this dissertation. In the proposed probabilistic framework of Chapter 2, we assumed that the position of the transmitter is fixed. A possible interesting extension is to understand channel predictability when both the transmitter and receiver are moving. Also, we assumed narrowband channels. Considering the impact of wideband channels is also another possible extension. In Chapter 4, we mentioned that binary consensus has several potential applications in cooperative network operations. One potential application is cooperative spectrum sensing in cognitive radio networks. In cooperative spectrum sensing, the secondary users (unlicensed users) sense the spectrum power of the primary user (licensed user) and utilize the spectrum when the primary user is in an idle mode. The secondary users will then communicate with each other to improve their detection performance by exploiting their spatial diversity. Then, since the goal of the network is to reach an agreement on an occurrence of a binary event, i.e. the presence or absence of the primary user, this problem directly falls into the binary consensus category. Therefore, the binary consensus framework of this dissertation can be extended to address spectrum sensing in cognitive radio networks.

References

[1] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, pp. 1289–1306, April 2006.

[2] D. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Transactions on Information Theory*, vol. 47, pp. 2845–2862, November 2001.

[3] E. J. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math*, vol. 59, no. 8, pp. 1207–1223, 2005.

[4] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," tech. rep., 2007.

[5] W. M. Smith, *Urban propagation modeling for wireless systems*. PhD thesis, Stanford University, 2004.

[6] A. Ghaffarkhah and Y. Mostofi, "Communication-aware motion planning in mobile networks," *IEEE Transactions on Automatic Control*, vol. 56, pp. 2478 –2485, oct. 2011.

[7] W. Wang, V. Srinivasan, B. Wang, and K. Chua, "Coverage for target localization in wireless sensor networks," *IEEE Trans. on Wireless Communications*, vol. 7, pp. 667 –676, feb. 2008.

[8] J. Lu and T. Suda, "Differentiated surveillance for static and random mobile sensor networks," *IEEE Trans. on Wireless Communications*, vol. 7, pp. 4411 –4423, nov. 2008.

[9] K. Hung and K. Lui, "On perimeter coverage in wireless sensor networks," *IEEE Trans. on Wireless Communications*, vol. 9, pp. 2156 –2164, jul. 2010.

[10] A. Jadbabaie, J. Lin, and A. S. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Transactions on Automatic Control*, vol. 48, pp. 988–1001, 2003.

[11] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

[12] T. S. Rappaport, *Wireless Communications: Principles and Practice*. Upper Saddle River, NJ, USA: Prentice-Hall, 2001.

[13] W. C. Jakes, *Microwave Mobile Communications*. New York: Wiley-IEEE Press, 1994.

[14] H. Hashemi, "The indoor radio propagation channel," *Proceedings of the IEEE*, vol. 81, pp. 943–968, Jul 1993.

[15] K. Remley, H. Anderson, and A. Weisshar, "Improving the accuracy of ray-tracing techniques for indoor propagation modeling," *IEEE Transactions on Vehicular Technology*, vol. 49, pp. 2350–2358, Nov 2000.

[16] C. F. Yang, B. C. Wu, and C. J. Ko, "A ray-tracing method for modeling indoor wave propagation and penetration," *IEEE Transactions on Antennas and Propagation*, vol. 46, pp. 907–919, Jun 1998.

[17] A. Schwab and P. Fischer, "Maxwell, hertz, and german radio-wave history," *Proc. of the IEEE*, vol. 86, jul. 1998.

[18] P. Zakharov, R. Dudov, E. Mikhailov, A. Korolev, and A. Sukhorukov, "Finite integration technique capabilities for indoor propagation prediction," pp. 369 –372, nov. 2009.

[19] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. The MIT Press, Cambridge, Massachusetts, London, England, 2006.

[20] J. Cortes, "Distributed kriged kalman filter for spatial estimation," in *IEEE Trans. on Automatic Control*, pp. 2816 – 2827, 2009.

[21] J. Choi, J. Lee, , and S. Oh., "Biologically-inspired navigation strategies for swarm intelligence using spatial gaussian processes," in *17th International Federation of Automatic Control (IFAC)*, July 2008.

[22] J. Choi, J. Lee, , and S. Oh., "Swarm intelligence for achieving the global maximum using spatio-temporal gaussian processes," in *American Control Conference*, June 2008.

[23] S. Martínez, "Distributed interpolation schemes for field estimation by mobile sensor networks," *IEEE Transactions on Control Systems Technology*. In press, 2009.

[24] S.-J. Kim, E. Dall'Anese, and G. Giannakis, "Cooperative spectrum sensing for cognitive radios using kriged kalman filtering," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 5, pp. 24 –36, feb. 2011.

[25] A. Schmidt and J. M. F. Moura, "Field inversion by consensus and compressed sensing," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, p. 24172420, April 2009.

[26] N. Cressie, "Kriging nonstationary data," *Journal of the American Statistical Association*, vol. 81, no. 395, pp. 625–634, 1986.

[27] N. Cressie, *Statistics for Spatial Data*. Wiley-Interscience, January 1993.

[28] A. Stein and L. C. A. Corsten, "Universal kriging and cokriging as a regression procedure," *Biometrics*, vol. 47, no. 2, pp. 575–587.

[29] Y. Yan and Y. Mostofi, "Robotic router formation in realistic communication environments," *IEEE Transactions on Robotics*, vol. 28, pp. 810 – 827, August 2012.

[30] A. Ghaffarkhah and Y. Mostofi, "Path planning for networked robotic surveillance," *IEEE Transactions on Signal Processing*, vol. 60, pp. 3560 –3575, july 2012.

[31] M. Malmirchegini and Y. Mostofi, "On the spatial predictability of communication channels," *IEEE Transactions on Wireless Communications*, vol. 11, pp. 964–978, March 2012.

[32] Y. Mostofi, M. Malmirchegini, and A. Ghaffarkhah, "Estimation of communication signal strength in robotic networks," in *Proceedings of IEEE International Conference on Robotics and Automation*, pp. 1946 –1951, May 2010.

[33] Y. Tsaig and D. L. Donoho, "Extensions of compressed sensing," *Signal Processing, Elsevier*, vol. 86, pp. 549–571, March 2006.

[34] C. E. Shannon, "Communication in the presence of noise," *Proc. Institute of Radio Engineers*, vol. 37, pp. 10–21, January 1949.

[35] S. Cotter and B. Rao, "Sparse channel estimation via matching pursuit with application to equalization," *IEEE Transactions on Communications*, vol. 50, pp. 374 –377, mar 2002.

[36] J. Meng, W. Yin, H. Li, E. Houssain, and Z. Han, "Collaborative spectrum sensing from sparse observations using matrix completion for cognitive radio networks," in *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, pp. 3114 –3117, march 2010.

[37] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proceedings of the IEEE*, vol. 98, pp. 1058 –1076, june 2010.

[38] D. Baron, S. Sarvotham, and R. G. Baraniuk, "Bayesian compressive sensing via belief propagation," *IEEE Transactions on Signal Processing*, vol. 58, pp. 269 –280, jan. 2010.

[39] V. K. Goyal, A. K. Fletcher, and S. Rangan, "Compressive sampling and lossy compression," *IEEE Signal Processing Magazine*, vol. 25, pp. 48 –56, march 2008.

[40] B. Babadi, N. Kalouptsidis, and V. Tarokh, "SPARLS: The sparse RLS algorithm," *IEEE Transactions on Signal Processing*, vol. 58, pp. 4013 –4025, aug. 2010.

[41] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, pp. 83 –91, march 2008.

[42] W. U. Bajwa, J. D. Haupt, G. M. Raz, S. J. Wright, and R. D. Nowak, "Toeplitz-structured compressed sensing matrices," in *IEEE/SP 14th Workshop on Statistical Signal Processing*, pp. 294 –298, aug. 2007.

[43] W. U. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, "Joint source channel communication for distributed estimation in sensor networks," *IEEE Transactions on Information Theory*, vol. 53, pp. 3629 –3653, oct. 2007.

[44] W. U. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, "Compressive wireless sensing," in *The Fifth International Conference on Information Processing in Sensor Networks (IPSN)*, pp. 134 –142, 2006.

[45] M. Rabbat, J. Haupt, A. Singh, and R. Nowak, "Decentralized compression and predistribution via randomized gossiping," in *The Fifth International Conference on Information Processing in Sensor Networks (IPSN)*, pp. 51 –59, 2006.

[46] M. R. Duarte, M. Wakin, D. Baron, and R. Baraniuk, "Universal distributed sensing via random projections," in *The Fifth International Conference on Information Processing in Sensor Networks (IPSN)*, pp. 177 –185, 2006.

[47] W. U. Bajwa, A. M. Sayeed, and R. Nowak, "Learning sparse doubly-selective channels," in *46th Annual Allerton Conference on Communication, Control, and Computing*, pp. 575 –582, 2008.

[48] M. Sharp and A. Scaglione, "Application of sparse signal recovery to pilot-assisted channel estimation," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3469 –3472, april 2008.

[49] G. Taubock and F. Hlawatsch, "A compressed sensing technique for OFDM channel estimation in mobile environments: Exploiting channel sparsity for reducing pilots," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 31 2008-april 4 2008.

[50] M. Malmirchegini and Y. Mostofi, "An integrated sparsity and model-based probabilistic framework for estimating the spatial variations of communication channels," *special issue of Elsevier Physical Communication Journal on Compressive Sensing in Communications*, pp. 1658–1663, June 2012.

[51] D. Grunbaum, A. Okubo, and S. A. Levin, "Modelling social animal aggregations," in *Frontiers in theoretical Biology: Lecture notes in biomathematics*, Springer-Verlag, 1994.

[52] C. M. Breder, "Equations descriptive of fish schools and other animal aggregations," *Ecology*, vol. 35, pp. 361–370, 1954.

[53] A. Okubo, "Dynamical aspects of animal grouping: swarms, schools, flocks, and herds.," *Adv Biophys*, vol. 22, pp. 1–94, 1986.

[54] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet, "Novel type of phase transition in a system of self-driven particles," *Physical Review Letters*, vol. 75, no. 6, pp. 1226–1229, 1995.

[55] C. W. Reynolds, "Flocks, herds, and schools: A distributed behavioral model," *Computer Graphics*, vol. 21, pp. 25–34, 1987.

[56] J. A. Fax and R. M. Murray, "Information flow and cooperative control of vehicle formations," *IEEE Transactions on Automatic Control*, vol. 49, pp. 1465–1476, September 2004.

[57] J. A. Marshall, M. E. Broucke, and B. A. Francis, "A pursuit strategy for wheeled-vehicle formations," in *Proceedings of the 42nd IEEE Conference on Decision and Control*, (Maui, HI), pp. 2555–2560, December 2003.

[58] H. G. Tanner, A. Jadbabaie, and G. J. Pappas, "Stable flocking of mobile agents, part I: Fixed topology," in *Proceedings of CDC*, (Maui, Hawaii), pp. 2010–2015, December 2003.

[59] H. G. Tanner, A. Jadbabaie, and G. J. Pappas, "Stable flocking of mobile agents, part II: Dynamic topology," in *Proceedings of CDC*, (Maui, Hawaii), pp. 2016–2021, December 2003.

[60] J. Lin, A. S. Morse, and B. D. O. Anderson, "The multi-agent rendezvous problem," in *Proceedings of the 42nd IEEE Conference on Decision and Control*, (Maui, Hawaii), pp. 1508–1513, December 2003.

[61] "DARPA Grand Challenge." http://www.darpa.mil/grandchallenge/index.asp.

[62] A. Rodriguez-Angeles, "Cooperative synchronization of robots via estimated state feedback," in *Proceedings of the 42nd IEEE Conference on Decision and Control*, (Maui, Hawaii), pp. 1514–1519, December 2003.

[63] A. Jadbabaie, N. Motee, and M. Barahona, "On the stability of the Kuramoto model of coupled nonlinear oscillators," in *Proceedings of the 2004 American Control Conference*, (Boston, MA), pp. 4296–4301, 2004.

[64] W. Ren and R. W. Beard, "Consensus seeking in multi-agent systems under dynamically changing interaction topologies," *IEEE Transactions on Automatic Control*, vol. 50, no. 5, pp. 655–661, 2005.

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

[66] Y. Hatano and M. Mesbahi, "Agreement over random networks," *IEEE Transactions on Automatic Control*, vol. 50, no. 11, pp. 1867–1872, 2005.

[67] M. Mesbahi, "State-dependent graphs," in *Proceedings of 42nd IEEE conference on Decision and Control*, 2003.

[68] J. Lin, A. S. Morse, and B. Anderson, "The multi-agent rendezvous problem - part I: the synchronous case," *SIAM Journ. on Cont. & Opt.*, 2006.

[69] J. Cortes, S. Martinez, and F. Bullo, "Robust rendezvous for mobile autonomous agents via proximity graphs in arbitrary dimensions," *IEEE Transactions on Automatic Control*, vol. 51, no. 8, pp. 1289–1298, 2006.

[70] V. D. Blondel, J. M. Hendrick, A. Olshevsky, and J. N. Tsitsiklis, "Convergence in multiagent coordination, consensus, and flocking," in *IEEE Conf. on Decision and Control and European Control Conference*, pp. 2996–3000, December 2005.

[71] D. P. Spanos, R. Olfati-Saber, and R. M. Murray, "Approximate distributed Kalman filtering in sensor networks with quantifiable performance," in *the proceedings of 4th international conf. on Information Processing in Sensor Networks*, pp. 133–139, 2005.

[72] W. Ren, R. Beard, and E. Atkins, "A survey of consensus problems in multi-agent coordination," in *Proceedings of the American Control Conference*, pp. 1859–1864 vol. 3, June 2005.

[73] W. Ren, R. Beard, and D. Kingston, "Multi-agent Kalman consensus with relative uncertainty," in *Proceedings of American Control Conf.*, 2005.

[74] T. C. Aysal, M. E. Yildiz, A. D. Sarwate, and A. Scaglione, "Broadcast gossip algorithms for consensus," *IEEE Transactions on Signal Processing*, vol. 57, pp. 2748–2761, July 2009.

[75] S. Kar and J. Moura, "Distributed consensus algorithms in sensor networks with imperfect communication: Link failures and channel noise," *IEEE Transactions on Signal Processing*, vol. 57, pp. 355–369, Jan. 2009.

[76] G. Scutari and S. Barbarossa, "Distributed consensus over wireless sensor networks affected by multipath fading," *IEEE Transactions on Signal Processing*, vol. 56, pp. 4100–4106, Aug. 2008.

[77] 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, pp. 350–364, January 2008.

[78] I. D. Schizasa, G. B. Giannakis, S. D. Roumeliotis, and A. Ribeiro, "Consensus in Ad Hoc WSNs with noisy links - part II: Distributed estimation and smoothing of random signals," *IEEE Trans. on Signal Processing*, vol. 56, pp. 1650–1666, April 2008.

[79] Y. Mostofi, "Binary Consensus with Gaussian Communication Noise: A Probabilistic Approach," in *Proceedings of the 46th IEEE Conference on Decision and Control (CDC)*, December 2007.

[80] Y. Hong, A. Scaglione, and P. Varshney, "A communication architecture for reaching the consensus in decision for a large network," in *IEEE Statistical Signal Processing Workshop, SSP2005*, (Bordeaux, France), pp. 17–22, July 2005.

[81] R. Olfati-Saber, E. Franco, E. Frazzoli, and J. S. Shamma, "Belief consensus and distributed hypothesis testing in sensor networks," in *Network Embedded Sensing and Control*, vol. 331, pp. 169–182, 2006.

[82] M. E. Yildiz and A. Scaglione, "The limiting rate behavior and rate allocation strategies for average consensus problems with bounded convergence," in *Proc. ICASSP*, (Las Vegas, NV), April 2008.

[83] P. Frasca, R. Carli, F. Fagnani, and S. Zampieri, "Average consensus by gossip algorithms with quantized communication," in *Proceedings of the 47th IEEE Conference on Decision and Control (CDC)*, December 2008.

[84] A. Kashyap, T. Basar, and R. Srikant, "Quantized consensus," in *Proc. of Intl. Symp. on Information Theory*, July 2006.

[85] R. Carli, F. Fagnani, P. Frasca, and S. Zampieri, "Gossip consensus algorithms via quantized communication," in *review*, 2009.

[86] T. Aysal, M. Coates, and M. Rabbat, "Distributed average consensus with dithered quantization," *IEEE Transactions on Signal Processing*, vol. 56, pp. 4905–4918, Oct. 2008.

[87] Y. Ruan and Y. Mostofi, "Binary consensus with soft information processing in cooperative networks," in *47th IEEE Conference on Decision and Control (CDC)*, Dec. 2008.

[88] J. Liu, V. Yadav, H. Sehgal, J. Olson, H. Liu, and N. Elia, "Phase transitions on fixed connected graphs and random graphs in the presence of noise," in *the proceedings of 44th IEEE Conference on Decision and Control*, pp. 2996–3000, 2005.

[89] Y. Mostofi and Y. Yuan, "Impact of heterogeneous link qualities and network connectivity on binary consensus," in *American Control Conference*, June 2009.

[90] M. Malmirchegini, Y. Ruan, and Y. Mostofi, "Binary consensus over fading channels: A best affine estimation approach," in *IEEE Global Telecommunications Conference*, pp. 1–6, Dec. 2008.

[91] Y. Mostofi and M. Malmirchegini, "Binary consensus over fading channels," *IEEE Transactions on Signal Processing*, vol. 58, pp. 6340–6354, Dec. 2010.

[92] M. Malmirchegini and Y. Mostofi, "An integrated diversity and fusion framework for binary consensus over fading channels," *in revision, EURASIP Journal on Advances in Signal Processing*, 2012.

[93] K. W. Cheung, "Error in estimating local average power of multipath signals," in *Electronics letters*, vol. 27, 1991.

[94] C. Anderson, T. Rappaport, K. Bae, A. Verstak, N. Ramakrishnan, W. Tranter, C. Shaffer, and L. Watson, "In-building wideband multipath characteristics at 2.5 and 60 GHz," in *Proceedings of Vehicular Technology Conference*, vol. 1, 2002.

[95] J. G. Proakis, *Digital Communications*. McGraw-Hill, 2001.

[96] "Rssi in madwifi." http://madwifi-project.org/wiki/UserDocs/RSSI.

[97] S. Cotton and W. Scanlon, "Higher order statistics for lognormal small-scale fading in mobile radio channels," *Antennas and Wireless Propagation Letters, IEEE*, vol. 6, pp. 540 –543, 2007.

[98] M. Gudmundson, "Correlation model for shadow fading in mobile radio systems," *Electronics Letters*, vol. 27, Nov. 1991.

[99] R. Horn and C. Johnson, *Matrix Analysis*. New York: Cambridge University Press, 1999.

[100] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Prentice Hall, 2000.

[101] J. Dattorro, *Convex Optimization and Euclidean Distance Geometry*. Meboo, 2005.

[102] I. J. Schoenberg, "Metric spaces and positive definite functions," *Trans. of the American Mathematical Society*, 1938.

[103] E. Kreyszig, *Introductory Functional Analysis with Applications*. Wiley, 1989.

[104] Y. Mostofi, A. Gonzalez-Ruiz, A. Ghaffarkhah, and D. Li, "Characterization and Modeling of Wireless Channels for Networked Robotic and Control Systems - A Comprehensive Overview," *Proc. of Int. Conf. on Intelligent Robots and Systems*, October 2009.

[105] D. Akerberg, "Properties of a tdma pico cellular office communication system," in *IEEE Conf. on Global Communications*, 1988.

[106] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. on Information Theory*, vol. 52, pp. 489–509, February 2006.

[107] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Compte Rendus de l'Academie des Sciences*, vol. 346, pp. 589–592, 2008.

[108] F. Santosa and W. W. Symes, "Linear inversion of band-limited reflection seismograms," *SIAM Journal on Scientific and Statistical Computing*, vol. 7, no. 4, pp. 1307–1330, 1986.

[109] E. J. Candes, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique, Elsevier*.

[110] D. Needell and R. Vershynin, "Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit," 2007. Preprint.

[111] M. Rudelson and R. Vershynin, "Sparse reconstruction by convex relaxation: Fourier and Gaussian measurements," 2006. Preprint.

[112] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[113] Y. Mostofi and P. Sen, "Compressive Cooperative Mapping in Mobile Networks," in *Proceedings of the 28th American Control Conference (ACC)*, (St. Louis, MO), pp. 3397–3404, June 2009.

[114] H. Rauhut and R. Ward, "Sparse Legendre expansions via l_1-minimization," *Preprint*, 2010.

[115] J. J. Fuchs, "More on sparse representations in arbitrary bases," *IEEE Trans. on Information Theory*, pp. 1341–1344, 2004.

[116] A. Zymnis, S. Boyd, and E. Candes, "Compressed sensing with quantized measurements," *IEEE Trans. on Signal Processing*, pp. 149–152, Feb 2010.

[117] S. Wright, R. Nowak, and M. Figueiredo, "Sparse reconstruction by separable approximation," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 3373–3376, April 2008.

[118] Y. Mostofi and Y. Ruan, "Binary Consensus over AWGN Channels," *in revision, IEEE Transactions on Automatic Control*, Dec. 2008.