University of Arkansas, Fayetteville
ScholarWorks@UARK

12-2013

Distortion-Tolerant Communications with Correlated Information

Ning Sun
University of Arkansas, Fayetteville

Follow this and additional works at: http://scholarworks.uark.edu/etd

Part of the Systems and Communications Commons

Recommended Citation

Distortion-Tolerant Communications with Correlated Information

Distortion-Tolerant Communications with Correlated Information

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Electrical Engineering

By

Ning Sun
Shandong University of Science and Technology
Bachelor of Science in Electrical Engineering, 2005
Shandong University of Science and Technology
Master of Science in Electrical Engineering, 2008

December 2013
University of Arkansas

This dissertation is approved for recommendation to the Graduate Council.

_____

Dr.Jingxian Wu
Dissertation Director

_____   _____

Dr. Randy Brown   Dr. Dale R. Thompson
Committee Member   Committee Member

_____

Dr. Jing Yang
Committee Member

Abstract

This dissertation is devoted to the development of distortion-tolerant communication techniques by exploiting the spatial and/or temporal correlation in a broad range of wireless communication systems under various system configurations. Signals observed in wireless communication systems are often correlated in the spatial and/or temporal domains, and the correlation can be used to facilitate system designs and to improve system performance. First, the optimum node density, *i.e.*, the optimum number of nodes in a unit area, is identified by utilizing the spatial data correlation in the one- and two-dimensional wireless sensor networks (WSNs), under the constraint of fixed power per unit area. The WSNs distortion is quantized as the mean square error between the original and the reconstructed signals. Then we extend the analysis into WSNs with spatial-temporally correlated data. The optimum sampling in the space and time domains is derived. The analytical optimum results can provide insights and guidelines on the design of practical WSNs. Second, distributed source coding schemes are developed by exploiting the data correlation in a wireless network with spatially distributed sources. A new symmetric distributed joint source-channel coding scheme (DJSCC) is proposed by utilizing the spatial source correlation. Then the DJSCC code is applied to spatial-temporally correlated sources. The temporal correlated data is modeled as the Markov chain. Correspondingly, two decoding algorithms are proposed. The first multi-codeword message passing algorithm (MCMP) is designed for spatially correlated memoryless sources. In the second algorithm, a hidden Markov decoding process is added to the MCMP decoder to effectively exploit the data correlation in both the space and time domains. Third, we develop distortion-tolerant high mobility wireless communication systems by considering correlated channel state information (CSI) in the time domain, and study the optimum

designs with imperfect CSI. The pilot-assisted channel estimation mean square error is expressed as a closed-form expression of various system parameters through asymptotic analysis. Based on the statistical properties of the channel estimation error, we quantify the impacts of imperfect CSI on system performance by developing the analytical symbol error rate and a spectral efficiency lower bound of the communication system.

Acknowledgements

I would like to gratefully and sincerely express my gratitude to all the people who help me to complete this dissertation.

First, I would like to give my deepest appreciation to my supervisor, Dr. Jingxian Wu, who is the best advisor and mentor I had so far throughout all my studies. His unremitting zeal and diligence to the scientific research inspired me deeply to pursue the truth in areas of communications. Without his intelligent guidance, caring, patience, and persistent help, this dissertation would not have been finished.

Second, I am also extremely grateful to my dissertation committee members, Dr. Scott C. Smith, Dr. Dale R. Thompson, Dr. Randy Brown, and Dr. Jing Yang. Their careful reviews and valuable comments have helped me to improve the quality of the dissertation. Special thanks goes to Dr. Randy Brown, who was willing to serve in my final defense committee at the last moment.

Last but not least, I would like to thank my parents, husband, and little boy Enoch. They were always there supporting me and encouraging me. The warmth of the family love helps me to overcome all the difficulties, not only in the study but also in the daily life.

Contents

List of Papers

- **Chapter 2**, Jingxian Wu and Ning Sun, "Optimum sensor density in distortion tolerant wireless sensor networks," *IEEE Trans. Wireless Commun.*, vol. 11, pp. 2056-2064, June 2012.

- **Chapter 3**, Ning Sun and Jingxian Wu, "Optimum sampling in spatial-temporally correlated wireless sensor networks," *EURASIP J. Wireless Commun. Networking*, doi:10.1186/1687-1499-2013-5, Jan. 2013.

- **Chapter 4**, Ning Sun , Jingxian Wu, and Hai Lin, "Distributed joint source and channel code with correlated information sources," in *Proc. IEEE Intern. Conf. Commun. China ICCC'12*, Aug. 2012

- **Chapter 5**, Ning Sun , Jingxian Wu, and Guoqing Zhou, "Distributed joint source-channel code for spatial-temporally correlated Markov sources," in *Proc. IEEE Intern. Conf. Commun. ICC'13*, June 2013.

- **Chapter 6**, Ning Sun and Jingxian Wu, "Maximizing spectral efficiency for high mobility systems with imperfect channel state information," accepted by *IEEE Trans. Wireless Commun.*, 2013.

Chapter 1

Introduction

1.1 Background and Motivation

The well celebrated Shannon channel coding theorem states that given a noisy communication channel, if the transmitted data rate is less than the channel capacity, then there exists channel codes that can achieve an arbitrarily small error probability at the receiver [1]. In other words, theoretically, we might be able to transmit the information without error as long as the data rate is below the channel capacity. Under this *error-free* communication assumption, most works perform their analysis by using the design metric of network capacity, which is the maximum throughput supported by a network. However, as demonstrated in [2], such an error-free assumption is over-restrictive in the design of practical systems. Due to noise and interference during data transmission, distortion is inevitable in practical communication systems. In reality, a small amount of errors might be acceptable for real world applications such as target detection [3], [4], [5], target localization [6], information coverage [7], [8], and information recovery [9]–[13], etc. With the growing demands for those applications, The concept of *distortion-tolerant communications* [13][14] is used in this dissertation. Under the distortion-tolerant communication framework, controlled distortion is allowed at the receiver as long as the transmitted information can be recovered beyond a certain fidelity requirement, in exchange of energy and spectral efficient communications. Compared to conventional communication techniques that focus on distortion-free communications or error minimization, the distortion-tolerant systems can provide extra degrees-of-freedom and additional

benefits in the system designs, such as higher spectral efficiency, lower power consumption, and better flexibility.

Specifically, the distortion-tolerant communication systems will be developed by exploiting the correlation among the signals observed in wireless communication networks. There are two types of signal correlations. The first type is the spatial-temporal correlation among the data to be transmitted in a wireless system. The spatial-temporal data correlations can be found in a wide range of practical applications, such as environment monitoring with temperature and humidity correlated in the space and time domains, soil and water quality monitoring with the chemical compositions correlated in the space and time domains, and structure health monitoring with spatial-temporally correlated vibration information of the civil structure, etc. The second type correlation is contributed by the propagation environment of the wireless channel. For example, the wireless fading channel changes with respect to time due to the relative movement between the transmitter and receiver, and the fading coefficients at different moments are usually correlated, with the correlated coefficient determined by the movement speed and the carrier frequency. Due to the correlation among the transmitted data as well as the correlation in the channel, signals observed at the communication receivers are usually correlated. This space-time correlation introduces unique challenges and opportunities that can be used to achieve significant system performance gains. Therefore, how to exploit the space and/or time correlated information to design various distortion-tolerant communication systems becomes a promising research topic. In this dissertation, both types of signal correlations will be studied and utilized to facilitate the design of distortion-tolerant communication systems.

Distortion-tolerant wireless sensor network with correlated source information: Wireless sensor network consists of many small, low-power and low-cost wireless sensor nodes. Each

node collects the data from a measurement field, and process and transmit the collected data to a fusion center directly or through multi-hop transmissions with the help of other nodes in the network. In dense wireless sensor networks (WSNs), due to the physical properties of the measured object, the collected information between the spatially distributed sensor nodes may be highly correlated, which is denoted as *spatial correlation*. Meanwhile, in reality, the nature of the energy-radiating physical phenomenon determines the correlation between each consecutive observation of a sensor node, *i.e.*, the physical phenomenon under monitoring changes with respect to time, and the consecutive observations of a sensor node are often correlated temporally. This is *temporal correlation* in WSNs. The existence of the spatial and temporal correlations can bring significant potential improvement for the performance of WSNs.

In this dissertation, we study distortion-tolerant wireless sensor network designs by utilizing the spatial-temporal source correlation. The correlation among the sources can be used to facilitate system design from different perspectives. For example, intuitively, nodes in a low power sensor network need to be densely deployed to enable reliable communication as well as to obtain more spatial samples of the measurement field. However, given a fixed transmission energy per unit area, a higher spatial node density means less transmission energy per node, which degrades the communication performance. The spatial correlation among the collected data samples can be utilized to reduce the node density requirement. Therefore, there exists an optimum node density that can achieve the optimum performance given certain distortion constraints. Similarly, an optimum temporal sampling rate can be obtained if we consider the temporal correlation among the data samples collected by a single node. One approach to exploit the source correlation is through distributed source coding, where spatially distributed sensors can compress their respective information separately without sharing information with each other, yet achieve a compression ratio

approaching the one as if the information is shared among sensors and compressed jointly [15]. Several optimum system designs that can efficiently utilize the spatial-temporal source correlation are presented in this dissertation, and they can lead to energy and spectral efficient designs of distortion-tolerant wireless sensor networks.

Distortion-tolerant high mobility systems with correlated channel information: Next generation wireless networks need to be able to provide high mobility communication to support a wide range of emerging applications such as communication in high speed trains and aircrafts. One of the main challenges faced by high mobility communications is the fast time-varying fading caused by the Doppler shift, which could be as high as 1,000 Hz for a 2.4 GHz system operating at a speed of 450 km/hr. In a high mobility system, the accurate estimation and tracking of the fast time-varying fading are critical to reliable system operations. The time-varying fading can be estimated and tracked by utilizing the temporal correlation among the fading coefficients. Most conventional communication systems are designed by assuming perfect knowledge of the channel state information (CSI) at the receiver. However, in high mobility communications, this assumption might no longer be true, because the fast time-varying fading makes it is almost impossible to achieve error free channel estimation. As a result, channel estimation error is usually inevitable in high mobility systems, and they might have significant impacts on the system performance and designs. How to effectively utilize the temporal correlation among the fading coefficients to design distortion-tolerant high mobility communication systems in the presence of channel estimation error is another topic that will be studied in this dissertation.

1.2 Objectives

The goal of this dissertation is to develop new distortion-tolerant communication techniques and system structures by exploiting the spatial and/or temporal correlation among the signals observed in wireless communication systems. The goal is achieved through the following two objectives.

The first objective is to develop energy and spectral efficient WSNs by utilizing the spatial-temporal correlation among the data collected by the spatially distributed sensors. Theoretical studies are performed to analytically analyze the optimum node density in the space domain and the optimum sampling rate in the time domain for power constrained distortion-tolerant wireless networks. The analytical results can quantitatively identify the interactions among the various system parameters and the estimation fidelity, and provide guidelines on practical system designs. Then new distributed source coding schemes are designed to improve communication performance by utilizing the source correlation. According to the Slepian-Wolf theorem, the distributed sources with correlated information can perform encoding separately, yet achieve a code rate that is the same as when the information is encoded jointly. However, the Slepian-Wolf theorem is not constructive, *i.e.*, it provides no practical coding scheme to achieve the optimum performance. Hence, we aim to develop a practical distributed joint source-channel code, which can combine the source compression and the channel protection in one framework to achieve extra performance gains over conventional systems.

The second objective of our research focuses on the design of the distortion-tolerant high mobility communication system by exploring the correlation among the channel coefficients. The fast time-varying fading is tracked at the receiver by employing the pilot-assisted channel estimation. Since channel estimation error is inevitable in high mobility systems, we aim to study the

5

impacts of the channel estimation error on system performance through both analytical and simulation results. Various parameters such as the Doppler spread, the signal-to-noise ratio (SNR) at pilot locations, and the percentage of pilot symbols in the transmitted symbols are intended to be optimized based on several design metrics, such as spectral efficiency and symbol error rate (SER).

These two objectives are incarnated throughout this dissertation by means of practical design, theoretical analysis and extensive simulations under various system configurations.

1.3   Dissertation Outline

The outline of the rest of the dissertation is given as follows.

Chapter 2: In this chapter, the optimum sensor node density for one- and two dimensional WSNs with *spatial source correlation* is studied. The distortion is quantized as the mean square error (MSE) between the original and recovered information. Under the constraint of fixed power per unit area, the impacts of node density and spatial data correlation on the network performance are investigated for both small networks with finite number of nodes, and large networks with infinite area, infinite node numbers, but finite node density through asymptotic analysis. The optimum node densities are identified for two different objectives of data-gathering, one needs to recover data only at discrete locations, and one needs to recover data at arbitrary locations.

Chapter 3: The optimum sampling in the one- and two-dimensional WSNs with *spatial-temporally correlated data* is studied in this chapter. The impacts of the node density in the space domain and the sampling rate in the time domain are investigated asymptotically by considering a large network with infinite area but finite node density and finite temporal sampling rate, under the constraint of fixed power per unit area. Similarly to Chapter 2, the impact of space-time sampling

on network performances is investigated for two different objectives of data-gathering.

Chapter 4: A new distributed joint source-channel code (DJSCC) is proposed for a communication network with multiple *spatially correlated* information sources. The DJSCC is performed by puncturing the information bits of a linear block code but leaving the parity bits intact. A new unequal energy allocation scheme is proposed for the delivery of the codeword from transmitter to the receiver. The DJSCC can effectively utilize the spatial data correlation, and protect the information from distortions in the channel.

Chapter 5: In this chapter, the new DJSCC code in chapter 4 is extended for a communication network with *spatial-temporally correlated* sources, where the temporal data correlation is modeled as the first order Markov chain. A new multi-codeword message passing (MCMP) decoding algorithm is proposed for the distributed memoryless sources. For the spatially correlated Markov sources, in recognition that the signals at the receiver are distorted observations of the Markov source and thus can be modeled by a hidden Markov model (HMM), a HMM decoding module is added to the MCMP decoder to exploit the temporal data correlation.

Chapter 6: This chapter studies the optimum system design that can maximize the spectral efficiency of high mobility wireless communication systems with imperfect CSI. The channel estimation MSE is analytical derived in closed-form based on pilot-assisted channel stimulation methods. We quantify the impacts of imperfect CSI on system performance by developing the analytical SER and a spectral efficiency lower bound of the communication system. The optimum pilot percentage that can maximize the spectral efficiency lower bound is identified through both analytical and simulation results.

Chapter 7: Conclusion remarks are drawn in this chapter. The major contributions of this research proposal are summarized and some future research topics are listed.

1.4   References

[1] Thomas M. Cover and Joy A. Thomas, *Elements of Information Theory 2nd Edition*, Wiley-Interscience, July 2006.

[2] D. Marco, E.J. Duarte-Melo, M. Liu, and D. L. Neuhoff, "On the many-to-one transport capacity of a dense wireless sensor network and the compressibility of its data," in *Proc. Intern. Conf. Info. Processing Sensor Networks ISPN'03*, pp. 1-16, Apr. 2003.

[3] A. Anandkumar, L. Tong, and A. Swami, "Optimal node density for detection in energy-constrained random networks," *IEEE Trans. Signal Processing*, vol. 56, pp. 5232-5245, Oct. 2008.

[4] T. Wimalajeewa and S. K. Jayaweera, "Impact of mobile node density on detection performance measures in a hybrid sensor network," *IEEE Trans. Wireless Commun.*, vol. 9, pp. 1760-1769, May 2010.

[5] J. F. Chamberland, and V. V. Veeravalli, "How dense should a sensor network be for detection with correlated observations?" *IEEE Trans. Inform. theory*, vol. 52, pp. 5099-5106, Nov. 2006.

[6] W. Wang, V. Srinivasan, B. Wang, and K.-C. Chua, "Coverage for target localization in wireless sensor networks," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 667-676, Feb. 2008.

[7] B. Wang, K.-C. Chua, V. Srinivasan, and W. Wang, "Information coverage in randomly deployed wireless sensor networks," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 2994-3004, Aug. 2007.

[8] L. Liu, X. Zhang, and H. Ma, "Localization-oriented coverage in wireless camera sensor networks," *IEEE Trans. Wireless Commun.*, vol. 10, pp. 484-494, Feb. 2011.

[9] R. Cristescu, and M. Vetterli, "On the optimal density for real-time data gathering of spatio-temporal processes in sensor networks," in *Proc. IPSN 2005*, pp. 159-164, April, 2005.

[10] X. Zhang, H. Wang, F. N. Abdesselam, and A. A. Khokhar, "Distortion Analysis for Real-Time Data Collection of Spatially Temporally Correlated Data Fields in Wireless Sensor Networks," *IEEE Trans. Vehicular Technology*, vol. 58, pp. 1583-1594, March 2009.

[11] M. Gastpar, and M. Vetterli, "Power, Spatio-Temporal Bandwidth, and Distortion in Large Sensor Networks," *IEEE J. Selected Areas Commun.*, vol. 23, pp. 745-755, April 2005.

[12] Y. Sung, H. V. Poor, and H. Yu, "Optimal node density for two-dimensional sensor arrays," in *Proc. IEEE Sensor Array Multichannel Signal Processing Workshop*, pp. 271-274, 2008.

[13] J. Wu and N. Sun, "Optimal sensor density in a distortion-tolerant linear wireless sensor network," in *Proc. IEEE Global Telecommun. Conf. GLOBECOM'10*, Dec. 2010.

[14] J. Wu and N. Sun, "Optimum sensor density in distortion tolerant wireless sensor networks," *IEEE Trans. Wireless Commun.*, vol. 11, pp. 2056-2064, June 2012.

[15] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Info. Theory*, vol. 19, pp. 471-480, July 1973.

Chapter 2

Optimum Sensor Density in Distortion-tolerant Wireless Sensor Networks

Jingxian Wu and Ning Sun

2.1 Abstract

The optimum sensor node density for one- and two-dimensional (1-D and 2-D) wireless sensor networks (WSNs) with spatial source correlation is studied in this chapter. The WSN attempts to reconstruct a spatially correlated signal field by collecting the location-dependent measurements from the distributed sensor nodes. The WSN is designed to minimize the mean square error (MSE) distortion between the reconstructed and the original signal field under the constraint of fixed power per sample per unit area. The impacts of node density and spatial data correlation on the network performance are investigated for both small networks with finite number of nodes, and large networks with infinite area, infinite node numbers, but finite node density through asymptotic analysis. The interactions among the various network parameters and their impacts on the system performance are quantitatively identified with exact analytical expressions, many of which are in closed-form. Under fixed power per unit area, it is discovered that: 1) for applications that only need to recover data at discrete locations, placing exact one sensor at the desired measurement locations will generate the optimum performance; 2) for applications that need to recover data at arbitrary locations in the measurement field, the optimum node density is a function of the spatial data correlation; and 3) the 1-D and 2-D networks have similar performance trends with respect to node density, and their performance difference diminishes as the spatial correlation increases.

2.2 Introduction

A wireless sensor network (WSN) provides autonomous monitoring of physical or environmental conditions by using a group of spatially distributed sensor nodes transmitting measured data to a fusion center (FC) [1]. One of the primary challenges faced by the design of a large WSN is to determine the node density, *i.e.*, the number of nodes in a unit area, to optimize the network performance under the energy and/or cost constraints [1]–[12]. Given a fixed transmission power per unit area, a higher node density means less power available to each node, which degrades the network performance due to the reduced signal-to-noise ratio (SNR) for the signal transmitted by each node. On the other hand, a higher node density can obtain more data samples per unit area, which can benefit the system performance. Such a trade-off relationship necessitates the study of the optimum node density in the design of practical WSNs.

There have been considerable works in the literature investigating the impacts of node density on the network performance for both one-dimensional (1-D) and two-dimensional (2-D) networks [1]-[15]. The seminal work by Gupta and Kumar [13] discovers that the per node throughput in an ad hoc network scales with $\mathcal{O}\left(\frac{1}{\sqrt{N \log N}}\right)$, with N being the number of nodes per unit area, *i.e.*, the node density. The result in [13] does not consider the spatial data correlation. Data collected in the real world often contain redundancies due to the spatial correlation inherent in the monitored object(s). In [14], a Wiener process is used to model the spatial correlation of an 1-D field. It is demonstrated that, due to the spatial data correlation, distortion-free communication can be achieved even if the per node throughput tends to 0 as $N \to \infty$. The above study is for peer-to-peer networks, where there are equal numbers of sources and destinations. For many-to-one networks such as a WSN, it is demonstrated in [15] that no compression scheme is sufficient to achieve

11

distortion-free communications in a many-to-one network.

The analysis in most of the previous works is performed by using the design metric of network capacity, which is the maximum throughput supported by a network with *error-free* communication. As demonstrated in [15], such an *error-free* assumption is over-restrictive in the design of practical WSNs. In reality, a small amount of errors might be acceptable for real world applications such as target detection [2], [3], [7], target localization [4], information coverage [5], [6], and information recovery [8]–[12], etc. In [2], the optimum node density maximizing the Neyman-Pearson detection error exponent is characterized subject to a constraint on the average energy consumption per node. In [7], the optimum network density of an 1-D network is studied by using the detection error probability as the performance metric. Both [2] and [7] are developed for WSNs with binary detection at the FC. In [8], an arbitrary point on a continuous measurement field is estimated by interpolating the samples collected by the spatially discrete sensors, and there is a finite optimum node density that minimizes the estimation mean square error (MSE) over the measurement field. The model in [8] is extended in [9] by considering realistic transmission schemes, such as limited transmission range and practical network/routing parameters. The studies in [8] and [9] only consider the distortion from the spatial interpolation, and the distortions introduced by the noisy channel are not incorporated in the analysis. A distortion lower bound is derived in [10] for a network with a finite number of correlated sources as a function of the number of sensors and spatial-temporal communication bandwidth. The analysis in [10] is only applicable to a measurement field with finite degree-of-freedom. In [11], the asymptotic optimum sensor density that can maximize the information collected from a 2-D correlated random fields is obtained under a total energy constraint as the number of nodes and the measurement field tends to infinity. A total energy constraint in a large network often leads to negligible energy per node or per unit

area, yet in reality the energy budget per unit area is non-trivial. A constraint on fixed power per unit area, or equivalently, fixed energy per sample per unit area, is adopted in our previous work [12] to obtain the asymptotically optimum node density as the number of nodes and the size of the measurement field tends to infinity. The power or energy per sample constraint will limit the instantaneous power of a sensor node, and it can be easily translated to the energy constraint of a node under a fixed sampling rate. The constraint on power per unit area ensures a fair comparison between networks with different sizes and different node densities. The results in [12] are only applicable to 1-D networks, and the analysis is based on an approximation of the minimum mean square error (MMSE) spatial interpolation.

In this chapter, we investigate the optimum node density in the 1-D and 2-D distortion-tolerant networks under the constraint of fixed power per unit area. The WSN attempts to reconstruct a spatially correlated signal field by collecting measurements from distributed sensor nodes. There is no assumption on the statistical properties of the field, other than that it forms a continuous random process that is wide sense stationary (WSS) in the space domain. In recognition of the distortion-tolerance of many practical applications, the optimization is performed with respect to the MSE distortion between the recovered and the original signal fields. Each sensor node collects spatially correlated samples of the measurement field, and forwards the information to a data FC through a noisy link. The FC reconstructs an estimate of the signal field by exploiting the spatial data correlation with an optimum MMSE receiver. Under the constraints of fixed transmission power per unit area, the impacts of node density and spatial data correlation on MSE are investigated for both small networks with finite number of nodes and large networks with infinite area but finite node density. Exact analytical expressions are obtained to describe the interactions and tradeoff relationship between SNR per node, which is inversely proportional to node density, and spatial

13

sample correlation, which increases with the node density, for both 1-D and 2-D networks. Based on the analytical results, the optimum node densities for two different objectives of data-gathering, one needs to recover data only at discrete locations, and one needs to recover data at arbitrary locations, are identified through the analysis. It is observed that the 1-D and 2-D networks have similar performance trends, regardless whether the node number is finite or infinite, and their performance difference diminishes as the spatial correlation increases.

The remainder of this chapter is organized as follows. Section 2.3 introduces the system model and a two-step MMSE estimation method. Sections 2.4 and 2.5 study the impacts of node density on the performance of 1-D and 2-D networks, respectively, by following the two-step MMSE method. In these two sections, the optimum node densities in various networks with different node number are identified, and numerical examples are presented to demonstrate the interactions among the various system parameters. Section 2.6 concludes the chapter.

2.3 Problem Formulation

2.3.1 System Model

Consider a WSN with N sensor nodes evenly distributed over a measurement field Ω_η, as shown in Fig. 2.1(a) for an 1-D network, and Fig. 2.1(b) for a 2-D network. The smallest distance between two adjacent sensors is d. The n-th node is placed at a location with coordinate $\eta_n \in \Omega_\eta$, for $n = 1, 2, \cdots, N$. The even node distribution shown in Fig. 2.1(a) is adopted here for mathematical tractability, and similar configurations have been widely used in the literature [7]–[9], [11], [12], and [14]. Define the node density, δ, as the number of nodes in a unit area. The node densities are $\delta = \frac{1}{d}$ and $\delta = \frac{1}{d^2}$ for the 1-D and 2-D networks, respectively.

14

(a) An 1-D network

(b) A 2-D network
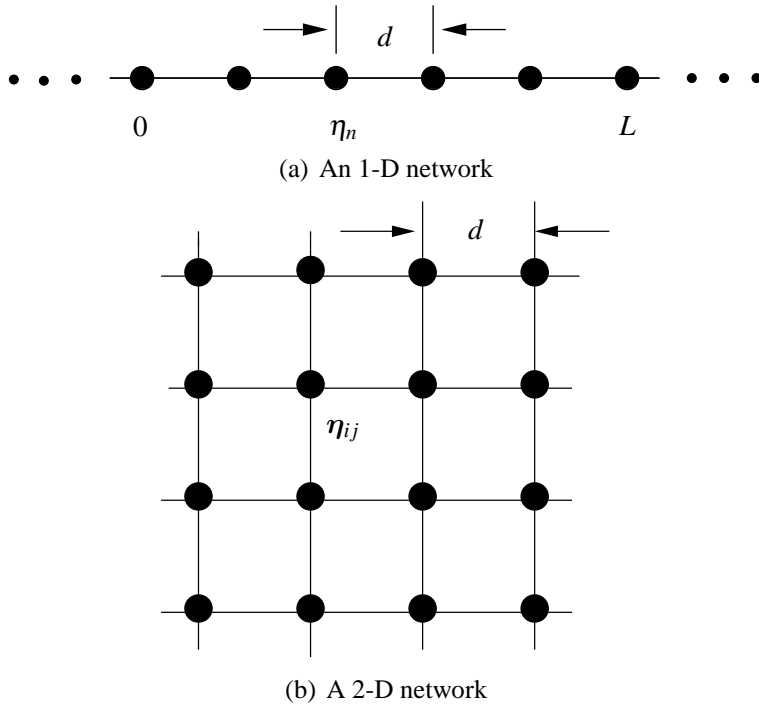
Figure 2.1: The one- and two-dimensional networks with evenly distributed nodes.

Each sensor will measure a location dependent physical quantity, $x(\boldsymbol{\eta}_n)$, such as vibration intensity of a bridge, humidity, and pH value in soil, etc. Data collected from two sensors that are close to each other are often strongly correlated due to the spatial redundancy of the measured object. In this chapter, it is assumed that the correlation between two data samples can be modeled as follows

$$\mathbb{E}\left[x(\boldsymbol{\eta}_p)x(\boldsymbol{\eta}_q)\right] = \rho^{\|\boldsymbol{\eta}_p - \boldsymbol{\eta}_q\|}, \tag{2.1}$$

where $\rho \in [0,1]$ is the spatial correlation coefficient, $\mathbb{E}(\cdot)$ represents mathematical expectation, and $\|\mathbf{a}\| = \sqrt{\mathbf{a}\mathbf{a}^T}$ is the L_2-norm of the column vector $\mathbf{a}$ with $(\cdot)^T$ representing matrix transpose. In (2.1), $\|\boldsymbol{\eta}_p - \boldsymbol{\eta}_q\|$ measures the Euclidean distance between the two points with coordinates $\boldsymbol{\eta}_p$ and $\boldsymbol{\eta}_q$.

It is assumed that sensors deliver the measured data to the FC through an orthogonal media access control (MAC) scheme, such as the frequency division multiplexing access, such that collision-free communication is achieved at the FC. The signal observed by the FC from the n-th sensor node is

$$y_n = \sqrt{P_n}x(\boldsymbol{\eta}_n) + z_n, \tag{2.2}$$

where P_n is the average transmission power, or averange energy per sample, of the n-th node, and z_n is the additive white Gaussian noise (AWGN) with variance σ_z^2. The system model does not consider the effects of fading or pathloss. The subsequent analysis can be easily extended to systems with fading. It is assumed that the total power per unit area is fixed at P_0 to ensure a fair comparison among networks with different node densities and different sizes. Given a network with node density δ, the power per node is then $P_n = \frac{P_0}{\delta}$.

2.3.2 Optimum MMSE Detection

The FC will obtain an estimate of the spatially continuous quantity, $x(\boldsymbol{\eta})$, $\forall \boldsymbol{\eta} \in \Omega_{\boldsymbol{\eta}}$, over the entire measurement field, by extracting information transmitted from the N sensors, $\mathbf{y} = [y_1, \cdots, y_N]^T \in \mathscr{R}^{N \times 1}$, where $\mathscr{R}$ is the set of real numbers. The MSE at any location $\boldsymbol{\eta}$ is

$$\sigma_{\boldsymbol{\eta}}^2 = \mathbb{E}\left[\hat{x}(\boldsymbol{\eta}) - x(\boldsymbol{\eta})\right]^2, \boldsymbol{\eta} \in \Omega_{\boldsymbol{\eta}} \tag{2.3}$$

where $\hat{x}(\boldsymbol{\eta})$ is an estimate of $x(\boldsymbol{\eta})$ at the FC.

The optimum linear receiver that minimizes $\sigma_{\boldsymbol{\eta}}^2$ is the MMSE receiver described as follows [18]

$$\hat{x}(\boldsymbol{\eta}) = \sqrt{P_n}\mathbf{r}_{\boldsymbol{\eta}}^T \left(P_n\mathbf{R}_{ss} + \sigma_z^2\mathbf{I}_N\right)^{-1}\mathbf{y}, \tag{2.4}$$

where $\mathbf{r}_{\boldsymbol{\eta}} = \mathbb{E}[x(\boldsymbol{\eta})\mathbf{x}_s] = [\rho^{\|\boldsymbol{\eta}-\boldsymbol{\eta}_1\|}, \cdots, \rho^{\|\boldsymbol{\eta}-\boldsymbol{\eta}_N\|}]^T \in \mathscr{R}^{N \times 1}$ with $\mathbf{x}_s = [x(\boldsymbol{\eta}_1), \cdots, x(\boldsymbol{\eta}_N)]^T$, $\mathbf{R}_{ss} = \mathbb{E}[\mathbf{x}_s \mathbf{x}_s^T] \in \mathscr{R}^{N \times N}$ with the (p,q)-th element being $\rho^{\|\boldsymbol{\eta}_p - \boldsymbol{\eta}_q\|}$ as defined in (2.1), and $\mathbf{I}_N$ is a size-N identity matrix.

With the optimum MMSE receiver given in (2.4), the MSE $\sigma_{\boldsymbol{\eta}}^2$ can be calculated as [18]

$$\sigma_{\boldsymbol{\eta}}^2 = 1 - \mathbf{r}_{\boldsymbol{\eta}}^T \left(\mathbf{R}_{ss} + \frac{\delta}{\gamma_0} \mathbf{I}_N \right)^{-1} \mathbf{r}_{\boldsymbol{\eta}}, \tag{2.5}$$

where $\gamma_0 = \frac{P_0}{\sigma_z^2}$ is the SNR per unit area. The MSE $\sigma_{\boldsymbol{\eta}}^2$ given in (2.5) is a function of the location $\boldsymbol{\eta}$, the SNR γ_0, the spatial correlation coefficient ρ, and the node density δ.

Given a fixed transmission power per unit area, the node density, δ, plays a critical role on the MSE $\sigma_{\boldsymbol{\eta}}^2$. A smaller node density means more transmission power per node, thus a better SNR per sample. On the other hand, a smaller node density means less samples per unit area, or a smaller correlation among the samples, and this will degrade the estimation accuracy.

In order to distinguish the opposite impacts of the node density on the SNR per sample and the spatial sample correlation, we decompose the MMSE receiver described in (2.4) into two steps as follows.

Definition 2.1: **Two-Step MMSE**:

1) The FC first obtains an estimate of the data at the sensor locations: $\mathbf{x}_s = [x(\boldsymbol{\eta}_1), \cdots, x(\boldsymbol{\eta}_N)]^T \in \mathscr{R}^{N \times 1}$, with a linear MMSE receiver as

$$\hat{\mathbf{x}}_s = \mathbf{W}_s^T \mathbf{y}, \tag{2.6}$$

where $\hat{\mathbf{x}}_s = [\hat{x}(\boldsymbol{\eta}_1), \cdots, \hat{x}(\boldsymbol{\eta}_N)]^T$ is an estimate of $\mathbf{x}_s$. The MMSE matrix $\mathbf{W}_s \in \mathscr{R}^{N \times N}$ is designed to minimize the MSE per node:

$$\sigma_{s,N}^2 = \frac{1}{N} \mathbb{E}\left[\|\hat{\mathbf{x}}_s - \mathbf{x}_s\|^2 \right]. \tag{2.7}$$

2) The FC obtains an estimate of the data at an arbitrary location, $\hat{x}(\boldsymbol{\eta})$, $\forall \boldsymbol{\eta} \in \Omega_{\boldsymbol{\eta}}$, by interpolating $\hat{\mathbf{x}}_s$ with the MMSE criterion,

$$\hat{x}(\boldsymbol{\eta}) = \mathbf{w}_{sl}^T \hat{\mathbf{x}}_s, \tag{2.8}$$

where the vector, $\mathbf{w}_{sl} \in \mathscr{R}^{N \times 1}$, is designed to minimize the MSE $\sigma_{\boldsymbol{\eta}}^2 = \mathbb{E}\left[\hat{x}(\boldsymbol{\eta}) - x(\boldsymbol{\eta})\right]^2$.

Lemma 2.1: The two-step MMSE receiver described in Definition 2.1 is equivalent to the optimum MMSE given in (2.4).

Proof: The proof is in Appendix 2.7.1. ∎

Decomposing the optimum MMSE of (2.4) into the two-step MMSE allows us to study the two opposite effects of the node density on the MSE separately. In the next two sections, we will investigate, respectively, the impacts of the node density on the 1-D and 2-D networks by following the two-step MMSE.

2.4 Optimum Node Density in One-dimensional Networks

In this section, we study the optimum node density in an 1-D network, where the N sensor nodes are evenly distributed over a length-L linear section as shown in Fig. 2.1(a). In a linear network, the n-th node is placed at a location with the coordinate $\eta_n = \eta_n = (n-1)d$. Following the two-step MMSE given in Definition 2.1, we will study in the next two subsections the impacts of node density on the performance of the two steps: MMSE estimation of the data at the sensor locations, and MMSE interpolation for the estimation of the data at arbitrary locations.

2.4.1 MMSE Estimation at Sensor Locations

We study in this subsection the performance of the MMSE estimation of the data at the sensor locations as described in (2.6), which corresponds to the first step of the two-step MMSE estimation.

For the MMSE estimation described in (2.6), the optimum $\mathbf{W}_s$ that minimizes the MSE, $\sigma_{s,N}^2$, can be found through the orthogonal principal, $\mathbb{E}\left[(\hat{\mathbf{x}}_s - \mathbf{x}_s)\mathbf{y}^T\right] = 0$. The result is

$$\mathbf{W}_s^T = \sqrt{P_n}\mathbf{R}_{ss}\left(P_n\mathbf{R}_{ss} + \sigma_z^2\mathbf{I}_N\right)^{-1}. \tag{2.9}$$

The error correlation matrix, $\mathbf{R}_{ee}^{(s)} = \mathbb{E}\left[\mathbf{e}_s\mathbf{e}_s^T\right]$, with $\mathbf{e}_s = \hat{\mathbf{x}}_s - \mathbf{x}_s$, can be calculated by

$$\mathbf{R}_{ee}^{(s)} = \mathbf{R}_{ss} - \mathbf{R}_{ss}\left(\mathbf{R}_{ss} + \frac{\delta}{\gamma_0}\mathbf{I}_N\right)^{-1}\mathbf{R}_{ss} = \left(\mathbf{R}_{ss}^{-1} + \frac{\gamma_0}{\delta}\mathbf{I}_N\right)^{-1}, \tag{2.10}$$

where the orthogonal principal is used in the first equality, and the second equality is based on the identity $\mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} = (\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}$. The MSE can then be calculated as $\sigma_{s,N}^2 = \frac{1}{N}\text{trace}\left(\mathbf{R}_{ee}^{(s)}\right)$, where trace($\mathbf{A}$) returns the trace of the matrix $\mathbf{A}$. The calculation of the MSE involves matrix inversion and the trace operation. Performing the eigenvalue decomposition of $\mathbf{R}_{ss}$ in (2.10), we can rewrite the MSE as

$$\sigma_{s,N}^2 = \frac{1}{N}\sum_{n=1}^{N}\left(\frac{1}{\lambda_n} + \frac{\gamma_0}{\delta}\right)^{-1}, \tag{2.11}$$

where λ_n is the n-th eigenvalue of $\mathbf{R}_{ss}$.

Fig. 2.2 plots the numerical and simulated MSE of the data samples as a function of node density, δ, under various values of the node number N. The spatial correlation coefficient is $\rho = 0.5$ and the SNR per unit area is $\gamma_0 = 10$ dB. Data samples are assumed to be a zero-mean Gaussian process with the spatial auto-correlation function given in (2.1). It can be seen from the figure that the MSE is an increasing function of the node density. At a given node density, the MSE decreases as N increases. When N is large, $e.g.$, $N \geq 30$, increasing N further only leads to a slight improvement in the MSE. It is thus postulated that the MSE converges to a constant value as $N \to \infty$.
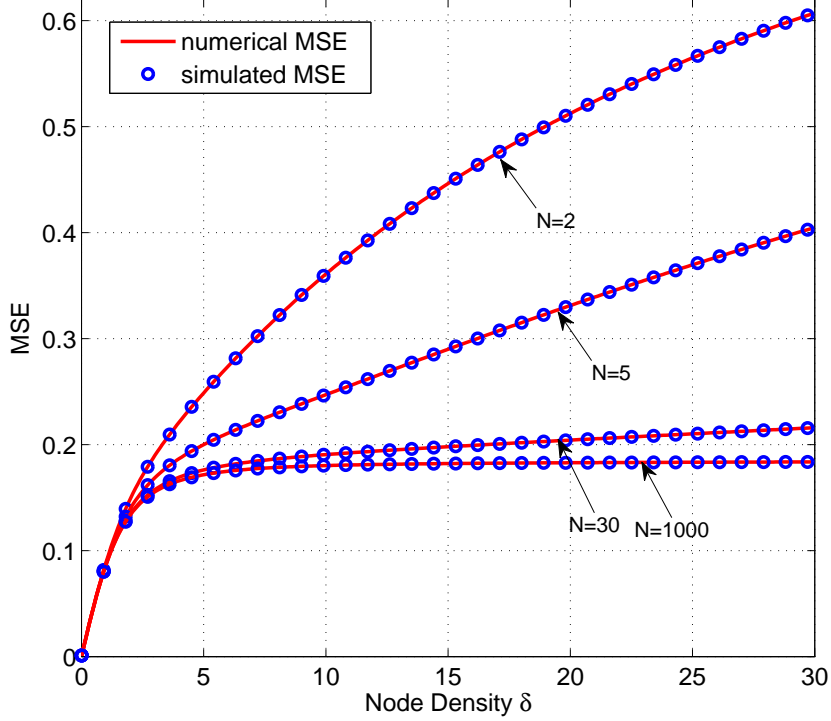
Figure 2.2: The MSE of the data samples under various node number N in an 1-D network ($\rho = 0.5$, $\gamma_0 = 10$ dB).

In order to explicitly identify the impacts of the node density and the spatial data correlation on the MSE, we resort to the asymptotic analysis by letting $N \to \infty$ while keeping a finite node density δ. The results are presented as follows.

Proposition 2.1: When $N \to \infty$ while keeping a finite δ, the asymptotic MSE of the estimated data at the sensor locations in an 1-D network is

$$\sigma_s^2 \triangleq \lim_{N \to \infty} \sigma_{s,N}^2 = \left[\left(1 + \frac{\gamma_0}{\delta}\right)^2 + \frac{4\gamma_0 \rho^{\frac{2}{\delta}}}{\delta \left(1 - \rho^{\frac{2}{\delta}}\right)} \right]^{-\frac{1}{2}}. \tag{2.12}$$

Proof: The proof is given in Appendix 2.7.2. ■

In (2.12), the opposite effects of node density on the asymptotic MSE are manifested in the form of two functions, $f_1(\delta) \triangleq \frac{\gamma_0}{\delta}$, and $f_2(\delta) \triangleq \frac{\rho^{\frac{2}{\delta}}}{1 - \rho^{\frac{2}{\delta}}}$. The first function, $f_1(\delta)$, can be interpreted

20
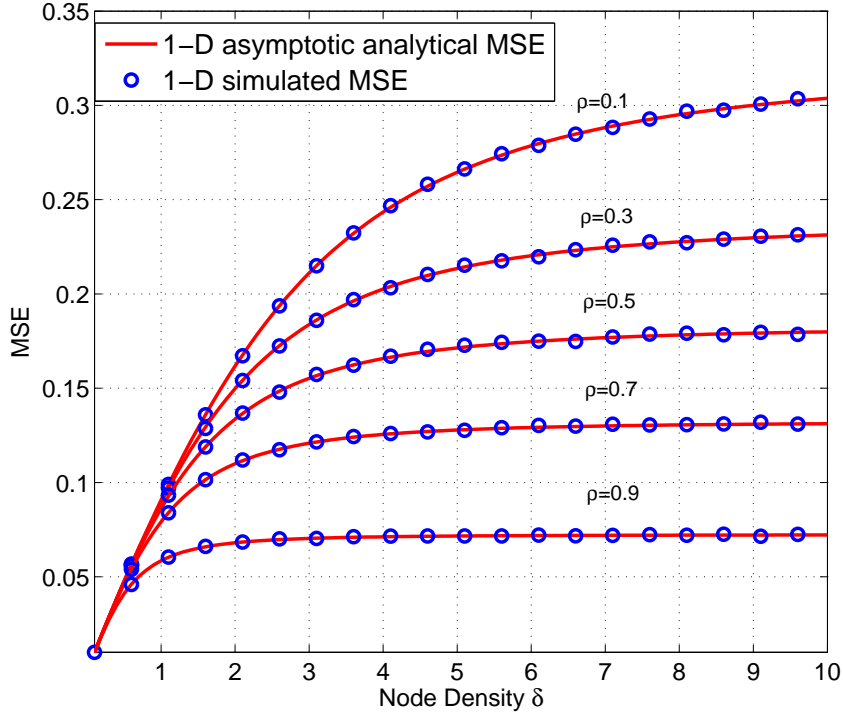
Figure 2.3: The asymptotic MSE of the data samples in an 1-D network ($\gamma_0 = 10$ dB).

as the SNR per node, which is inversely proportional to δ. Thus $f_1(\delta)$ translates a positive correlation between δ and the asymptotic MSE. The second function, $f_2(\delta)$, is related to the spatial correlation among sensors, and it is an increasing function of δ. Hence, $f_2(\delta)$ translates a negative correlation between δ and the asymptotic MSE. Therefore, δ exhibits two opposite effects on MSE through $f_1(\delta)$ and $f_2(\delta)$. For the estimation of data at the sensor node locations, it is shown in the following corollary that the effect of the SNR per node, $f_1(\delta)$, dominates that of the spatial correlation, $f_2(\delta)$.

Corollary 2.1: The asymptotic MSE given in (2.12) is a monotonic increasing function of the node density, δ.

Proof: The proof is in Appendix 2.7.3. ∎

The result in Corollary 2.1 indicates that, the asymptotic MSE for estimating data at the sensor

node locations can benefit from a smaller density. *Therefore, if we only want to obtain the data at some discrete locations, we should use a node density that is as small as allowed by the application, i.e., placing exactly one sensor at each desired measurement location will obtain the optimum performance.*

Fig. 2.3 shows the asymptotic MSE as a function of the node density, δ, under various values of the spatial correlation coefficient, ρ. The SNR per unit area is $\gamma_0 = 10$ dB. The simulation results are obtained by using $N = 1,000$ nodes to approximate infinite nodes. Excellent agreement is observed between the asymptotic analytical results with $N \to \infty$ and the simulation results with $N = 1000$. As pointed out by Corollary 2.1, the MSE increases monotonically as δ increases, indicating the dominance of the SNR per node over the spatial correlation. It can be seen from Fig. 2.3 that the MSE approaches a constant value as $\delta \to \infty$. This indicates a balance between the opposite effects between $f_1(\delta)$ and $f_2(\delta)$ as $\delta \to \infty$, which is corroborated by the following corollary.

Corollary 2.2: For the estimation of the data at the sensor locations, the asymptotic MSE is upper bounded by

$$\sigma_s^2 \leq \left(1 - \frac{2\gamma_0}{\log \rho}\right)^{-\frac{1}{2}} \tag{2.13}$$

Proof: Eqn. (2.13) can be directly proved by setting $\lim_{\delta \to \infty} \sigma_s^2$ in (2.12). ∎

The asymptotic MSE upper bound is determined by the spatial correlation and the SNR per unit area. No matter how low the SNR per node is, the MSE is always upper bounded due to the spatial correlation among the sensors.
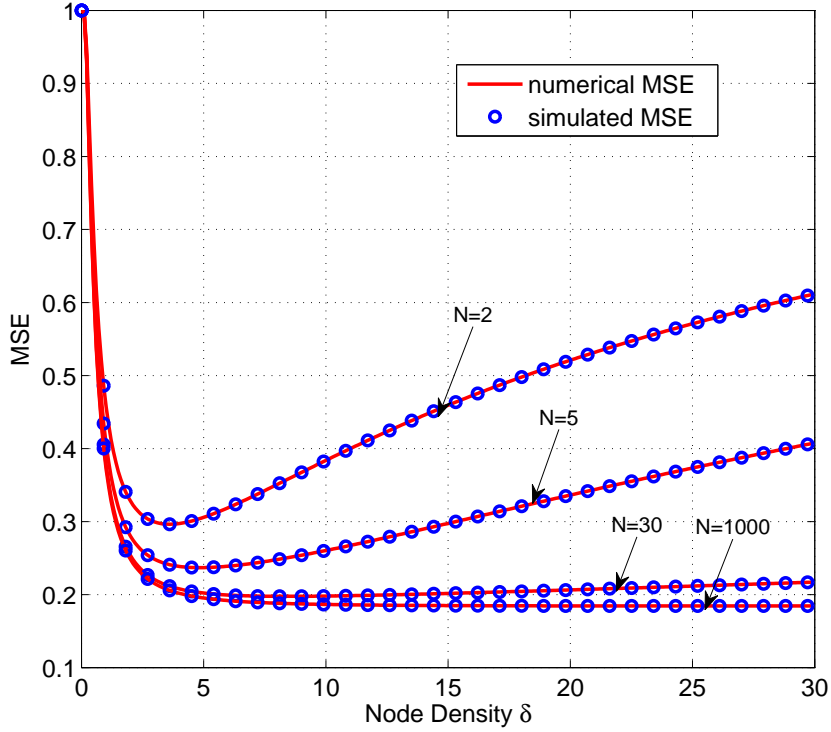
Figure 2.4: The MSE of the spatial interpolations under various node number N in an 1-D network ($\rho = 0.5$, $\gamma_0 = 10$ dB).

2.4.2 MMSE Spatial Interpolation

Once the estimates of the data at the sensor locations are obtained, they can be interpolated to get an estimate of the data at any arbitrary location in the measurement field. The impacts of the node density on the spatial interpolation are investigated in this subsection.

As discussed in Definition 2.1 and Lemma 3.3.2, MMSE spatial interpolation can obtain the optimum performance by minimizing σ_η^2 given in (2.3). The MSE given in (2.3) depends on the location η. Since we are interested in the reconstruction fidelity of the entire measurement field, the worst case scenario will be considered by estimating the data located in the middle of two neighboring sensors, with coordinate $\eta'_n = (n - \frac{1}{2})d$. Correspondingly, denote the data vector to be estimated through interpolation as $\mathbf{x}_d = [x(\eta'_1), \cdots, x(\eta'_N)]^T \in \mathscr{R}^{N \times 1}$.

Following the orthogonal principal, $\mathbb{E}\left[(\hat{\mathbf{x}}_d - \mathbf{x}_d)\hat{\mathbf{x}}_s^T\right] = \mathbf{0}$, where $\hat{\mathbf{x}}_d$ is an estimate of $\mathbf{x}_d$, the MMSE spatial interpolation can be expressed by

$$\hat{\mathbf{x}}_d = \mathbf{R}_{d\hat{s}}\mathbf{R}_{\hat{s}\hat{s}}^{-1}\hat{\mathbf{x}}_s, \tag{2.14}$$

where

$$\mathbf{R}_{d\hat{s}} \triangleq \mathbb{E}(\mathbf{x}_d\hat{\mathbf{x}}_s^T) = \sqrt{P_n}\mathbf{R}_{ds}\mathbf{W}_s, \tag{2.15a}$$

$$\mathbf{R}_{\hat{s}\hat{s}} \triangleq \mathbb{E}(\hat{\mathbf{x}}_s\hat{\mathbf{x}}_s^T) = \mathbf{W}_s^T(P_n\mathbf{R}_{ss} + \sigma_z^2\mathbf{I})\mathbf{W}_s, \tag{2.15b}$$

with $\mathbf{R}_{ds} \triangleq \mathbb{E}(\mathbf{x}_d\hat{\mathbf{x}}_s)$ being a Toeplitz matrix expressed as follows

$$\mathbf{R}_{ds} = \rho^{\frac{d}{2}}\begin{bmatrix} \rho^0 & \rho^0 & \rho^d & \cdots & \rho^{(N-2)d} \\ \rho^d & \rho^0 & \rho^0 & \cdots & \rho^{(N-3)d} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \rho^{(N-2)d} & \rho^{(N-3)d} & \cdots & \rho^0 & \rho^0 \\ \rho^{(N-1)d} & \rho^{(N-2)d} & \cdots & \rho^d & \rho^0 \end{bmatrix}. \tag{2.16}$$

Combining (2.14) with (2.15), we have

$$\hat{\mathbf{x}}_d = \sqrt{P_n}\mathbf{R}_{ds}\left(P_n\mathbf{R}_{ss} + \sigma_z^2\mathbf{I}\right)^{-1}\mathbf{y}. \tag{2.17}$$

The corresponding error correlation matrix, $\mathbf{R}_{ee}^{(d)} \triangleq \mathbb{E}\left[(\hat{\mathbf{x}}_d - \mathbf{x}_d)(\hat{\mathbf{x}}_d - \mathbf{x}_d)^T\right]$, can then be calculated by

$$\mathbf{R}_{ee}^{(d)} = \mathbf{R}_{ss} - \mathbf{R}_{ds}\left(\mathbf{R}_{ss} + \frac{\delta}{\gamma_0}\mathbf{I}_N\right)^{-1}\mathbf{R}_{sd}, \tag{2.18}$$

where $\mathbf{R}_{dd} = \mathbb{E}(\mathbf{x}_d\mathbf{x}_d^T) = \mathbf{R}_{ss}$ is used in the above equation, and $\mathbf{R}_{sd} = \mathbf{R}_{ds}^T$. The MSE for spatial interpolation is $\sigma_{d,N}^2 = \frac{1}{N}\text{trace}\left(\mathbf{R}_{ee}^{(d)}\right)$.
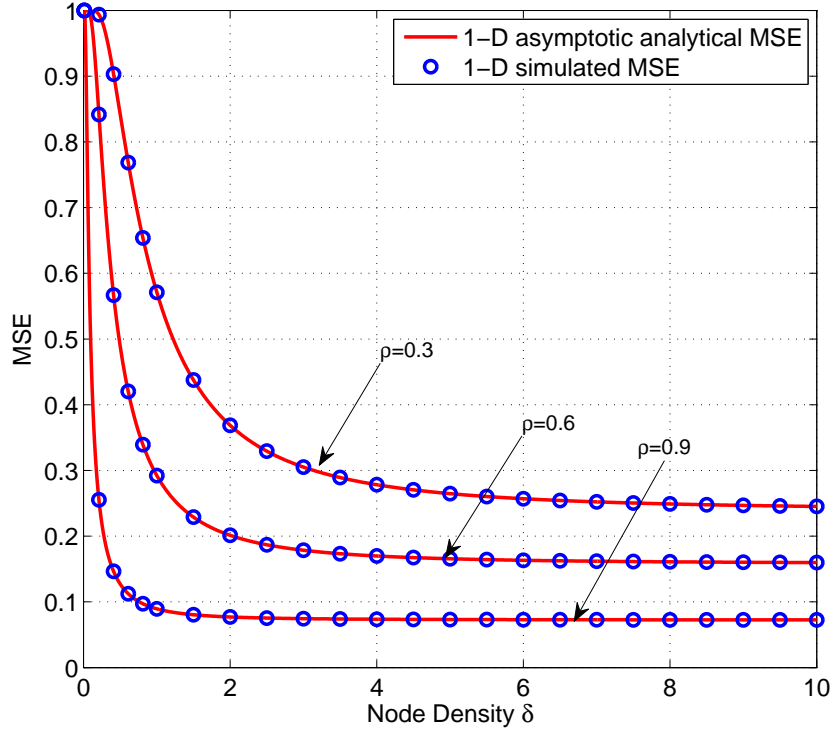
Figure 2.5: The asymptotic MSE of the spatial interpolation in an 1-D network ($\gamma_0 = 10$ dB).

The MSE for the spatial interpolation is numerically evaluated for different values of δ and N, and the results are shown in Fig. 2.4. All the parameters are the same as those in Fig. 2.2. When N is small, *e.g.*, $N < 30$, the MSE is convex in δ. When N is large, *e.g.*, $N \geq 30$, the spatial interpolation MSE becomes a monotonically decreasing function of δ. In addition, similar to the results in Fig. 2.2, increasing N from 30 to 1000 only leads to a marginal performance improvement. Such an observation further necessitates the asymptotic study of the system behavior as $N \rightarrow \infty$, and the results are given in the following proposition.

Proposition 2.2: When $N \rightarrow \infty$ while keeping a finite δ, the MSE of the MMSE interpolation in an 1-D network is

$$\sigma_d^2 \triangleq \lim_{N \rightarrow \infty} \sigma_{d,N}^2 = \left(\frac{\delta}{\gamma_0} + \frac{1 - \rho^{\frac{1}{\delta}}}{1 + \rho^{\frac{1}{\delta}}} \right)^{\frac{1}{2}} \left(\frac{\delta}{\gamma_0} + \frac{1 + \rho^{\frac{1}{\delta}}}{1 - \rho^{\frac{1}{\delta}}} \right)^{-\frac{1}{2}}. \tag{2.19}$$

Proof: The proof is in Appendix 2.7.4. ∎

Define $f_3(\delta) \triangleq \frac{1+\rho^{\frac{1}{\delta}}}{1-\rho^{\frac{1}{\delta}}}$. It can be easily shown that $f_3(\delta)$ is an increasing function of δ, and its impact on the asymptotic MSE is opposite to the SNR per node, $f_1(\delta)$. The following corollary shows that σ_d^2 is dominated by the effects of $f_3(\delta)$.

Corollary 2.3: The asymptotic MSE given in (2.19) is a monotonic decreasing function of the node density, δ.

Proof: The proof is in Appendix 2.7.5. ∎
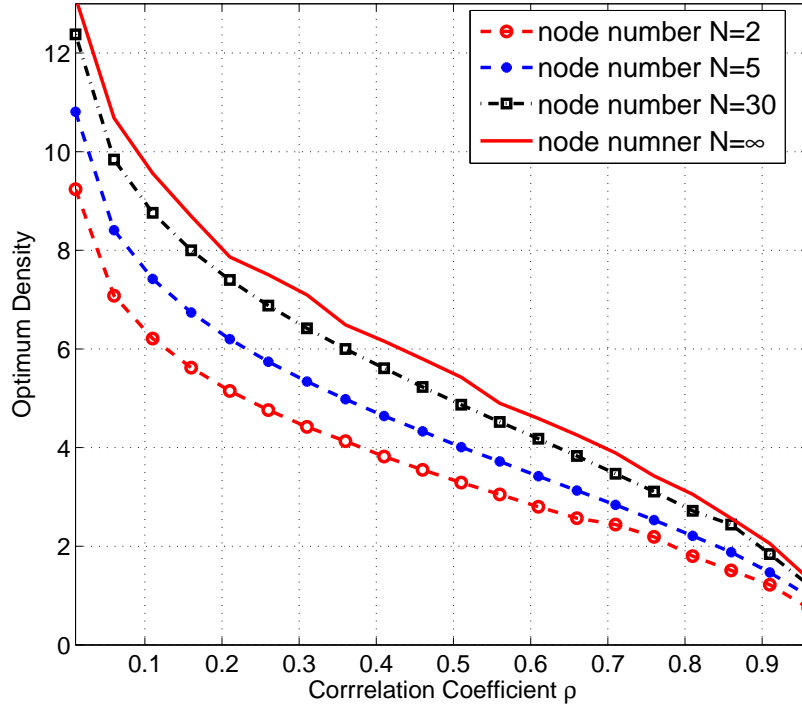
Figure 2.6: Optimum node density v.s. correlation coefficient ρ in an 1-D network ($\gamma_0 = 10$ dB).

From Corollaries 2.1 and 2.3, it is apparent that δ has opposite impacts on σ_s^2 and σ_d^2. The results in Corollary 2.3 can be intuitively explained by the fact that the spatial interpolation depends mainly on the spatial correlation among the data samples, and a higher density means a stronger correlation among the data samples, thus a better estimation fidelity.

26

The asymptotic MSE of the data interpolation is shown in Fig. 2.5, where it is apparent that σ_d^2 is a decreasing function of δ. The simulation parameters are the same as those in Fig. 2.3. Again, perfect agreement is observed between the simulation results with $N = 1000$ and asymptotic analytic results with $N \to \infty$. When $\delta \to \infty$, σ_d^2 is lower bounded, and this is described in the following corollary.

Corollary 2.4: The following relationship holds for σ_s^2 and σ_d^2

$$\sigma_d^2 \geq \left(1 - \frac{2\gamma_0}{\log \rho} \right)^{-\frac{1}{2}} \geq \sigma_s^2 \tag{2.20}$$

Proof: Since σ_d^2 is a decreasing function of δ, its minimum value can be obtained by letting $\delta \to \infty$ in (2.19), and (2.20) follows immediately. ∎

The result in (2.20) indicates that σ_d^2 is always bigger than σ_s^2 and they converge when $\delta \to \infty$. This can be explained by the fact that the estimation of $\mathbf{x}_d$ is based on the estimation accuracy of $\mathbf{x}_s$, thus the fidelity of $\hat{\mathbf{x}}_d$ can not exceed that of $\hat{\mathbf{x}}_s$. *This result further corroborates that, for the estimation of data at a discrete location, a sensor node needs to be placed at the desired location to ensure the optimum performance, because interpolation will always lead to an inferior performance.*

It is observed from Fig. 2.5 that, when δ is small, the MSE decreases dramatically as δ increases. When δ reaches a certain threshold, no apparent performance gain can be achieved by increasing δ further, *i.e.*, the slope of σ_d^2 approaches zero as δ increases. Therefore, the optimum node density can be chosen as the point such that $\left| \frac{\partial \sigma_d^2}{\partial \delta} \right| \leq \varepsilon$, with ε being a small number.

From (2.19), the slope of σ_d^2 can be calculated as

$$\frac{\partial \sigma_d^2}{\partial \delta} = \frac{\sigma_d^2}{2} \cdot \left[\frac{\frac{1}{\gamma_0} + \frac{2\log\rho \cdot \rho^{\frac{1}{\delta}}}{\delta^2 \cdot (1+\rho^{\frac{1}{\delta}})^2}}{\frac{\delta}{\gamma_0} + \frac{1-\rho^{\frac{1}{\delta}}}{1+\rho^{\frac{1}{\delta}}}} - \frac{\frac{1}{\gamma_0} - \frac{2\log\rho \cdot \rho^{\frac{1}{\delta}}}{\delta^2 \cdot (1-\rho^{\frac{1}{\delta}})^2}}{\frac{\delta}{\gamma_0} + \frac{1+\rho^{\frac{1}{\delta}}}{1-\rho^{\frac{1}{\delta}}}} \right] \tag{2.21}$$

The optimum node density can then be obtained by numerically solving the equation $\left| \frac{\partial \sigma_d^2}{\partial \delta} \right|_{\delta_0} = \varepsilon$.

Fig. 2.6 shows the optimum node density as functions of the spatial correlation coefficient ρ and the number of nodes N. The SNR per unit area is $\gamma_0 = 10$ dB. When N is small ($N = 2$ and $N = 5$), the MSE is convex in δ as shown in Fig. 2.4, and the optimum node density is obtained by finding the value of δ that minimizes the MSE. When N is large ($N = 30$ and $N \to \infty$), the optimum node density is obtained by solving $\left| \frac{\partial \sigma_d^2}{\partial \delta} \right| \leq \varepsilon$ with $\varepsilon = 10^{-3}$. The optimum node density decreases as ρ increases for all the systems. This is intuitive because a field with stronger spatial correlation can operate with a smaller node density. In addition, given δ, the optimum node density increases as N increases, and it is upper bounded by the asymptotic case with $N \to \infty$. The optimum node density for $N = 30$ is very close to the asymptotic results with $N \to \infty$.

2.5 Optimum Node Density in Two-dimensional Networks

The impacts of node density on the estimation fidelity in a 2-D network are studied in this section. In a 2-D network, consider the N sensor nodes located over a grid on a square region with area $\sqrt{N}d \times \sqrt{N}d$ as shown in Fig. 2.1(b). The coordinates for the nodes are $\boldsymbol{\eta}_{ij} = [id, jd]^T$, for $i, j = 0, \cdots, K-1$ with $K = \sqrt{N}$. The correlation between any two data samples is $\mathbb{E}\left[x(\boldsymbol{\eta}_{ij})x(\boldsymbol{\eta}_{mn})\right] = \rho^{\|\boldsymbol{\eta}_{ij} - \boldsymbol{\eta}_{mn}\|}$. It should be noted that the node density in a 2-D network is $\delta = \frac{1}{d^2}$, which is different from the 1-D case.

Stacking the data from all the sensors into a column vector, we have $\boldsymbol{\xi}_s = [\mathbf{x}_0^T, \cdots, \mathbf{x}_{K-1}^T]^T \in$

$\mathscr{R}^{N\times 1}$, where $\mathbf{x}_m = [x(\boldsymbol{\eta}_{m0}), \cdots, x(\boldsymbol{\eta}_{m(K-1)})]^T \in \mathscr{R}^{K\times 1}$. The auto-correlation matrix, $\blacksquare_{ss} = \mathbb{E}\left[\boldsymbol{\xi}_s\boldsymbol{\xi}_s^T\right] \in$

$\mathscr{R}^{N\times N}$, can be expressed in the form of a block Toeplitz matrix as

$$\blacksquare_{ss} = \begin{pmatrix} \mathbf{R}_0 & \mathbf{R}_1 & \cdots & \mathbf{R}_{K-1} \\ \mathbf{R}_1 & \mathbf{R}_0 & \cdots & \mathbf{R}_{K-2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_{K-1} & \mathbf{R}_{K-2} & \cdots & \mathbf{R}_0 \end{pmatrix}, \tag{2.22}$$

where $\mathbf{R}_m = \mathbb{E}(\mathbf{x}_m\mathbf{x}_0) \in \mathscr{R}^{K\times K}$ is a symmetric Toeplitz matrix with the first row and first column

being $\mathbf{r}_m = [r_{m,0}, r_{m,1}, \cdots, r_{m,K-1}]^T \in \mathscr{R}^{k\times 1}$, and $r_{m,k} = \mathbb{E}[x(\boldsymbol{\eta}_{mk})x(\boldsymbol{\eta}_{00})]$ can be expressed as

$$r_{m,k} = \rho^{\|\boldsymbol{\eta}_{mk}-\boldsymbol{\eta}_{00}\|} = \rho^{\sqrt{\frac{k^2+m^2}{\delta}}}. \tag{2.23}$$

The matrix, $\blacksquare_{ss}$, assumes the form of a Toeplitz-block-Toeplitz (TBT) matrix [17], *i.e.*, $\blacksquare_{ss}$ is a

block Toeplitz matrix, and each sub-matrix is also a Toeplitz matrix.

Mirroring the analysis in the 1-D case, we will study, in the next two subsections, the MMSE

estimation at the sensor locations, and the MMSE interpolation for the estimation at arbitrary

locations, respectively.

2.5.1 MMSE Estimation at Sensor Locations

Following the same procedure as in Section 2.4.1, we have the MSE, $\vartheta_{s,N}^2 \triangleq \frac{1}{N}\mathbb{E}\left[\|\hat{\boldsymbol{\xi}}_s - \boldsymbol{\xi}_s\|^2\right]$, as

$$\vartheta_{s,N}^2 = \frac{1}{N}\text{trace}\left(\blacksquare_{ss}^{-1} + \frac{\gamma_0}{\delta}\mathbf{I}_N\right)^{-1}, \tag{2.24}$$

where $\hat{\boldsymbol{\xi}}_s$ is the MMSE estimate of $\boldsymbol{\xi}_s$. Performing the eigenvalue decomposition of $\blacksquare_{ss}$ in (2.22),

we can rewrite the MSE in (2.24) as

$$\vartheta_{s,N}^2 = \frac{1}{N}\sum_{k=1}^{K-1}\sum_{m=1}^{K-1}\left(\frac{1}{\lambda_{k,m}} + \frac{\gamma_0}{\delta}\right)^{-1}, \tag{2.25}$$
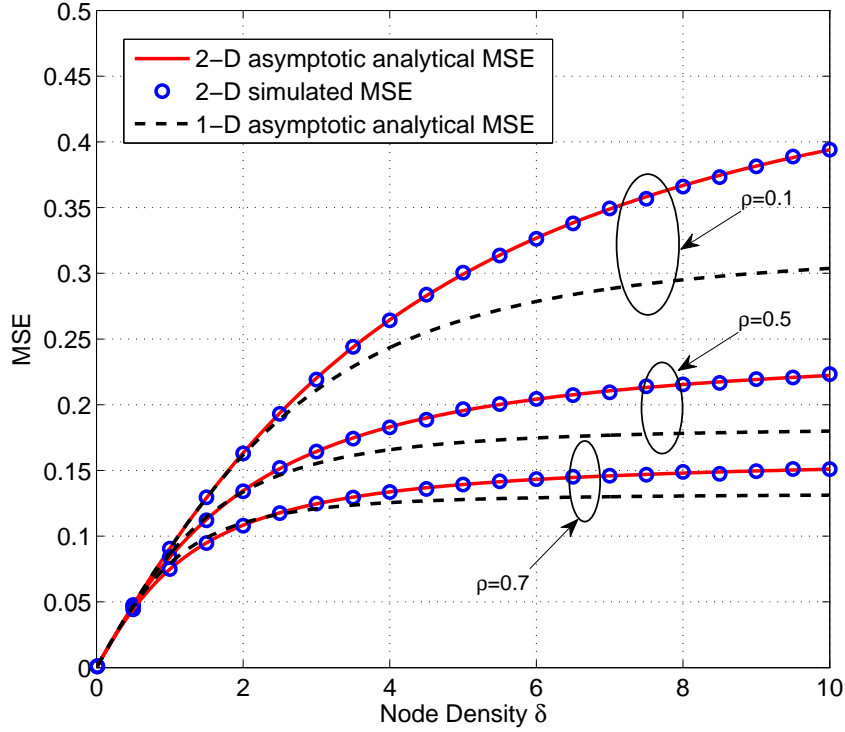
Figure 2.7: The asymptotic MSE of the data samples estimation in the 1-D and 2-D networks ($\gamma_0 = 10$ dB).

where $\lambda_{k,m}$, for $k, m = 0, 1, \cdots, K-1$, are eigenvalues of $\blacksquare_{ss}$.

The asymptotic behavior of $\vartheta_{s,N}^2$ when $N \to \infty$ at finite node density is presented in the following proposition.

Proposition 2.3: When $N \to \infty$ while keeping a finite node density δ, the asymptotic MSE of the estimated data at the sensor locations in a 2-D network with spatial correlation coefficient ρ is

$$\vartheta_s^2 \triangleq \lim_{N \to \infty} \vartheta_{s,N}^2 = \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} \left[\frac{1}{\Lambda_{ss}'(f_1, f_2)} + \frac{\gamma_0}{\delta} \right]^{-1} df_1 df_2, \tag{2.26}$$

where

$$\Lambda_{ss}'(f_1, f_2) = \sum_{k=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} \rho^{\sqrt{\frac{k^2+m^2}{\delta}}} e^{-j2\pi(kf_1+mf_2)}. \tag{2.27}$$
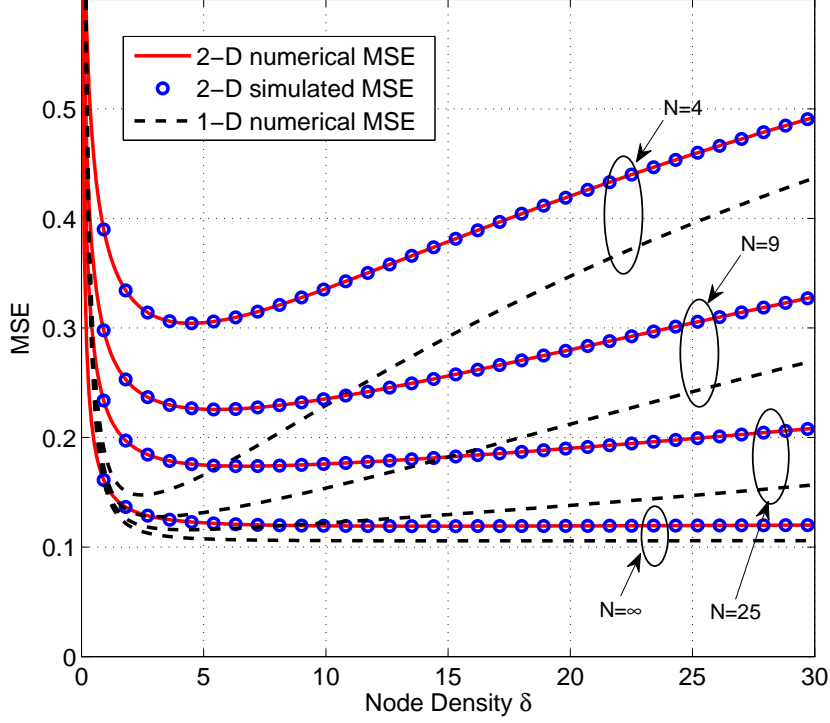
Figure 2.8: The MSE of the spatial interpolations under various node number N in the 1-D and 2-D networks ($\rho = 0.8$, $\gamma_0 = 10$ dB).

Proof: The results in (2.26) can be obtained by applying [17, Theorem 1] to (2.25), which is the extension of the Szego's theorem to TBT matrices. ∎

In Proposition 2.3, the impacts of the node density are expressed through the SNR per node, $\frac{\gamma_0}{\delta}$, and the 2-D DTFT, $\Lambda'_{ss}(f_1, f_2)$. The expression of ϑ_s^2 eludes a closed-form. The expression in (2.26) can be easily evaluated numerically given that the integrals are of finite limits. Even though $\Lambda'_{ss}(f_1, f_2)$ is expressed as the sum of an infinite series, the value of $\rho^{\sqrt{\frac{k^2+m^2}{\delta}}}$ decreases exponentially as k and m increase, thus $\Lambda'_{ss}(f_1, f_2)$ can be accurately approximated with moderate limits on k and m.

Fig. 2.7 shows the asymptotic MSE ϑ_s^2 as a function of the node density δ. For comparison, the MSE σ_s^2 for an 1-D network is also shown in this figure. The SNR is $\gamma_0 = 10$ dB. Similar to

the 1-D case, the asymptotic MSE ϑ_s^2 is a monotonic increasing function of δ. In Fig. 2.7, given

a fixed node density, the asymptotic 2-D MSE is larger (worse) than the asymptotic 1-D MSE.

Therefore, the 2-D network has a worse estimation fidelity than its 1-D counterpart under the same

node density. This can be explained by the fact that each node in the 2-D network needs to cover

a larger area than its 1-D counterpart. The difference between the 1-D and 2-D cases gradually

diminishes as the spatial correlation coefficient, ρ, increases, because the impact of node distance

decreases as ρ getting close to 1.

2.5.2 MMSE Spatial Interpolation

The performance of the spatial interpolation in a 2-D network is studied in this subsection. Similar

to the 1-D case, during spatial interpolation, we consider the worst case scenario by interpolating

the data located in the middle of the square formed by four adjacent sensor nodes, with coordi-

nates of the data points to be estimated being $\boldsymbol{\eta}'_{ij} = [(i + \frac{1}{2})d, (j + \frac{1}{2})d]^T$, for $i, j = 0, \cdots, K-1$.

Correspondingly, the data vector can be expressed as $\boldsymbol{\xi}_d = [\mathbf{x}'^T_0, \cdots, \mathbf{x}'^T_{K-1}]^T \in \mathscr{R}^{N \times 1}$, where $\mathbf{x}'_m =$

$[x(\boldsymbol{\eta}'_{m0}), \cdots, x(\boldsymbol{\eta}'_{m(K-1)})]^T \in \mathscr{R}^{K \times 1}$.

Following the same procedure as in the 1-D case, the error correlation matrix, $\blacksquare^{(d)}_{ee} = \mathbb{E}[(\hat{\boldsymbol{\xi}}_d - \boldsymbol{\xi}_d)(\hat{\boldsymbol{\xi}}_d - \boldsymbol{\xi}_d)^T]$,

with $\hat{\boldsymbol{\xi}}_d$ being the MMSE estimate of $\boldsymbol{\xi}_d$, can be calculated by

$$\blacksquare^{(d)}_{ee} = \blacksquare_{ss} - \blacksquare_{ds} \left(\blacksquare_{ss} + \frac{\delta}{\gamma_0} \mathbf{I}_N \right)^{-1} \blacksquare_{sd}, \tag{2.28}$$

where $\blacksquare_{dd} = \mathbb{E}(\boldsymbol{\xi}_d \boldsymbol{\xi}_d^T) = \blacksquare_{ss}$ is used in the above equation, $\blacksquare_{ds} = \mathbb{E}\left[\boldsymbol{\xi}_d \boldsymbol{\xi}_s^T\right]$, and $\blacksquare_{ds} = \blacksquare^T_{sd}$. The

cross-correlation matrix, $\blacksquare_{ds}$, can be expressed as

$$\blacksquare_{ds}=\begin{pmatrix} \mathbf{R}'_0 & \mathbf{R}'_0 & \mathbf{R}'_1 & \cdots & \mathbf{R}'_{(K-2)} \\ \mathbf{R}'_1 & \mathbf{R}'_0 & \mathbf{R}'_0 & \cdots & \mathbf{R}'_{(K-3)} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{R}'_{(K-2)} & \mathbf{R}'_{(K-3)} & \cdots & \mathbf{R}'_0 & \mathbf{R}'_0 \\ \mathbf{R}'_{(K-1)} & \mathbf{R}'_{(K-2)} & \cdots & \mathbf{R}'_1 & \mathbf{R}'_0 \end{pmatrix}, \tag{2.29}$$

where $\mathbf{R}'_m = \mathbb{E}(\mathbf{x}'_m\mathbf{x}_0) \in \mathscr{R}^{K\times K}$ is a non-symmetric Toeplitz matrix with the first row being $\mathbf{r}'_m = [r'_{m,0}, r'_{m,0}, r'_{m,1}, \cdots, r'_{m,K-2}]^T \in \mathscr{R}^{k\times 1}$ and first column $\mathbf{r}'_m = [r'_{m,0}, r'_{m,1}, \cdots, r'_{m,K-1}]^T \in \mathscr{R}^{k\times 1}$, and $r'_{m,k} = \mathbb{E}[x(\boldsymbol{\eta}'_{mk})x(\boldsymbol{\eta}_{00})]$ can be expressed as

$$r'_{m,k} = \rho^{\|\boldsymbol{\eta}'_{mk}-\boldsymbol{\eta}_{00}\|} = \rho^{\sqrt{\frac{(k+\frac{1}{2})^2+(m+\frac{1}{2})^2}{\delta}}}. \tag{2.30}$$

The matrix $\blacksquare_{ds}$ is in the form of a non-symmetric TBT matrix.

From (2.28), the MSE of the 2-D interpolation in a network with N nodes can be calculated as $\vartheta_{d,N}^2 = \frac{1}{N}\text{trace}\left(\blacksquare_{ee}^{(d)}\right)$. When $N \to \infty$, the asymptotic MSE is given in the following proposition.

Proposition 2.4: When $N \to \infty$ while keeping a finite δ, the asymptotic MSE of the data estimated through spatial interpolations in a 2-D network with the spatial correlation coefficient ρ is

$$\vartheta_d^2 \triangleq \lim_{N\to\infty} \vartheta_{d,N}^2 = \int_{-\frac{1}{2}}^{\frac{1}{2}}\int_{-\frac{1}{2}}^{\frac{1}{2}}\left[\Lambda'_{ss}(f_1,f_2) - \frac{|\Lambda'_{ds}(f_1,f_2)|^2}{\Lambda'_{ss}(f_1,f_2)+\frac{\delta}{\gamma_0}}\right]df_1 df_2, \tag{2.31}$$

where $\Lambda'_{ss}(f_1, f_2)$ is given in (2.27), and $\Lambda'_{ds}(f_1, f_2)$ is computed as

$$\Lambda'_{ds}(f_1,f_2) = \sum_{k=-\infty}^{+\infty}\sum_{m=-\infty}^{+\infty} \rho^{\sqrt{\frac{(k+\frac{1}{2})^2+(m+\frac{1}{2})^2}{\delta}}} e^{-j2\pi(kf_1+mf_2)}. \tag{2.32}$$

Proof: The proof is in Appendix 2.7.6. ■

33
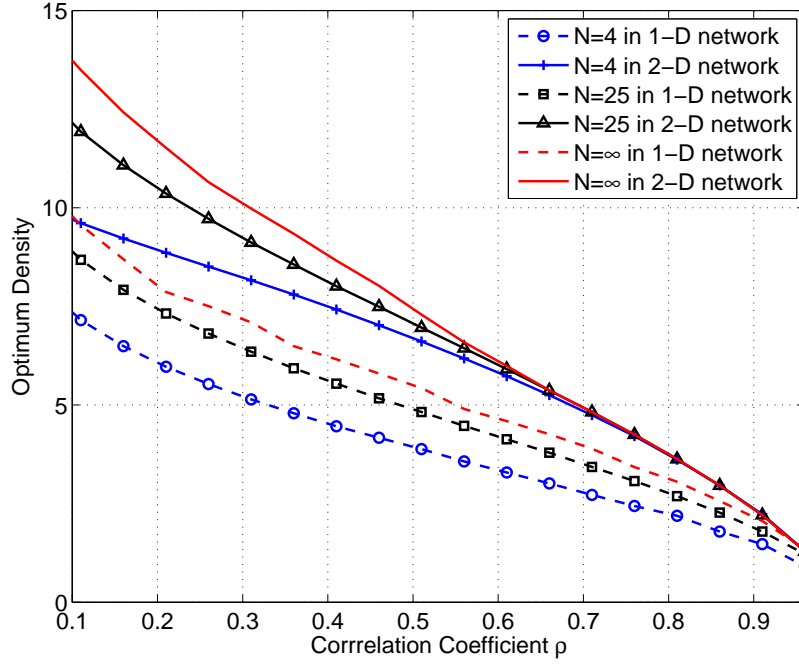
Figure 2.9: The optimum or asymptotically optimum node density v.s. spatial correlation coefficient ρ in the 1-D and 2-D networks ($\gamma_0 = 10$ dB).

Fig. 2.8 shows the numerical and simulated MSE of the spatial interpolation in the 2-D network, and the results for the 1-D case are also shown in the figure for comparison. The correlation coefficient is $\rho = 0.8$. The SNR is $\gamma_0 = 10$ dB. Simulation results with $N = 1600$ are used to approximate the asymptotic results with $N \to \infty$. For both the 1-D and 2-D cases, the MSE is convex in δ when N is small, and it becomes a monotonic decreasing function of δ when N is large. It can be seen that the 2-D network has a worse MSE compared to its 1-D counterpart. The performance difference between the 1-D and 2-D networks becomes smaller as either N or δ increases.

When N is small such that $\vartheta_{d,N}^2$ is convex in δ, the optimum node density in a 2-D network can be numerically identified by solving $|\frac{\partial \vartheta_d^2}{\partial \delta}| = 0$. When N is large or tends to infinity, the optimum or asymptotically optimum node density can be obtained by numerically solving solving $|\frac{\partial \vartheta_{d,N}^2}{\partial \delta}| = \varepsilon$, with ε being a very small number. Fig. 2.9 shows the optimum node density as a function of the

spatial correlation coefficient in the 1-D and 2-D networks. In the figure, $\varepsilon = 10^{-3}$ is used for both 1-D and 2-D networks when N is large. For both the 1-D and 2-D networks, the optimum node density decreases as ρ increases. For a fixed ρ, the optimum node density increases as N increases, and it is upper bounded by the asymptotic result. Therefore, the 1-D and 2-D networks have similar performance trends. For a given ρ, the optimum node density of the 2-D network is slightly higher than its 1-D counterpart.

From Figs. 2.7 – 2.9, it is observed that the performance of a 2-D network has similar trends as that of an 1-D network. Therefore, we can use the tools and results derived for an 1-D network to approximately predict the performance trend of a 2-D network, especially when the node density or the spatial correlation coefficient is large. Since the results of the 1-D network are expressed in closed-forms, using the 1-D results for the analysis of the 2-D network reduces the analysis complexity, and provides similar insights on the interactions between the parameters and performance.

2.6 Conclusions

In this chapter, the optimum sensor node densities for 1-D and 2-D WSNs with spatial source correlation were studied. The impacts of the node density on the MSE of the data reconstructed at the FC were investigated for both small networks with finite number of nodes, and large networks with the number of nodes tending to infinity. Exact analytical expressions of the MSE, many in closed-forms, were obtained for the 1-D and 2-D networks. The analytical results quantitatively identified the interactions among the various system parameters and the estimation fidelity, and the results provide insights and guidelines on the design of practical WSNs.

There were three observations from the analytical and numerical results. First, if the network

only needs to estimate spatially discrete data, placing exactly one sensor at the desired measurement locations will generate the optimum performance. Second, for the estimation of the data at arbitrary locations in the measurement field, the optimum node density can be found when the MSE-density slope is close to zero, and the optimum density decreases as the spatial correlation coefficient increases. Finally, the 1-D and 2-D networks have similar performance trends with respect to node density, and their performance difference diminishes as the spatial correlation coefficient increases.

2.7 Appendix of Proofs

2.7.1 Proof of Lemma 2.1

In Step 2, the MMSE vector $\mathbf{w}_{sl}$ that minimizes σ_η^2 can be obtained through the orthogonal principal [18], $\mathbb{E}\{[\mathbf{w}_{sl}^T\hat{\mathbf{x}}_s - x(\boldsymbol{\eta})]\hat{\mathbf{x}}_s^T\} = 0$, and the result is

$$
\begin{aligned}
\mathbf{w}_{sl}^T &= \mathbb{E}[x(\boldsymbol{\eta})\hat{\mathbf{x}}_s^T]\left[\mathbb{E}(\hat{\mathbf{x}}_s\hat{\mathbf{x}}_s^T)\right]^{-1}, \\
&= \sqrt{P_n}\mathbf{r}_{\boldsymbol{\eta}}\left(P_n\mathbf{R}_{xx} + \sigma_z^2\mathbf{I}_N\right)^{-1}\mathbf{W}_s^{-1},
\end{aligned}
\tag{2.33}
$$

where (2.2) and (2.6) are used for the second equality. Combining (2.6), (2.8), and (2.33) leads to (2.4).

2.7.2 Proof of Proposition 2.1

Based on Szego's Theorem [16], when $N \to \infty$, (2.11) can be rewritten as

$$
\sigma_s^2 = \lim_{N\to\infty}\sigma_{s,N}^2 = \int_{-\frac{1}{2}}^{\frac{1}{2}}\left[\frac{1}{\Lambda_{ss}(f)} + \frac{\gamma_0}{\delta}\right]^{-1}df,
\tag{2.34}
$$

36

where $\Lambda_{ss}(f) = \sum_{n=-\infty}^{+\infty} \rho^{|n|d} e^{-jn2\pi f}$ is the discrete-time Fourier transform (DTFT) of the sequence, $\left\{ \rho^{|n|d} \right\}_n$, which are elements of the Toeplitz matrix $\mathbf{R}_{ss}$. The DTFT, $\Lambda_{ss}(f)$, can be calculated by

$$\Lambda_{ss}(f) = \frac{1 - \rho^{2d}}{1 + \rho^{2d} - 2\rho^d \cos(2\pi f)}. \tag{2.35}$$

Substituting (2.35) into (2.34) leads to

$$\sigma_s^2 = \int_{-\frac{1}{2}}^{\frac{1}{2}} \left[\frac{1 + \rho^{2d} - 2\rho^d \cos(2\pi f)}{1 - \rho^{2d}} + \frac{\gamma_0}{\delta} \right]^{-1} df. \tag{2.36}$$

The above integral can be solved by using the identity [19, eqn. (2.553.3)]

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} [a + b\cos(2\pi f)]^{-1} df = \frac{1}{\sqrt{a^2 - b^2}}, \tag{2.37}$$

Combining (2.36) with (2.37) leads to (2.12).

2.7.3   Proof of Corollary 2.1

From (2.12), it is equivalent to show that $g_1(d) = (1 + \gamma_0 d)^2 + 4\gamma_0 d \frac{\rho^{2d}}{1 - \rho^{2d}}$ is a monotonic increasing

function of $d = \frac{1}{\delta}$. Taking the first derivative of $g_1(d)$, we have

$$g_1'(d) = \frac{2\gamma_0}{\left(1 - \rho^{2d}\right)^2} \times g_2(d, \gamma_0), \tag{2.38}$$

where $g_2(d, \gamma_0)$ is defined as

$$g_2(d, \gamma_0) \triangleq \left(1 - \rho^{2d}\right)^2 (1 + \gamma_0 d) + 2\rho^{2d}\left(1 - \rho^{2d}\right) + 4d\log(\rho)\rho^{2d} \tag{2.39}$$

From (2.38), in order to prove $g_1'(d) \geq 0$, it is sufficient to prove that $g_2(d, 0) \geq 0$ because

$g_2(d, \gamma_0) \geq g_2(d, 0)$. Let $v = \rho^{2d} \in [0, 1]$, then $g_2(d, 0)$ can be rewritten as

$$g_3(v) \triangleq g_2(d, 0) = 1 - v^2 + 2v\log(v), \quad 0 \leq v \leq 1 \tag{2.40}$$

It can be easily shown that $g_3''(v)=2(\frac{1}{v}-1)\geq 0, \forall v\in[0,1]$. Therefore $g_3(v)$ is quadratic on $[0,1]$ with the minimum value obtained at the solution of $g_3'(v)=-2v+2\log(v)+2=0$, which is $v=1$. Substituting $v=1$ into (2.40), we have $\min\{g_3(v)\}=0$. Therefore, $g_2(d,\gamma_0)\geq g_2(d,0)=g_3(v)\geq 0$, and this completes the proof.

2.7.4   Proof of Proposition 2.2

The Toeplitz matrix, $\mathbf{R}_{ds}$, is uniquely determined by the sequence $\mathbf{t}_{ds}=[t_{-(N-1)},\cdots,t_0,\cdots,t_{N-1}]^T$, where $t_n=\rho^{\frac{d}{2}}\rho^{|n+1|d}$ when $n<0$, and $t_n=\rho^{\frac{d}{2}}\rho^{nd}$ otherwise. When $N\to\infty$, the DTFT of the sequence $\mathbf{t}_{ds}$ can be calculated as

$$\Lambda_{ds}(f)=\rho^{\frac{d}{2}}\frac{(1-\rho^d)(1+e^{j2\pi f})}{1+\rho^{2d}-2\rho^d\cos(2\pi f)}. \tag{2.41}$$

Based on [16, Lemma 2], $\mathbf{R}_{ds}$ is asymptotically equivalent to a circulant matrix, $\mathbf{C}_{ds}=\mathbf{U}_N^H\mathbf{D}_{ds}\mathbf{U}_N$, where $\mathbf{U}_N^H$ is the unitary discrete Fourier transform (DFT) matrix with the (m,n)-th element being $(\mathbf{D}_{ds})_{m,n}=\frac{1}{\sqrt{N}}\exp\left[-j2\pi\frac{(m-1)(n-1)}{N}\right]$, and $\mathbf{D}_{ds}$ is a diagonal matrix with its k-th diagonal element being

$$(\mathbf{D}_{ds})_{k,k}=\Lambda_{ds}\left(\frac{k-1}{N}\right). \tag{2.42}$$

Similarly, the Toeplitz matrix, $\mathbf{R}_{ss}$, is asymptotically equivalent to a circulant matrix, $\mathbf{C}_{ss}=\mathbf{U}_N^H\mathbf{D}_{ss}\mathbf{U}_N$, where $\mathbf{D}_{ss}$ is a diagonal matrix with its k-th diagonal element being $(\mathbf{D}_{ss})=\Lambda_{ss}\left(\frac{k-1}{N}\right)$, with $\Lambda_{ss}(f)$ defined in (2.35).

Based on [20, Theorem 2.1], the error correlation matrix, $\mathbf{R}_{ee}^{(d)}$, is asymptotically equivalent to a circulant matrix, $\mathbf{C}_{ee}^{(d)}=\mathbf{C}_{ss}-\mathbf{C}_{ds}\left(\mathbf{C}_{ss}+\frac{\delta}{\gamma_0}\mathbf{I}\right)^{-1}\mathbf{C}_{ds}^H=\mathbf{U}_N^H\mathbf{D}_{ee}^{(d)}\mathbf{U}_N$, where $\mathbf{D}_{ee}^{(d)}=\mathbf{D}_{ss}-\mathbf{D}_{ds}\left(\mathbf{D}_{ss}+\frac{\delta}{\gamma_0}\mathbf{I}\right)^{-1}\mathbf{D}_{ds}^H$. It is apparent that $\mathbf{D}_{ee}^{(d)}$ is diagonal given that $\mathbf{D}_{ss}$ and $\mathbf{D}_{ds}$ are diagonal.

Based on Szego's Theorem, we have

$$\sigma_d^2 = \int_{-\frac{1}{2}}^{\frac{1}{2}} \left[\Lambda_{ss}(f) - \frac{|\Lambda_{ds}(f)|^2}{\Lambda_{ss}(f) + \frac{\delta}{\gamma_0}} \right] df. \tag{2.43}$$

Substituting (2.35) and (2.41) into the above equation and simplifying leads to (2.19).

2.7.5 Proof of Corollary 2.3

The MSE in (2.19) can be alternatively represented as

$$\sigma_d^2 = \left\{ 1 + \left[f_3^{-1}(\delta) - f_3(\delta) \right] / \left[f_1^{-1}(\delta) + f_3(\delta) \right] \right\}^{\frac{1}{2}}. \tag{2.44}$$

Since $f_1(\delta)$ is a decreasing function of δ and $f_3(\delta)$ is an increasing function of δ, it is straight-forward to show that $\left[f_3^{-1}(\delta) - f_3(\delta) \right] / \left[f_1^{-1}(\delta) + f_3(\delta) \right]$ is a decreasing function of δ, and this completes the proof.

2.7.6 Proof of Proposition 2.4

According to [17, Lemma 1], the TBT matrices, $\blacksquare_{ss}$ and $\blacksquare_{ds}$, are asymptotically equivalent to circulant-block-circulant (CBC) matrices, $\mathbf{B}_{ss}$ and $\mathbf{B}_{ds}$, respectively, where the eigenvalues of $\mathbf{B}_{ss}$ and $\mathbf{B}_{ds}$ are samples of $\Lambda'_{ss}(f_1, f_2)$ and $\Lambda'_{ds}(f_1, f_2)$, respectively [17, Theorem 3]. In addition, the CBC matrices, $\mathbf{B}_{ss}$ and $\mathbf{B}_{ds}$, share the same orthonormal eigenvectors [21]. Once the asymptotic equivalence is established, the rest of the proof follows the same procedure as described in Appendix D for the 1-D case.

2.8 Appendix of the Copyright

2.8.1 Documentation of multi-authored chapter

UNIVERSITY OF ARKANSAS

College of Engineering
Department of Electrical Engineering

July 31, 2013

To Whom It May Concern

This letter is to certify that Ms. Ning Sun, a Ph.D. candidate under my supervision at the Department of Electrical Engineering, has contributed more than 51% of the work for the following paper,

Jingxian Wu and Ning Sun, "Optimum sensor density in distortion tolerant wireless sensor networks," IEEE Transactions on Wireless Communications, vol. 11, pp. 2056-2064, June 2012.

The paper is included in her Ph.D. dissertation entitled "Distortion-Tolerant Communications with Correlated Information".

It should be noted that in my research group, the first author of a paper is usually the person who comes up with the original idea. For this paper, I came up with the original idea, and Ms. Ning Sun improved upon the idea and contributed more than 51% of the work to the paper.

Sincerely Yours,

Jingxian Wu
Director, Wireless Information Network Lab
Associate Professor of Electrical Engineering
University of Arkansas
Email: wuj@uark.edu
Tel: (479) 575-6584

3217 Bell Engineering Center • Fayetteville, AR 72701 • 479-575-3005 • Fax: 479-575-7967 • www.uark.edu
The University of Arkansas is an equal opportunity/affirmative action institution.

2.8.2   Copyright Clearance

Copyright Clearance Center **RightsLink®**

Home Create Account Help

IEEE
Requesting permission to reuse content from an IEEE publication

Title: Optimum Sensor Density in Distortion-Tolerant Wireless Sensor Networks
Author: Jingxian Wu; Ning Sun
Publication: Wireless Communications, IEEE Transactions on
Publisher: IEEE
Date: June 2012
Copyright © 2012, IEEE

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

2.9 References

[1] Q. Zhao and L. Tong, "Energy-efficient information retrieval for correlated source reconstruction in sensor networks," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 157-165, Jan. 2007.

[2] A. Anandkumar, L. Tong, and A. Swami, "Optimal node density for detection in energy-constrained random networks," *IEEE Trans. Signal Processing*, vol. 56, pp. 5232-5245, Oct. 2008.

[3] T. Wimalajeewa and S. K. Jayaweera, "Impact of mobile node density on detection performance measures in a hybrid sensor network," *IEEE Trans. Wireless Commun.*, vol. 9, pp. 1760-1769, May 2010.

[4] W. Wang, V. Srinivasan, B. Wang, and K.-C. Chua, "Coverage for target localization in wireless sensor networks," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 667-676, Feb. 2008.

[5] B. Wang, K.-C. Chua, V. Srinivasan, and W. Wang, "Information coverage in randomly deployed wireless sensor networks," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 2994-3004, Aug. 2007.

[6] L. Liu, X. Zhang, and H. Ma, "Localization-oriented coverage in wireless camera sensor networks," *IEEE Trans. Wireless Commun.*, vol. 10, pp. 484-494, Feb. 2011.

[7] J. F. Chamberland, and V. V. Veeravalli, "How dense should a sensor network be for detection with correlated observations?" *IEEE Trans. Inform. theory*, vol. 52, pp. 5099-5106, Nov. 2006.

[8] R. Cristescu, and M. Vetterli, "On the optimal density for real-time data gathering of spatio-temporal processes in sensor networks," in *Proc. IPSN 2005*, pp. 159-164, April, 2005.

[9] X. Zhang, H. Wang, F. N. Abdesselam, and A. A. Khokhar, "Distortion Analysis for Real-Time Data Collection of Spatially Temporally Correlated Data Fields in Wireless Sensor Networks," *IEEE Trans. Vehicular Technology*, vol. 58, pp. 1583-1594, March 2009.

[10] M. Gastpar, and M. Vetterli, "Power, Spatio-Temporal Bandwidth, and Distortion in Large Sensor Networks," *IEEE J. Selected Areas Commun.*, vol. 23, pp. 745-755, April 2005.

[11] Y. Sung, H. V. Poor, and H. Yu, "Optimal node density for two-dimensional sensor arrays," in *Proc. IEEE Sensor Array Multichannel Signal Processing Workshop*, pp. 271-274, 2008.

[12] J. Wu and N. Sun, "Optimal sensor density in a distortion-tolerant linear wireless sensor network," in *Proc. IEEE Global Telecommun. Conf. GLOBECOM'10*, Dec. 2010.

[13] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. on Inform. Theory*, vol. 46, pp. 388-404, Mar. 2000.

[14] S. D. Servetto, "On the feasibility of large scale wireless sensor networks," in *Proc. IEEE Int. Conf. commu., control, and computing*, Oct. 2002.

[15] D. Marco, E.J. Duarte-Melo, M. Liu, and D. L. Neuhoff, "On the many-to-one transport capacity of a dense wireless sensor network and the compressibility of its data," in *Proc. Intern. Conf. Info. Processing Sensor Networks ISPN'03*, pp. 1-16, Apr. 2003.

[16] H. Gazzah, P. A. Regalia, and J.P. Delmas, "Asymptotic eigenvalue distribution of block Toeplitz matrices and application to blind SIMO channel identification," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1243-1251, Mar. 2001.

[17] P. A. Voois, "A theorem on the asymtotic eigenvalue distribution of Toeplitz-block-Toeplitz matrix," *IEEE Trans. Signal Processing*, vol. 44, pp. 1837-1841, July 1996.

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

[19] I. S. Gradshteyn, and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 6nd ed. Academic Press, July 2000.

[20] R. M. Gray. Toeplitz and Circulant Matrices: a Review, NOW Publishers, 2006 (available online at http://ee.stanford.edu/ gray/).

[21] T. DE Mazancourt and D. Gerlic, "The inverse of a block-circulant matrix," *IEEE Trans. Antennas and Propagation*, vol. AP-31, pp. 808-810, Sept. 1983.

Chapter 3

Optimum Sampling in Spatial-Temporally Correlated Wireless Sensor Networks

Ning Sun and Jingxian Wu

3.1 Abstract

The optimum sampling in the one- and two-dimensional (1-D and 2-D) wireless sensor networks (WSNs) with spatial-temporally correlated data is studied in this chapter. The impacts of the node density in the space domain, the sampling rate in the time domain, and the space-time data correlation on the network performance are investigated asymptotically by considering a large network with infinite area but finite node density and finite temporal sampling rate, under the constraint of fixed power per unit area. The impact of space-time sampling on network performances is investigated in two cases. The first case studies the estimations of the space-time samples collected by the sensors, and the samples are discrete in both the space and time domains. The second case estimates an arbitrary data point on the space-time hyperplane by interpolating the discrete samples collected by the sensors. Optimum space-time sampling is obtained by minimizing the mean square error distortion at the network fusion center. The interactions among the various network parameters, such as spatial node density, temporal sampling rate, measurement noise, channel fading, and their impacts on the system performance are quantitatively identified with analytical and numerical studies.

3.2 Introduction

Data collected by a wireless sensor network (WSN) often contain redundancy due to the spatial and temporal correlation inherent in the monitored object(s). The spatial-temporal data correlations can be found in a wide range of practical applications, such as environment monitoring with temperature and humidity correlated in the space and time domains, soil and water quality monitoring with the chemical compositions correlated in the space and time domains, and structure health monitoring with spatial-temporally correlated vibration information of the civil structure [1], etc. The space-time redundancy/correlation is important to the performance and design of practical WSNs, which attempt to reconstruct a spatial-temporally correlated signal field by collecting the data samples from the sensors. Given a fixed transmission power per unit area, a higher spatial node density or temporal sampling rate means less transmission energy per sample, which usually degrades performance due to a lower signal-to-noise ratio (SNR) at the receiver. On the other hand, the system performance might benefit from more data samples per unit area per unit time by exploiting the space-time redundancy. Therefore, it is critical to identify the optimum space-time sampling, i.e., the optimum spatial node density and temporal sampling rate, in a WSN with spatial-temporally correlated data.

There have been considerable works in the literature studying the impacts of spatial node density on the network performance [2] – [6]. In [2], the optimum node density of a many-to-one linear network is analyzed by using the detection probability of a binary event as the performance metric. In [3], a Wiener process is used to model the spatial correlation of an one-dimensional (1-D) field. It is demonstrated that, due to the spatial data correlation, distortion-free communication can be achieved even if the per node throughput tends to 0 as $N \rightarrow \infty$. The optimum node densities in

45

both 1-D and two-dimensional (2-D) networks are obtained by minimizing the mean square error (MSE) between the recovered information and the original information under a distortion-tolerant communication framework [5],[6]. Most existing studies focus only on the spatial data correlation, and they do not consider the variation of the data in the time domain. In reality, the physical phenomenon under monitoring changes with respect to time, and the consecutive observations of a sensor node are often correlated temporally [7].

There are limited works on the study of WSNs with spatial-temporally correlated data [8], [12]. In [8], an arbitrary point on a continuous measurement field is estimated by performing space-time interpolation over the samples collected by the spatially discrete sensors, and there is a finite optimum node density to minimize the estimation MSE over the measurement field. The model in [8] is extended in [9] by considering realistic transmission schemes, such as limited transmission range and practical network/routing parameters. In [8],[9], the temporal data correlation is only utilized to perform time domain interpolation, and they do not consider the effects of optimum time domain sampling. The effects of both space and time domain sampling are studied in [10] by using the network energy as a performance metric, through the study of a collision free network protocols. All of the aforementioned studies consider an error-free communication channel between the transmitter and the receiver. The impacts of additive white Gaussian noise (AWGN) are considered in [11], which obtains a lower bound on the distortion as a function of the number of sensors and spatial-temporal communication bandwidth. However, the analysis is only applicable to a measurement field with finite degree-of-freedom and is discrete in the time domain. In addition, it does not consider the optimum sampling rate in the time domain. The optimum space-time sampling of continuous data in an 1-D network with AWGN channel is studied in [12].

In this chapter, we investigate the optimum space-time sampling for both 1-D and 2-D WSNs

46

with spatial-temporally correlated data. The 1-D network can be used to model practical WSNs designed for highways and tunnels. The 2-D network models WSNs that cover a large area, such as a farmland. There is no limitation on the statistical properties of the field, other than that it forms a *continuous* random process that is wide sense stationary (WSS) in both the space and time domain. Each sensor node collect samples of the field, and forward the information to a data fusion center (FC) through an one-hop AWGN or fading channel. Similar one-hop network structures are used in [2], [5], [6], [12] – [15]. The FC attempts to reconstruct the time-varying and spatially continuous data field from the discrete sensor samples by exploiting the data correlation in both the space and time domains with the minimum mean square error (MMSE) receiver. The impacts of the spatial node density, the temporal sampling rate, and the space-time data correlation on the reconstruction MSE are investigated asymptotically in a large network with infinite area, infinite time period, but finite node density and finite temporal sampling rate, under the constraint of fixed transmission power per unit area.

Compared to existing studies in the literature, this chapter has the following main contributions. First, to the best of our knowledge, this chapter is the first that explicitly quantifies the interactions between the performance of networks with spatial-temporally correlated data and various system parameters, such as spatial node density, temporal sampling rate, measurement noise, and channel distortions, for both 1-D and 2-D networks. Second, the optimum spatial-temporal sampling for two types of networks, one needs to recover only the discrete space-time samples collected by the sensors through their noisy observation, and one needs to recover an arbitrary data point on the space-time hyperplane, are identified through the asymptotic analysis. Third, the impacts of various practical factors, such as measurement noise, channel fading, and random network topology, on the performance of networks with spatial-temporally correlated data are studied through

47

numerical analysis and simulations.

The remainder of this chapter is organized as follows. Section 3.3 introduces the system model and a two-step MMSE estimation method. Sections 3.4 and 3.5 studies the impacts of spatial-temporal sampling on 1-D and 2-D networks, respectively, by following the two-step MMSE method. In these two sections, the optimum spatial-temporal samplings in various networks are identified with asymptotic analysis and simulations. Both analytical and numerical results are presented in Sections 3.4 and 3.5 to demonstrate the interactions among the various system parameters. Section 3.6 concludes the chapter.

3.3   Problem formulation

3.3.1   System model

Consider a WSN with N_s sensor nodes uniformly placed over a measurement field. Data collected by the sensors are spatially correlated, and they change with respect to time. We first study a network with a deterministic topology, where the sensors are placed over an equal-distance grid as shown in Fig. 3.1, with the distance between two adjacent nodes being d. Such a deterministic topology can be used to model networks that can be carefully planned beforehand and has no limitation on sensor locations. The performance of networks with deterministic topology will be compared to those with randomly distributed nodes. Networks with random topology can be used to model ad hoc networks or networks with mobile nodes. The results obtained for these two types of networks can serve as performance bounds for practical networks, which usually use a combination of these two topologies.

Each sensor node collects data samples with a sampling rate of $\theta = \frac{1}{T_s}$ Hz. In the space domain,

(a) The 1-D sensor network with 2-D space-time samples.

(b) The 2-D sensor network with 3-D space-time samples.
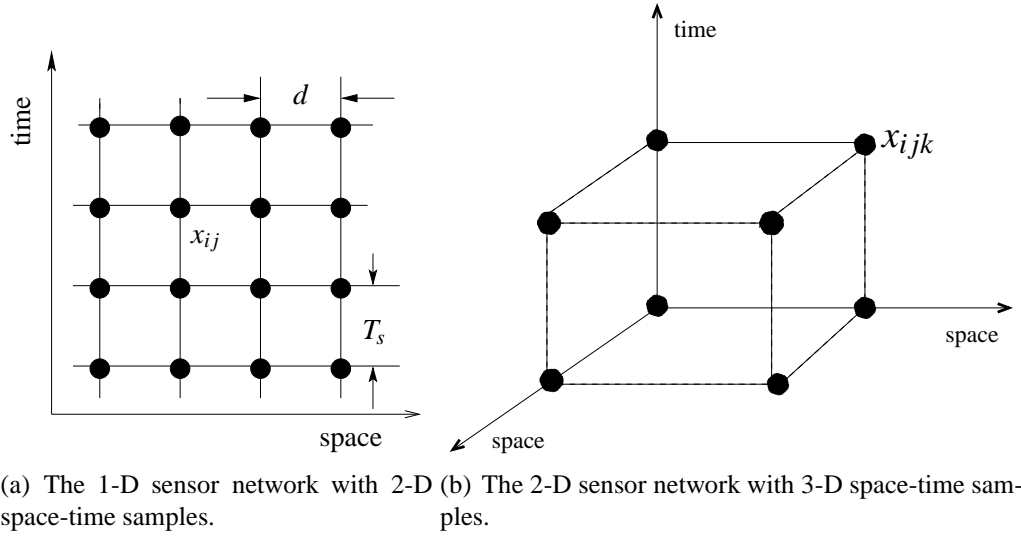
Figure 3.1: The spatial-temporally correlated sensor networks.

define the spatial node density, δ, as the number of nodes in a unit area. The spatial node densities are $\delta = \frac{1}{d}$ and $\delta = \frac{1}{d^2}$ for the grid-based 1-D and 2-D networks, respectively. Let $\boldsymbol{\eta} = [\mathbf{c}^T, t]^T$ represent the coordinate in the space-time hyperplane, where $\mathbf{A}^T$ denotes matrix transpose, $\mathbf{c}$ is the coordinate vector in the space domain, and t is the time variable.

Each sensor node will measure a spatial-temporally dependent physical quantity, $x(\boldsymbol{\eta_n})$, such as the temperature, humidity, or the vibration density of a bridge, etc. It is assumed that the physical quantities to be measured form a random process that is WSS in both the space and time domains. Due to the spatial-temporal redundancy of the measurement field, the spatial-temporal correlation function between any two arbitrary data samples is assumed as

$$\mathbb{E}\left[x(\boldsymbol{\eta_1})x(\boldsymbol{\eta_2})\right] = \rho_s^{\|\mathbf{c}_1 - \mathbf{c}_2\|} \cdot \rho_t^{|t_1 - t_2|} \tag{3.1}$$

where $\boldsymbol{\eta_n} = [\mathbf{c}_n^T, t_n]^T$, $\rho_s \in [0, 1]$ and $\rho_t \in [0, 1]$ are defined as the spatial correlation coefficient and the temporal correlation coefficient, respectively, and $\mathbb{E}(\cdot)$ represents mathematical expectation. In (3.1), the l_2 norm $\|\mathbf{c}_1 - \mathbf{c}_2\|$ measures the Euclidean distance between the two points with the

49

coordinates $\mathbf{c}_1$ and $\mathbf{c}_2$ in the space domain.

It is assumed that sensors deliver the measured data to the FC through an orthogonal media access control (MAC) scheme, such as the deterministic frequency division multiple access (FDMA), or the random exponentially-interval MAC (EI-MAC) [16], such that collision-free communication is achieved at the FC. The signal observed by the FC from the nth data sample is

$$y_n = \sqrt{\frac{E_n}{1 + \sigma_w^2}} \cdot h(\boldsymbol{\eta}_n) \cdot [x(\boldsymbol{\eta}_n) + w_n] + z_n, \tag{3.2}$$

where E_n is the average transmission energy per sample, $h(\boldsymbol{\eta}_n)$ represents the quasi-static fading coefficient, w_n is the measurement noise with variance σ_w^2, and z_n is the AWGN with variance σ_z^2. It is assumed that the total power per unit area is fixed at P_0. Given a network with a node density δ and a sample rate θ, the transmission energy per sample can be calculated as $E_n = \frac{P_0}{\theta \delta}$. It is assumed here that the sensor-FC distance is much larger than the sensor-sensor distance, such that all the sensors have approximately the same distance to the FC. Therefore, signals from all the sensors experience similar pathloss, such that they can employ the same transmission energy.

3.3.2 Optimum MMSE detection

The FC will obtain an estimate of the spatial-temporally continuous quantity, $x(\boldsymbol{\eta})$, $\forall \boldsymbol{\eta} \in \Omega_{\boldsymbol{\eta}}$, by using $N = N_s N_t$ discrete space-time samples received at the FC, where N_s is the number of the sensor nodes and N_t is the number of time-domain samples collected by each node. Define the space-time data sample vector as $\mathbf{x}_{st} = [\mathbf{x}_1^T, \ldots, \mathbf{x}_{N_s}^T]^T \in \mathscr{R}^{N \times 1}$, where $\mathbf{x}_i = [x_{i1}, \ldots, x_{iN_t}]^T \in \mathscr{R}^{N_t} \times 1$ is the time domain sample vector collected by the ith sensor node, and $\mathscr{R}$ is the set of real numbers. The corresponding signal observed by the FC can then be represented as $\mathbf{y} = [\mathbf{y}_1^T, \ldots, \mathbf{y}_{N_s}^T]^T \in \mathscr{R}^{N \times 1}$, with $\mathbf{y}_i = [y_{i1}, \ldots, y_{iN_t}]^T \in \mathscr{R}^{N_t \times 1}$.

The MSE for $x(\boldsymbol{\eta})$ is

$$\sigma_{\boldsymbol{\eta}}^2 = \mathbb{E}\left[\hat{x}(\boldsymbol{\eta}) - x(\boldsymbol{\eta})\right]^2, \boldsymbol{\eta} \in \Omega_{\boldsymbol{\eta}} \tag{3.3}$$

where $\hat{x}(\boldsymbol{\eta})$ is the estimate of $x(\boldsymbol{\eta})$ based on $\mathbf{y}$ at the FC.

The optimum linear receiver that minimizes $\sigma_{\boldsymbol{\eta}}^2$ is the MMSE receiver described as follows [17]

$$\hat{x}(\boldsymbol{\eta}) = \sqrt{\frac{E_n}{1 + \sigma_w^2}} \mathbf{r}_{\boldsymbol{\eta}}^H \mathbf{H}^H \left[\frac{E_n}{1 + \sigma_w^2} \mathbf{H} \mathbf{R}_{xx} \mathbf{H}^H + \frac{E_n \sigma_w^2}{1 + \sigma_w^2} \mathbf{H} \mathbf{H}^H + \sigma_z^2 \mathbf{I}_N\right]^{-1} \mathbf{y}, \tag{3.4}$$

where $\mathbf{r}_{\boldsymbol{\eta}} = \mathbb{E}\left[x(\boldsymbol{\eta})\mathbf{x}_{st}\right] \in \mathscr{R}^{N \times 1}$, $\mathbf{R}_{xx} = \mathbb{E}\left[\mathbf{x}_{st}\mathbf{x}_{st}^H\right] \in \mathscr{R}^{N \times N}$ with the element defined in (3.1), and $\mathbf{A}^H$ denotes the matrix Hermitian operation. The channel coefficient matrix, $\mathbf{H} \in \mathscr{C}^{N \times N}$, is a diagonal matrix with the diagonal elements being $\mathbf{h} = [\mathbf{h}_1^T, \ldots, \mathbf{h}_{N_s}^T]^T \in \mathscr{C}^{N \times 1}$, where $\mathbf{h}_i = h_i \mathbf{I}_{N_t} \in \mathscr{C}^{N_t \times 1}$ with h_i corresponding to the fading coefficient between the ith node and the FC, $\mathbf{I}_{N_t}$ is a size-N_t identity matrix, and $\mathscr{C}$ is the set of complex numbers.

With the optimum MMSE receiver given in (3.4), the MSE $\sigma_{\boldsymbol{\eta}}^2$ can be calculated as

$$\sigma_{\boldsymbol{\eta}}^2 = \mathbb{E}_{\mathbf{H}} \left\{ 1 - \mathbf{r}_{\boldsymbol{\eta}}^H \left[\mathbf{R}_{xx} + \sigma_w^2 + (1 + \sigma_w^2)\frac{\theta \delta}{\gamma_0}(\mathbf{H}^H \mathbf{H})^{-1}\right]^{-1} \mathbf{r}_{\boldsymbol{\eta}} \right\}, \tag{3.5}$$

where $\gamma_0 = \frac{P_0}{\sigma_z^2}$ is the signal-to-noise ratio (SNR) per unit area with AWGN, and the expectation operation is performed with respect to $\mathbf{H}$. The MSE $\sigma_{\boldsymbol{\eta}}^2$ given in (3.5) is a function of the space-time coordinate $\boldsymbol{\eta}$, the SNR γ_0, the measurement noise variance σ_w^2, the spatial correlation coefficient ρ_s, the temporal correlation coefficient ρ_t, the spatial node density δ, the temporal sampling rate θ, and the fading coefficient $\mathbf{H}$.

Given a fixed transmission power per unit area, the spatial-temporal sampling rate, δ and θ, play a critical role on the MSE $\sigma_{\boldsymbol{\eta}}^2$. A smaller node density and/or temporal sampling rate means more transmission energy per sample, thus a better SNR per sample, which can benefit the system

performance. On the other hand, a smaller node density and/or sampling rate means less samples per unit area per unit time, thus a smaller correlation among the data collected by the nodes, and this might degrade the estimation performance.

In order to distinguish the opposite impacts of the spatial-temporal sampling rates, we use an equivalent two-step MMSE method [6].

Lemma 3.1: The optimum MMSE given in (3.4) is equivalent to the two-step MMSE described as follows.

1) The FC first obtains an estimate of the N discrete space-time samples, $\mathbf{x}_{st}$, with a linear MMSE receiver as

$$\hat{\mathbf{x}}_{st} = \mathbf{W}_x^H \mathbf{y}, \tag{3.6}$$

where $\hat{\mathbf{x}}_{st} \in \mathscr{R}^{N \times 1}$ is the MMSE estimate of $\mathbf{x}_{st}$. The MMSE matrix $\mathbf{W}_x \in \mathscr{R}^{N \times N}$ is designed to minimize the average MSE per sample:

$$\sigma_{st,N}^2 = \frac{1}{N} \mathbb{E}\left[\|\hat{\mathbf{x}}_{st} - \mathbf{x}_{st}\|^2\right]. \tag{3.7}$$

2) The FC obtains an estimate of the data at an arbitrary location, $\hat{x}(\boldsymbol{\eta})$, $\forall \boldsymbol{\eta} \in \Omega_{\boldsymbol{\eta}}$, by interpolating $\hat{\mathbf{x}}_{st}$ with the MMSE criterion,

$$\hat{x}(\boldsymbol{\eta}) = \mathbf{w}_{sl}^H \hat{\mathbf{x}}_{st}, \tag{3.8}$$

where the vector, $\mathbf{w}_{sl} \in \mathscr{R}^{N \times 1}$, is designed to minimize the MSE $\sigma_{\boldsymbol{\eta}}^2 = \mathbb{E}\left[\hat{x}(\boldsymbol{\eta}) - x(\boldsymbol{\eta})\right]^2$.

Decomposing the optimum MMSE of (3.4) into the two-step MMSE allows us to study the two opposite effects of spatial-temporal sampling on the MSE separately. In the following two sections, we will investigate, respectively, the impacts of the node density on 1-D and 2-D networks by following the two-step MMSE.

3.4 Optimum space-time sampling in one-dimensional networks

In this section, we study the optimum space-time sampling in an 1-D network, where the N_s sensor nodes are evenly distributed over a length-L linear section as shown in Fig. 3.1(a). In this WSN, the space-time coordinate of the jth data sample collected by the ith sensor can then be represented as $[(i-1)d, (j-1)T_s]$. The spatial-temporal correlation matrix, $\mathbf{R}_{xx} = \mathbb{E}\left[\mathbf{x}_{st}\mathbf{x}_{st}^H\right] \in \mathscr{R}^{N \times N}$, can be expressed as

$$\mathbf{R}_{xx} = \mathbf{R}_s \otimes \mathbf{R}_t \tag{3.9}$$

where $\otimes$ denotes the Kronecker product, and $\mathbf{R}_s \in \mathscr{R}^{N_s \times N_s}$ and $\mathbf{R}_t \in \mathscr{R}^{N_t \times N_t}$ are the correlation matrices in the space domain and time domain, respectively. The space domain correlation matrix, $\mathbf{R}_s$, has the form of a symmetric Toeplitz matrix with the first row and first column being $\mathbf{r}_s = \left[1, \rho_s^d, \ldots, \rho_s^{(N_s-1)d}\right]^T$. Similarly, the time domain correlation matrix, $\mathbf{R}_t$, is a symmetric Toeplitz matrix with the first row and first column being $\mathbf{r}_t = \left[1, \rho_t^{T_s}, \ldots, \rho_t^{(N_t-1)T_s}\right]^T$. The matrix, $\mathbf{R}_{xx}$, has the form of a Toeplitz-block-Toeplitz (TBT) matrix [18], i.e., $\mathbf{R}_{xx}$ is a block Toeplitz matrix, and each sub-matrix is also a Toeplitz matrix.

3.4.1 MMSE estimation of the discrete samples

For the MMSE estimation described in (3.6), the optimum $\mathbf{W}_x$ that minimizes the MSE, $\sigma_{st,N}^2$, can be found through the orthogonal principal, $\mathbb{E}\left[(\hat{\mathbf{x}}_{st} - \mathbf{x}_{st})\mathbf{y}^H\right] = \mathbf{0}$. The result is

$$\mathbf{W}_x^H = \sqrt{\frac{E_n}{1+\sigma_w^2}}\mathbf{R}_{xx}\mathbf{H}^H\left[\frac{E_n}{1+\sigma_w^2}\mathbf{H}\mathbf{R}_{xx}\mathbf{H}^H + \frac{E_n\sigma_w^2}{1+\sigma_w^2}\mathbf{H}\mathbf{H}^H + \sigma_z^2\mathbf{I}_N\right]^{-1}, \tag{3.10}$$

The conditional error correlation matrix, $\mathbf{R}_{ee|\mathbf{H}}^{(x)} = \mathbb{E}\left[\mathbf{e}_s\mathbf{e}_s^H|\mathbf{H}\right]$, with $\mathbf{e}_s = \hat{\mathbf{x}}_{st} - \mathbf{x}_{st}$, can then be

calculated as

$$
\begin{aligned}
\mathbf{R}_{ee|\mathbf{H}}^{(x)} &= \mathbf{R}_{xx} - \mathbf{R}_{xx}\mathbf{H}^H \left[\mathbf{H}\mathbf{R}_{xx}\mathbf{H}^H + \sigma_w^2 \mathbf{H}\mathbf{H}^H + (1+\sigma_w^2)\frac{\theta\delta}{\gamma_0}\mathbf{I}_N \right]^{-1} \mathbf{H}\mathbf{R}_{xx} \\
&= \left[\mathbf{R}_{xx}^{-1} + \frac{\mathbf{H}\mathbf{H}^H}{\sigma_w^2\mathbf{H}\mathbf{H}^H + (1+\sigma_w^2)\frac{\theta\delta}{\gamma_0}\mathbf{I}_N} \right]^{-1},
\end{aligned}
\tag{3.11}
$$

where the orthogonal principal is used in the first equality, and the second equality is based on the identity $\mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} = (\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}$. The MSE can then be calculated as

$$
\sigma_{st,N}^2 = \frac{1}{N}\mathbb{E}_{\mathbf{H}}\left[\mathrm{trace}\left(\mathbf{R}_{ee|\mathbf{H}}^{(x)} \right) \right]
\tag{3.12}
$$

where $\mathrm{trace}\,(\mathbf{A})$ returns the trace of the matrix $\mathbf{A}$.

In Equations (3.11) and (3.12), the calculation of the MSE involves matrix inversion, the trace operation, and the expectation operation. The value of the MSE can be evaluated numerically. In order to explicitly identify the impacts of the node density and sampling rate on the MSE, we will first focus on the analysis of system operating in the AWGN channel, and this will allow us to express the MSE as a closed form expression of the node density and sampling rate. The MSE obtained under the AWGN channel will be compared to the MSE under the fading environment later in this section.

Proposition 3.1: When $N_s \to \infty$ and $N_t \to \infty$ while keeping both δ and θ finite, the MSE of the estimation of the discrete samples collected by the sensors and transmitted in an AWGN channel is

$$
\sigma_{st}^2 = \lim_{N\to\infty} \sigma_{st,N}^2 = \frac{\sqrt{2}}{\pi\sqrt{\beta}} \cdot K\left(\sqrt{\frac{\alpha}{\beta}} \right)
\tag{3.13}
$$

where $K(\cdot)$ is the complete elliptic integral of the first kind [19, Equation (8.112.1)], and

$$\alpha = \frac{8}{\sigma_w^2 + (1+\sigma_w^2)\frac{\theta\delta}{\gamma_0}} \cdot \frac{\rho_s^{\frac{1}{\delta}}}{1-\rho_s^{\frac{2}{\delta}}} \cdot \frac{\rho_t^{\frac{1}{\theta}}}{1-\rho_t^{\frac{2}{\theta}}}, \tag{3.14a}$$

$$\beta = \frac{1}{2} + \frac{1}{\sigma_w^2 + (1+\sigma_w^2)\frac{\theta\delta}{\gamma_0}} \cdot \left(1 + \frac{2\rho_s^{\frac{2}{\delta}}}{1-\rho_s^{\frac{2}{\delta}}}\right)\left(1 + \frac{2\rho_t^{\frac{2}{\theta}}}{1-\rho_t^{\frac{2}{\theta}}}\right) + \frac{1}{2}\left[\frac{1}{\sigma_w^2 + (1+\sigma_w^2)\frac{\theta\delta}{\gamma_0}}\right]^2 + \frac{\alpha}{2}. \tag{3.14b}$$

Proof: The proof is given in Appendix 3.7.1. ∎

In Proposition 3.1, the spatial-temporal sampling affects the MSE in the form of the following functions, $f_1(\rho_s,\delta) = \frac{\rho_s^{\frac{2}{\delta}}}{1-\rho_s^{\frac{2}{\delta}}}$, $g_1(\rho_s,\delta) = \frac{\rho_s^{\frac{1}{\delta}}}{1-\rho_s^{\frac{2}{\delta}}}$, $f_1(\rho_t,\theta) = \frac{\rho_t^{\frac{2}{\theta}}}{1-\rho_t^{\frac{2}{\theta}}}$, $g_1(\rho_t,\theta) = \frac{\rho_t^{\frac{1}{\theta}}}{1-\rho_t^{\frac{2}{\theta}}}$, and $f_2(\delta,\theta) = \frac{1}{\sigma_w^2+(1+\sigma_w^2)\frac{\theta\delta}{\gamma_0}}$. Among them, $f_1(\rho_s,\delta)$ and $g_1(\rho_s,\delta)$ are related to the spatial correlation, and they are increasing functions of δ. $f_1(\rho_t,\theta)$ and $g_1(\rho_t,\theta)$ are related to the temporal correlation, and they are increasing functions of θ. The function $f_2(\delta,\theta)$ is a decreasing function of both δ and θ.

In Proposition 3.1, if we assume that the data is spatially correlated but temporally uncorrelated, then the MSE of the spatial samples can be simplified as follows.

Corollary 3.1: If $\rho_t = 0$, the asymptotic MSE of the estimation for the spatially correlated samples is

$$\sigma_s^2 = \left[\left(1 + \frac{1}{\sigma_w^2 + (1+\sigma_w^2)\frac{\theta\delta}{\gamma_0}}\right)^2 + \frac{4\rho_s^{\frac{2}{\delta}}}{\left(\sigma_w^2+(1+\sigma_w^2)\frac{\theta\delta}{\gamma_0}\right)\left(1-\rho_s^{\frac{2}{\delta}}\right)}\right]^{-\frac{1}{2}}. \tag{3.15}$$

Proof: Setting $\rho_t = 0$ leads to $\alpha = 0$ and $\beta = 0.5 + [1+2f_1(\rho_s,\delta)]f_2(\delta,\theta) + 0.5f_2(\delta,\theta)^2$. Equation (3.15) can be obtained by substituting β into (3.13). ∎

When $\sigma_w^2 = 0$, the result in Corollary 3.1 coincides with [5, Equation (12)], where only the spatial samples are considered. It was shown in [5] analytically that σ_s^2 is an increasing function in δ.

Similarly, based on the symmetry between the space and time domains, we can get the MSE of the estimation of the temporal samples for a given node, by exchanging ρ_s with ρ_t, and δ with θ in (3.15).

Fig. 3.2 shows the asymptotic MSE as a function of the spatial node density, δ, under various values of the correlation coefficients, ρ_t and ρ_s, in an AWGN channel with SNR $\gamma_0 = 10\,\text{dB}$. Define $\gamma_w = \frac{P_0}{\sigma_w^2}$ as the measurement SNR per unit area. The temporal sampling rate is $\theta = 10$ sample/sec. Data samples are assumed to be a zero-mean Gaussian process with the auto-correlation function given in (3.1). The simulation results are obtained by using $N_s = N_t = 60$ samples to approximate infinite number of samples. Excellent match is observed between the simulation results with finite number of samples and the asymptotic results with infinite number of samples. As expected, the MSE performance improves as γ_w increases. When $\gamma_w = 10\,\text{dB}$, there is only a slight difference between the system with and without measurement noise. In addition, the MSE is an increasing function in node density for all configurations. This indicates that the MSE for the discrete data samples can benefit from a smaller spatial node density. Therefore, if we only want to obtain the data at some discrete locations, we should use a node density that is as small as allowed by the application, i.e., placing exactly one sensor at each desired measurement location will obtain the optimum performance. Due to the symmetry between the space and time domain, the above analysis is also true for the relationship between σ_{st}^2 and θ. In addition, the MSE approaches a constant as $\delta \to \infty$. The result is corroborated by the following corollary.

Corollary 3.2: For the estimation of the discrete samples collected by the sensors and transmitted in AWGN channels, given a sampling rate θ, when $\delta \to \infty$, the asymptotic MSE approaches
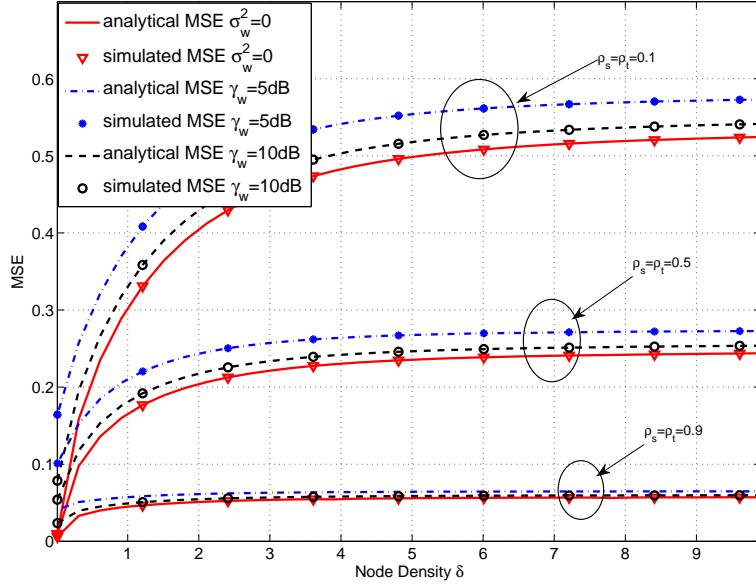
Figure 3.2: The asymptotic MSE of the estimated discrete data samples in the 1-D network under various values of measurement SNR γ_w (AWGN, $\gamma_0 = 10$ dB, $\theta = 10$ Hz).

a constant as

$$\lim_{\delta \to \infty} \sigma_{st}^2 = \frac{2}{\pi} \left[1 - \frac{2\gamma_0}{(1+\sigma_w^2)\theta \log(\rho_s)} \cdot \frac{1+\rho_t^{\frac{1}{\theta}}}{1-\rho_t^{\frac{1}{\theta}}} \right]^{-\frac{1}{2}} \cdot K(\Delta_\delta), \tag{3.16}$$

with $\Delta_\delta = \left[\dfrac{8\gamma_0 \rho_t^{\frac{1}{\theta}}}{2\gamma_0(1+\rho_t^{\frac{1}{\theta}})^2 - (1+\sigma_w^2)\theta \log(\rho_s)(1-\rho_t^{\frac{2}{\theta}})} \right]^{\frac{1}{2}}$.

Proof: The proof is in Appendix 3.7.2. ∎

Corollary 3.3: For the estimation of the discrete samples collected by the sensors and transmitted in AWGN channels, when both $\theta \to \infty$ and $\delta \to \infty$, we have

$$\lim_{\delta \to \infty, \theta \to \infty} \sigma_{st}^2 = \frac{2}{\pi} \left(1 + \frac{4}{\varpi} \right)^{-\frac{1}{2}} \cdot K \left(\sqrt{\frac{4}{4+\varpi}} \right), \tag{3.17}$$

where $\varpi = \frac{\log(\rho_s)\log(\rho_t)(1+\sigma_w^2)}{\gamma_0}$.

Proof: Equation (3.17) can be directly proved by substituting $\lim_{\theta \to \infty} \theta \left(1 - \rho_t^{\frac{1}{\theta}} \right) = -\log(\rho_t)$ into (3.45). ∎

57

In (3.17), when both θ and δ tend infinity, the limit depends on the correlation coefficients and the SNR. The relationship between the limit and ρ_s, ρ_t, γ_0 is given by the following corollary.

Corollary 3.4: The limit in Corollary 3.3 is proportional to ρ_s and ρ_t, and inversely proportional to the SNR γ_0.

Proof: The proof is in Appendix 3.7.3. ∎

We next compare in Fig. 3.3 the MSE for systems operating in AWGN channels and fading channels, respectively. The MSE in fading channels is obtained with a hybrid analytical and simulation method, i.e., given $\mathbf{H}$, the conditional MSE can be calculated by performing the trace operation over (3.11), and the unconditional MSE can then be obtained by averaging over a large number of independent implementations of $\mathbf{H}$. The parameters, γ_0 and θ, are the same as those in Fig. 3.2, and the variance of the measurement noise is $\sigma_w^2 = 0$. The fading MSE is lower bounded by its AWGN counterpart. The difference between the MSE of these two types of networks gradually diminishes as ρ_s and ρ_t increases. When $\rho_s = \rho_t = 0.9$, there is only a slight difference between the two, especially when the node density is high. In addition, both of the two networks have the same performance trend, i.e., the MSE is an increasing function in δ. Therefore, the analytical result in AWGN channel can provide a rough guideline on the design of systems with fading.

3.4.2 MMSE spatial-temporal interpolation

This section discusses the distortion performance of space-time interpolation, i.e., the estimation of any arbitrary point on the space-time plane by interpolating the N discrete space-time samples.

Since we are interested in the reconstruction fidelity of the entire space-time hyperplane, the worst case scenario is considered by estimating the data located in the middle of the square formed
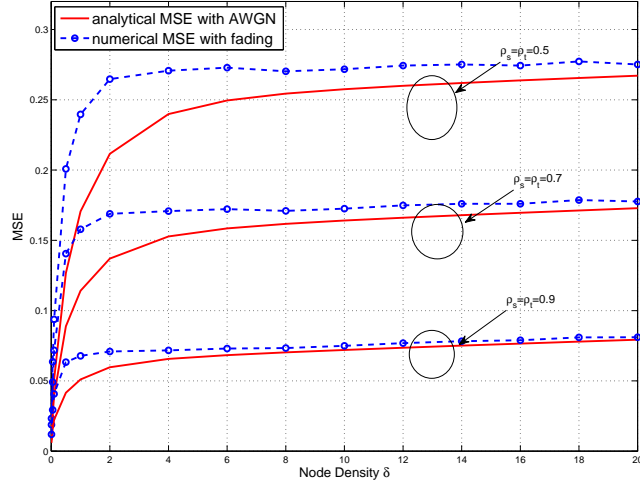
Figure 3.3: Impacts of fading on the asymptotic MSE of the estimated discrete data samples in 1-D networks ($\gamma_0 = 10$ dB, $\theta = 10$ Hz, $\sigma_w^2 = 0$).

by four neighboring samples, as shown in Fig. 3.4(a), with the data points to be estimated being $x'_{ij} = x[(i - \frac{1}{2})d, (j - \frac{1}{2})T_s]$, for $i = 1, \ldots, N_s$ and $j = 1, \ldots, N_t$. Define the interpolation data vector as $\mathbf{x}_{dt} = \left[\mathbf{x}_1^{'T}, \ldots, \mathbf{x}_{N_s}^{'T} \right]^T \in \mathscr{R}^{N \times 1}$, where $\mathbf{x}'_i = \left[x'_{i1}, x'_{i2}, \ldots, x'_{iN_t} \right]^T \in \mathscr{R}^{N_t \times 1}$.

(a) Space-time interpolation

(b) Space interpolation

(c) Time interpolation
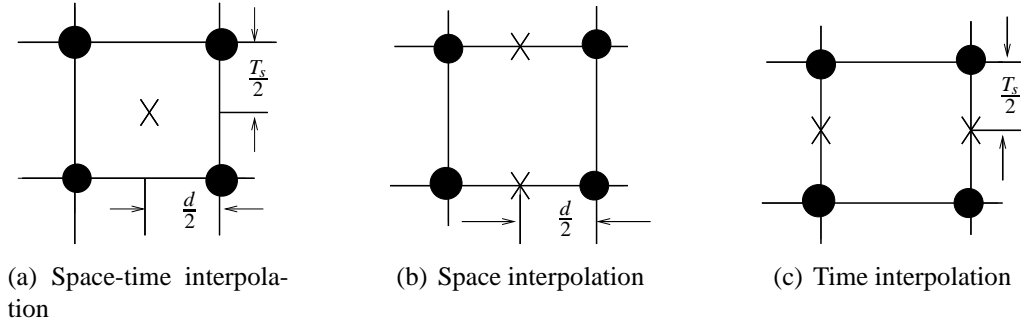
Figure 3.4: Three types of interpolations for the 1-D network.

Based on the orthogonal principal, $\mathbb{E}\left[(\hat{\mathbf{x}}_{dt} - \mathbf{x}_{dt})\hat{\mathbf{x}}_{st}^H \right] = \mathbf{0}$, where $\hat{\mathbf{x}}_{dt}$ is an estimate of $\mathbf{x}_{dt}$, the MMSE space-time interpolations can be expressed by

$$\hat{\mathbf{x}}_{dt} = \mathbf{R}_{d\hat{x}}\mathbf{R}_{\hat{x}\hat{x}}^{-1}\hat{\mathbf{x}}_{st}, \tag{3.18}$$

where

$$\mathbf{R}_{d\hat{x}} \triangleq \mathbb{E}(\mathbf{x}_{dt}\hat{\mathbf{x}}_{st}^H) = \sqrt{\frac{E_n}{1+\sigma_w^2}}\mathbf{R}_{dx}\mathbf{H}^H\mathbf{W}_x, \tag{3.19a}$$

$$\mathbf{R}_{\hat{x}\hat{x}} \triangleq \mathbb{E}(\hat{\mathbf{x}}_{st}\hat{\mathbf{x}}_{st}^H) = \mathbf{W}_x^H \left(\frac{E_n}{1+\sigma_w^2}\mathbf{H}\mathbf{R}_{xx}\mathbf{H}^H + \frac{E_n\sigma_w^2}{1+\sigma_w^2}\mathbf{H}\mathbf{H}^H + \sigma_z^2\mathbf{I}_N \right)\mathbf{W}_x, \tag{3.19b}$$

with $\mathbf{R}_{dx} \triangleq \mathbb{E}(\mathbf{x}_{dt}\mathbf{x}_{st}^H) = \mathbf{R}_s' \otimes \mathbf{R}_t'$ being a TBT matrix. The matrix $\mathbf{R}_s'$ is a Toeplitz matrix with the first row being $\rho_s^{\frac{d}{2}}[1, 1, \rho_s^d, \ldots, \rho_s^{(N_s-2)d|}]^T \in \mathscr{R}^{N_s \times 1}$, and the first column $\rho_s^{\frac{d}{2}}[1, \rho_s^d, \ldots, \rho_s^{(N_s-1)d|}]^T \in \mathscr{R}^{N_s \times 1}$. Similarly $\mathbf{R}_t'$ is a Toeplitz matrix with the first row being $\rho_t^{\frac{T_s}{2}}[1, 1, \rho_t^{T_s}, \ldots, \rho_t^{(N_t-2)T_s}]^T \in \mathscr{R}^{N_t \times 1}$, and the first column $\rho_t^{\frac{T_s}{2}}[1, \rho_t^{T_s}, \ldots, \rho_t^{(N_t-1)T_s}]^T \in \mathscr{R}^{N_t \times 1}$. Combining (3.18) with (3.19), we have

$$\hat{\mathbf{x}}_{dt} = \sqrt{\frac{E_n}{1+\sigma_w^2}}\mathbf{R}_{dx}\mathbf{H}^H \left[\frac{E_n}{1+\sigma_w^2}\mathbf{H}\mathbf{R}_{xx}\mathbf{H}^H + \frac{E_n\sigma_w^2}{1+\sigma_w^2}\mathbf{H}\mathbf{H}^H + \sigma_z^2\mathbf{I}_N \right]^{-1}\mathbf{y}. \tag{3.20}$$

The corresponding error correlation matrix, $\mathbf{R}_{ee}^{(d)} \triangleq \mathbb{E}\left[(\hat{\mathbf{x}}_{dt} - \mathbf{x}_{dt})(\hat{\mathbf{x}}_{dt} - \mathbf{x}_{dt})^H\right]$, can then be calculated by

$$\mathbf{R}_{ee}^{(d)} = \mathbb{E}_{\mathbf{H}} \left\{ \mathbf{R}_{xx} - \mathbf{R}_{dx}\mathbf{H}^H \left[\mathbf{H}\mathbf{R}_{xx}\mathbf{H}^H + \sigma_w^2\mathbf{H}\mathbf{H}^H + (1+\sigma_w^2)\frac{\theta\delta}{\gamma_0}\mathbf{I}_N \right]^{-1}\mathbf{H}\mathbf{R}_{xd} \right\}, \tag{3.21}$$

where $\mathbf{R}_{dd} = \mathbb{E}(\mathbf{x}_{dt}\mathbf{x}_{dt}^H) = \mathbf{R}_{xx}$ is used in the above equation, and $\mathbf{R}_{xd} = \mathbf{R}_{dx}^H$.

The MSE for the space-time interpolation when operating in a fading channel can be evaluated numerically by performing the trace operation over (3.21). To gain more insights on the impacts of node density and sampling rate, we next perform asymptotic analysis for systems operating in AWGN channels.

Proposition 3.2: When $N_s \to \infty$ and $N_t \to \infty$ while keeping both δ and θ finite, the MSE of the spatial-temporal interpolation for a network operating in AWGN channels is

$$\vartheta_{st}^2 \triangleq \lim_{N \to \infty} \vartheta_{st,N}^2 = \frac{1-\rho_t^{\frac{1}{\theta}}}{1+\rho_t^{\frac{1}{\theta}}} \cdot \left\{ 1 + \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{1+\cos(2\pi f)}{v - \cos(2\pi f)} \cdot \left[\frac{q - \cos(2\pi f)}{p - \cos(2\pi f)} \right]^{\frac{1}{2}} df \right\} \tag{3.22}$$

60

where

$$v = \frac{1 + \rho_t^{\frac{2}{\theta}}}{2\rho_t^{\frac{1}{\theta}}}, \ p = v + \frac{1}{2\left(\sigma_w^2 + (1 + \sigma_w^2)\frac{\theta\delta}{\gamma_0}\right)} \cdot \frac{1 - \rho_t^{\frac{2}{\theta}}}{\rho_t^{\frac{1}{\theta}}} \cdot \frac{1 + \rho_s^{\frac{1}{\delta}}}{1 - \rho_s^{\frac{1}{\delta}}},$$

$$q = v + \frac{1}{2\left(\sigma_w^2 + (1 + \sigma_w^2)\frac{\theta\delta}{\gamma_0}\right)} \cdot \frac{1 - \rho_t^{\frac{2}{\theta}}}{\rho_t^{\frac{1}{\theta}}} \cdot \frac{1 - \rho_s^{\frac{1}{\delta}}}{1 + \rho_s^{\frac{1}{\delta}}}. \tag{3.23}$$

Proof: The proof is in Appendix 3.7.4. ∎

The results in Proposition 3.2 illustrate the asymptotic MSE performance for the MMSE inter-polation in both the space and time domains. Even though the MSE in Proposition 3.2 is expressed as an explicit function of the correlation coefficients and the space-time sample rates, it is expressed in the form of an integral and eludes a closed-form expression. It should be noted that the integrand is composed for elementary functions, and the integration limit is finite. Therefore the integral can be easily evaluated numerically. To gain further insight on the impact of the space-time correlation on the estimation performance, we consider in the following section the interpolation in just one domain.

3.4.3   Interpolation in the space or time domain

In this section, we consider the MSE performance of interpolation in the space domain as in Fig. 3.4(b) or in the time domain as in Fig. 3.4(c), but not both. Studying the interpolation in one domain will help quantify the impact of node density or sampling rate on the estimation MSE. The analytical asymptotic study is performed for systems operating in AWGN channels.

Due to the symmetry between the space and time domains, it is sufficient to study the interpo-lation in the space domain. From Fig. 3.4(b), the coordinates of the data to be estimated during the spatial interpolation are $\left[\left(i + \frac{1}{2}\right)d, jT_s\right]$, for $i = 0, \ldots, N_s - 1$ and $j = 0, \ldots, N_t - 1$. The asymptotic

61

MSE of the spatial interpolation is given in the following proposition.

Proposition 3.3: When $N_t \to \infty$ and $N_s \to \infty$, while keeping δ and θ finite, the MSE of the estimated data during the spatial interpolation for a network operating in AWGN channels is

$$
\begin{aligned}
\vartheta_s^2 &= \frac{2}{\pi} \cdot \left[\sigma_w^2 + (1 + \sigma_w^2) \frac{\theta \delta}{\gamma_0} \right] \cdot \frac{1 - \rho_s^{\frac{1}{\delta}}}{1 + \rho_s^{\frac{1}{\delta}}} \cdot \frac{1}{\sqrt{(p-1)(q+1)}} \\
&\quad \cdot \left[(p-q)K(\alpha_1) + \frac{(q-v)(p+1)}{v+1} \Pi(\beta_1, \alpha_1) \right]
\end{aligned}
\tag{3.24}
$$

where v, p, q are defined in (3.23),

$$
\alpha_1 = \left[\frac{2(p-q)}{(p-1)(q+1)} \right]^{\frac{1}{2}}, \beta_1 = \frac{2(p-v)}{(p-1)(v+1)},
\tag{3.25}
$$

and $\Pi(\cdot)$ is the complete elliptic integral of the third kind [19].

Proof: The proof is in Appendix 3.7.5. ∎

If we assume the data samples are temporally uncorrelated ($\rho_t = 0$), and perform spatial interpolation based on the spatially correlated but temporally uncorrelated data samples, then the MSE given in Proposition 3.3 can be simplified as follows.

Corollary 3.5: If $\rho_t = 0$, the asymptotic MSE of the estimation for the spatial interpolation is

$$
\vartheta_s^2 = \left[\sigma_w^2 + (1 + \sigma_w^2) \frac{\delta \theta}{\gamma_0} + \frac{1 - \rho_s^{\frac{1}{\delta}}}{1 + \rho_s^{\frac{1}{\delta}}} \right]^{\frac{1}{2}} \left[\sigma_w^2 + (1 + \sigma_w^2) \frac{\delta \theta}{\gamma_0} + \frac{1 + \rho_s^{\frac{1}{\delta}}}{1 - \rho_s^{\frac{1}{\delta}}} \right]^{-\frac{1}{2}}
\tag{3.26}
$$

Proof: *Proof:* When $\rho_t = 0$, we have $\Lambda(\rho_t^{T_s}, f_2) = 1$. Substituting $\Lambda(\rho_t^{T_s}, f_2) = 1$ into (3.52) directly leads to (3.26). ∎

When $\sigma_w^2 = 0$, the result in Corollary 3.5 simplifies to [6, Proposition 2], where only the spatial data correlation is considered. It was proven in [6] that the MSE in (3.26) is a decreasing function of the node density δ.

Fig. 3.5 compares the asymptotic MSE performance between the spatial interpolation and the space-time interpolation. In the simulation, $\rho_t = 0.1$ and $\sigma_w^2 = 0$ and all other parameters are the same as those in Fig. 3.2. As expected, performing interpolation in the space domain alone leads to a better performance compared to interpolation in both the space and time domains. The difference increases as the spatial correlation coefficient, ρ_s, increases. Different from the results in Fig. 3.2, it is observed that the MSE of the spatial interpolation or space-time interpolation is a decreasing function of the spatial node density δ. This can be intuitively explained by the fact that the spatial interpolation depends mainly on the spatial correlation among the sensor nodes, and a higher node density means a stronger spatial correlation among the data samples, thus a better estimation fidelity.
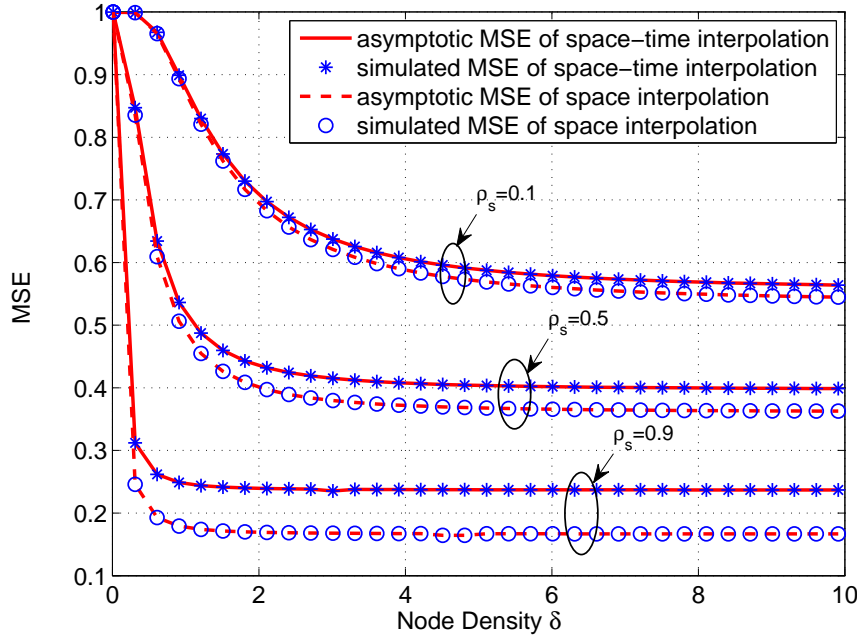

Figure 3.5: The asymptotic MSE of space-time interpolation and space interpolation in the 1-D network (AWGN, $\gamma_0 = 10$ dB, $\sigma_w^2 = 0$, $\rho_t = 0.1$, $\theta = 10$ Hz).

It can be seen from Fig. 3.5 that, when $\delta \rightarrow \infty$, the MSE approaches a lower bound, which is stated in the following corollary.

Corollary 3.6: The following relationship holds for the MSE of the estimation for the data samples σ_{st}^2 and the MSE of the spatial interpolation ϑ_s^2

$$\lim_{\delta\to\infty}\vartheta_s^2 = \left[\frac{2}{\pi}\left(1 - \frac{2\gamma_0}{(1+\sigma_w^2)\theta\log(\rho_s)}\cdot\frac{1+\rho_t^{\frac{1}{\theta}}}{1-\rho_t^{\frac{1}{\theta}}}\right)^{-\frac{1}{2}}\cdot K(\Delta_\delta)\right] \geq \lim_{\delta\to\infty}\sigma_{st}^2, \qquad (3.27)$$

with Δ_δ defined in Corollary 3.2.

Proof: The proof is in Appendix 3.7.6. ∎

Due to the symmetry between the space and the time domains, we can get the MSE of the time interpolation, as shown in Fig. 3.4(c), by exchanging ρ_s with ρ_t, and δ with θ in Proposition 3.3, and Corollaries 3.5 and 3.6.

3.4.4 Optimum spatial-temporal sampling

It can be seen from Fig. 3.5 that, when δ is small, the MSE decreases dramatically as δ increases. When δ reaches a certain threshold, no apparent performance gain can be achieved by increasing δ further, i.e., the slope of ϑ_{st}^2 approaches zero as δ increase. The above statement is also true for the sampling rate θ.

In the space domain, we can find the optimum node density, δ_0, by solving the equation $\left|\frac{\partial\vartheta_{st}^2}{\partial\delta}\right|_{\delta_0} = \varepsilon_s$, with ε_s being a small number. Fig. 3.6 shows the optimum node density in AWGN channels as a function of the spatial correlation coefficient ρ_s, under various values of the sampling rate θ. The parameters are $\rho_t = 0.5$, $\sigma_w^2 = 0$, and $\varepsilon_s = 10^{-3}$. The results in this figure demonstrate that the optimum node density decreases almost linearly as ρ_s increases. Therefore, for the estimation of the spatial interpolation, a smaller node density is required for a field with a stronger spatial correlation. Moreover, the optimal node density converges as the sampling rate θ increases, i.e.,
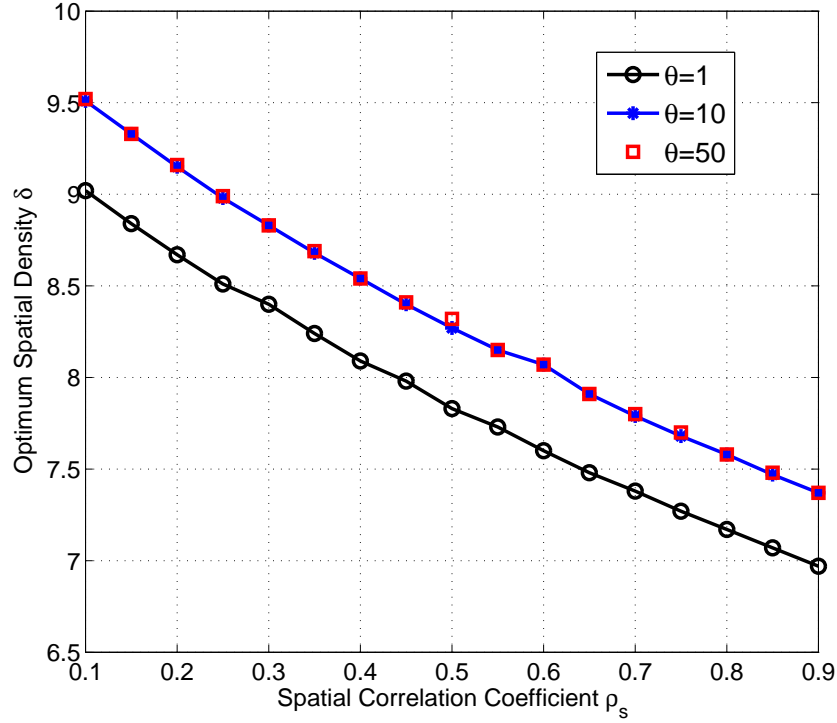
Figure 3.6: The asymptotically optimum spatial node density versus spatial correlation coefficient ρ_s in the 1-D network (AWGN, $\sigma_w^2 = 0$, $\rho_t = 0.5$, $\varepsilon_s = 10^{-3}$)

the optimum node densities are almost identical for $\theta = 10$ and $\theta = 50\,\text{Hz}$. This further corroborates that increasing the sampling rate beyond a certain threshold yields negligible performance gain. Similar results are observed for the optimum sampling rate due to the space-time symmetry.

3.5 Optimum node density in 2-D networks

The impacts of spatial-temporal sampling on the estimation fidelity in a 2-D network, as shown in Fig. 3.1(b), are studied in this section. In the space domain, the N_s sensor nodes are located on a square grid. In the time domain, each sensor collects N_t data samples. The space-time coordinate for the sample x_{ikm} is $[(i-1)d, (k-1)d, (m-1)T_s]$, for $i, k = 1, \ldots, M_s$, $m = 1, \ldots, N_t$, with $M_s = \sqrt{N_s}$. It should be noted that the spatial node density in a 2-D sensor network is $\delta = \frac{1}{d^2}$,

which is different from the 1-D case.

Stacking all the spatial-temporally correlated data samples into a column vector, we have $\boldsymbol{\xi}_{st} = [\mathbf{x}_{11}^T, \ldots, \mathbf{x}_{1M_s}^T, \ldots, \mathbf{x}_{M_s1}^T, \ldots, \mathbf{x}_{M_sM_s}^T]^T \in \mathscr{R}^{N\times1}$, where $\mathbf{x}_{m_1m_2} = [x_{m_1m_21}, \ldots, x_{m_1m_2N_t}]^T \in \mathscr{R}^{N_t\times1}$. The auto-correlation matrix, $\blacksquare_{xx} = \mathbb{E}\left[\boldsymbol{\xi}_{st}\boldsymbol{\xi}_{st}^H\right] \in \mathscr{R}^{N\times N}$, can be represented as

$$\blacksquare_{xx} = \mathbf{R}_{ss} \otimes \mathbf{R}_t \tag{3.28}$$

where $\mathbf{R}_{ss} \in \mathscr{R}^{N_s\times N_s}$ and $\mathbf{R}_t \in \mathscr{R}^{N_t\times N_t}$ are the correlation matrices in the space domain and time domain, respectively. The matrix, $\mathbf{R}_{ss}$, assumes the form of a TBT matrix as defined in [6, Equation (20)] for the 2-D spatially correlated network. The matrix $\mathbf{R}_t$ is a symmetric Toeplitz matrix as in Equation (3.9). Therefore, the matrix, $\blacksquare_{xx}$, is a 3-level Toeplitz matrix [20, Definition 1], i.e., $\blacksquare_{xx}$ has an outermost block Toeplitz structure, and each block is still a block Toeplitz matrix, down to the innermost block with the form of an ordinary Toeplitz matrix.

Mirroring the analysis in the 1-D case, we will study, in the following two sections, the optimum spatial-temporal sampling for the MMSE estimation of the discrete data samples, and the MMSE interpolation, respectively.

3.5.1 MMSE estimation of the discrete samples

With the first-step MMSE estimation in Lemma 3.1, we have the MSE, $\psi_{st,N}^2 = \frac{1}{N}\mathbb{E}\left[\|\hat{\boldsymbol{\xi}}_{st} - \boldsymbol{\xi}_{st}\|^2\right]$, as

$$\psi_{st,N}^2 = \frac{1}{N}\mathbb{E}_\mathbf{H}\left[\text{trace}\left(\blacksquare_{xx}^{-1} + \frac{\mathbf{H}\mathbf{H}^H}{\sigma_w^2\mathbf{H}\mathbf{H}^H + (1+\sigma_w^2)\frac{\theta\delta}{\gamma_0}\mathbf{I}_N}\right)^{-1}\right], \tag{3.29}$$

where $\hat{\boldsymbol{\xi}}_{st}$ is the MMSE estimate of $\boldsymbol{\xi}_{st}$. The above MSE in a fading channel can be evaluated numerically.

Following the same procedure as in 1-D networks, we derive the explicit form of the asymptotic MSE for the system in AWGN channels.

Proposition 3.4: When $N_s \to \infty$ and $N_t \to \infty$, while keeping δ and θ finite, the asymptotic MSE of the discrete space-time samples in a 2-D network transmitted through AWGN channels is

$$\psi_{st}^2 \triangleq \lim_{N \to \infty} \psi_{st,N}^2 = \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} \left[\frac{1}{\Lambda_{ss}(f_1,f_2)\Lambda(\rho_t^{T_s},f_3)} + \frac{1}{\sigma_w^2 + (1+\sigma_w^2)\frac{\theta\delta}{\gamma_0}} \right]^{-1} df_1 df_2 df_3, \quad (3.30)$$

where $\Lambda(a,f)$ is defined in (3.40) in Appendix 1, and

$$\Lambda_{ss}(f_1,f_2) = \sum_{i=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \rho_s^{\sqrt{(i^2+k^2)/\delta}} e^{-j2\pi(if_1+kf_2)}. \quad (3.31)$$

Proof: The proof is in Appendix 3.7.7. ∎

In Proposition 3.4, the impacts of the spatial-temporal sampling rate are expressed through the term, $\frac{1}{\sigma_w^2 + (1+\sigma_w^2)\frac{\theta\delta}{\gamma_0}}$, and the 3-D DTFT, $\Lambda_{xx}'(f_1,f_2,f_3) = \Lambda_{ss}(f_1,f_2)\Lambda(\rho_t^{T_s},f_3)$. The expression of ψ_{st}^2 eludes a closed-form. The non-closed-form expression in (3.30) can be easily evaluated numerically given that the integrals are of finite limits. Even though $\Lambda_{xx}'(f_1,f_2,f_3)$ is expressed as the sum of an infinite series, the value of $\rho_s^{\sqrt{(i^2+k^2)/\delta}}$ decreases exponentially as i and k increase, thus $\Lambda_{xx}'(f_1,f_2,f_3)$ can be accurately approximated with moderate limits on i and k.

If we assume that the data are temporally uncorrelated ($\rho_t = 0$), then the MSE of the data samples in proposition 3.4 can be simplified as follows.

Corollary 3.7: If $\rho_t = 0$, the asymptotic MSE of the data samples in a 2-D network with AWGN channels is

$$\psi_s^2 = \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} \left[\frac{1}{\Lambda_{ss}(f_1,f_2)} + \frac{1}{\sigma_w^2 + (1+\sigma_w^2)\frac{\theta\delta}{\gamma_0}} \right]^{-1} df_1 df_2, \quad (3.32)$$

where $\Lambda_{ss}(f_1,f_2)$ is defined in (3.31).

Proof: Setting $\rho_t = 0$ in (3.40) leads to $\Lambda(\rho_t^{T_s}, f_3) = 1$. Substituting $\Lambda(\rho_t^{T_s}, f_3) = 1$ into (3.30) and solving the integration with respect to f_3, we can obtain (3.32). ∎

The result in Corollary 3.7 simplifies to [6, Proposition 3] with $\sigma_w^2 = 0$, where only the spatial data correlation is considered.

The asymptotic MSE of the data samples in a 2-D network is plotted as a function of the temporal sampling rate θ in Fig. 3.7, under various values of temporal correlation coefficient ρ_t and measurement SNR γ_w. The parameters are $\rho_s = 0.5$ and $\gamma_0 = 10\,\text{dB}$. For comparison, the MSE in an 1-D network is also shown in the figure. It is interesting to note that when the measurement SNR is low ($\gamma_w = 5\,\text{dB}$) and the time correlation is high ($\rho_t = 0.9$), the MSE is decreasing in θ; for all other cases, the MSE is an increasing function in θ. This is because if σ_w^2 is large enough, the majority of the energy is used for transmitting measurement noise. In this case, when increasing θ for data with high temporal correlation, the benefit of data correlation outweighs the loss due to less energy per sample. The performance difference between $\gamma_w = 10\,\text{dB}$ and $\sigma_w^2 = 0$ is very small. In addition, 2-D MSE is larger (worse) than the 1-D MSE. This can be explained by the fact that, under the same spatial node density and temporal sampling rate, each node in the 2-D network needs to cover a larger area than the node in the 1-D network, thus leads to a worse performance.

The asymptotic MSE for 2-D networks in AWGN channels is compared to that in fading channels in Fig. 3.8. Similar to the 1-D case, the MSE with fading channels is worse than its AWGN counterpart. The networks with fading channels and AWGN channels have similar performance trend, and the performance difference between the two gradually diminishes as ρ_s increases. When $\rho_s = 0.9$, the performance in fading and AWGN channels are almost the same at high node density.
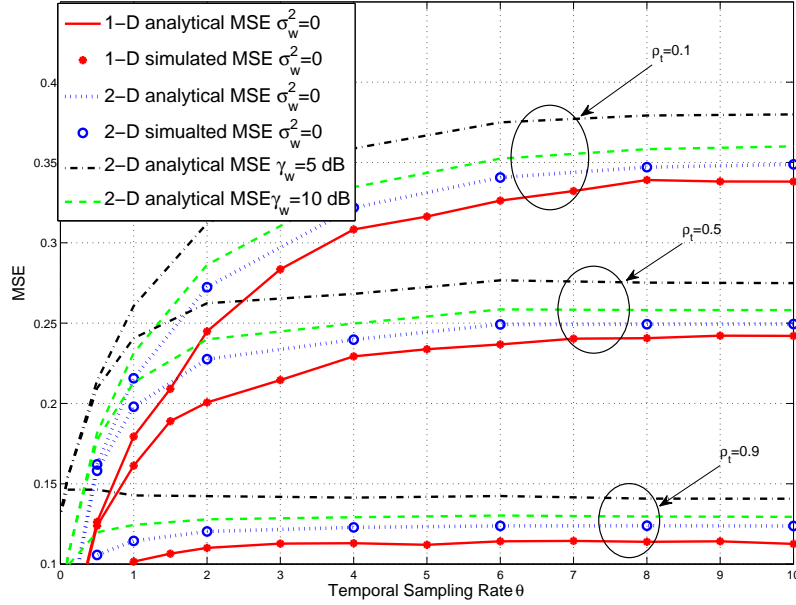
Figure 3.7: The MSE of the estimated discrete data samples in 2-D network. (AWGN, $\gamma_0 = 10$ dB, $\rho_s = 0.5$).

3.5.2 MMSE spatial-temporal interpolation

The performance of spatial-temporal interpolations in a 2-D network is studied in this section. Similar to the 1-D case, we consider the worst case by estimating the data located in the middle of the cube formed by eight adjacent data samples, with the data points to be estimated as $x'_{ikm} = x[(i - \frac{1}{2})d, (k - \frac{1}{2})d, (m - \frac{1}{2})T_s]$, for $i, k = 1, \ldots, \sqrt{N_s}$ and $m = 1, \ldots, N_t$. Correspondingly, the data vector can be expressed as $\boldsymbol{\xi}_{dt} = [\mathbf{x}'^T_{11}, \ldots, \mathbf{x}'_{1M_s}, \ldots, \mathbf{x}'^T_{M_s 1}, \ldots, \mathbf{x}'^T_{M_s M_s}]^T \in \mathscr{R}^{N \times 1}$, where $\mathbf{x}'_{m_1 m_2} = [x'_{m_1 m_2 1}, \ldots, x'_{m_1 m_2 N_t)}]^T \in \mathscr{R}^{N_t \times 1}$.

Following the same procedure as in the 1-D case, the error correlation matrix, $\blacksquare^{(d)}_{ee} = \mathbb{E}[(\hat{\boldsymbol{\xi}}_{dt} - \boldsymbol{\xi}_{dt})(\hat{\boldsymbol{\xi}}_{dt} - \boldsymbol{\xi}_{dt})^H]$, with $\hat{\boldsymbol{\xi}}_{dt}$ being the MMSE estimate of $\boldsymbol{\xi}_{dt}$, can be calculated by

$$\blacksquare^{(d)}_{ee} = \mathbb{E}_{\mathbf{H}} \left\{ \blacksquare_{xx} - \blacksquare_{dx} \mathbf{H}^H \left[\mathbf{H} \blacksquare_{xx} \mathbf{H}^H + \sigma_w^2 \mathbf{H} \mathbf{H}^H + (1 + \sigma_w^2) \frac{\theta \delta}{\gamma_0} \mathbf{I}_N \right]^{-1} \mathbf{H} \blacksquare_{xd}, \right\}, \qquad (3.33)$$

where $\blacksquare_{dd} = \mathbb{E}(\boldsymbol{\xi}_{dt} \boldsymbol{\xi}_{dt}^H) = \blacksquare_{xx}$ is used in the above equation. $\blacksquare_{dx} = \mathbb{E}\left[\boldsymbol{\xi}_{dt} \boldsymbol{\xi}_{st}^H\right]$, and $\blacksquare_{xd} = \blacksquare^H_{dx}$. The
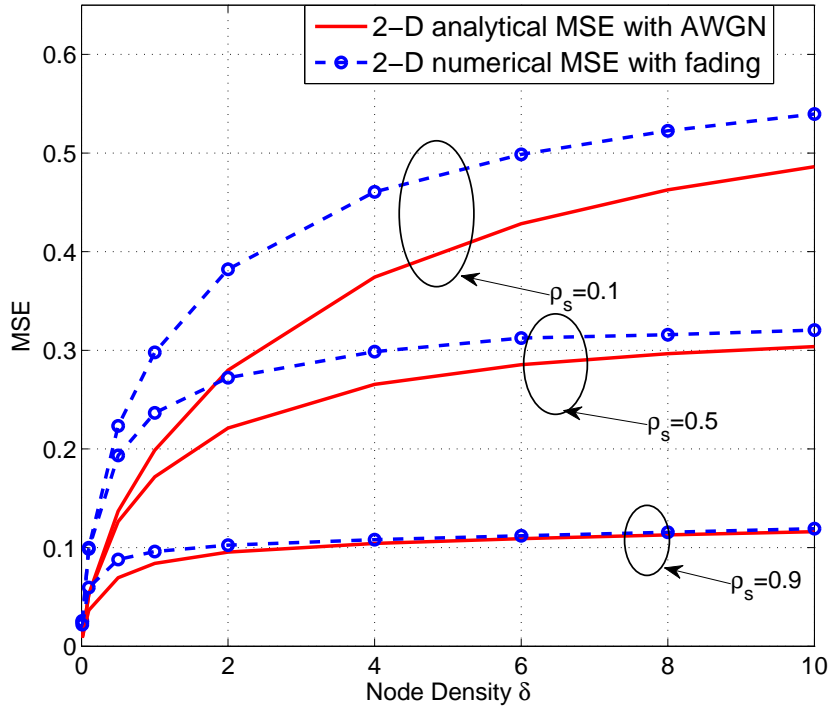
Figure 3.8: The MSE of the estimated discrete data samples in 2-D network with fading. ($\gamma_0 = 10$ dB, $\theta = 10$ Hz, $\sigma_w^2 = 0$).

cross-correlation matrix, $\blacksquare_{dx}$, can be expressed as

$$\blacksquare_{dx} = \mathbf{R}'_{ds} \otimes \mathbf{R}'_t \tag{3.34}$$

where $\mathbf{R}'_{ds} \in \mathscr{R}^{N_s \times N_s}$ and $\mathbf{R}'_t \in \mathscr{R}^{N_t \times N_t}$ are the cross-correlation matrices between the data samples and the interpolations in the space domain and time domain, respectively. The matrix, $\mathbf{R}'_{ds}$, has the form of a non-symmetric TBT matrix as defined in [6, Equation (27)] for the 2-D spatially correlated network. The matrix $\mathbf{R}'_t$ is a Toeplitz matrix defined in Section 3.4.2. The matrix, $\blacksquare_{dx}$, is a non-symmetric 3-level Toeplitz matrix.

For the AWGN case, the asymptotic MSE is given as follows.

Proposition 3.5: When $N_s \to \infty$ and $N_t \to \infty$, while keeping δ and θ finite, the asymptotic MSE

of the space-time interpolations in a 2-D network with AWGN channels is

$$\varphi_{st}^2 = \lim_{N \to \infty} \varphi_{st,N}^2 = \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} \left[\Lambda'_{xx}(f_1, f_2, f_3) - \frac{|\Lambda'_{dx}(f_1, f_2, f_3)|^2}{\Lambda'_{xx}(f_1, f_2, f_3) + \left(\sigma_w^2 + (1 + \sigma_w^2) \frac{\theta \delta}{\gamma_0} \right)} \right] df_1 df_2 df_3.$$

(3.35)

where $\Lambda'_{xx}(f_1, f_2, f_3)$ is defined in (3.55) in Appendix 7, and $\Lambda'_{dx}(f_1, f_2, f_3)$ is

$$\Lambda'_{dx}(f_1, f_2, f_3) = \frac{\rho_t^{\frac{T_s}{2}}(1 - \rho_t^{T_s})(1 + e^{j2\pi f_3})}{1 + \rho_t^{2T_s} - 2\rho_t^{T_s} \cos(2\pi f_3)} \cdot \sum_{i=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \rho_s^{\sqrt{[(i+\frac{1}{2})^2 + (k+\frac{1}{2})^2]/\delta}} e^{-j2\pi(if_1 + kf_2)}.$$

(3.36)

Proof: The proof is in Appendix 3.7.8. ∎

Fig. 3.9 compares the asymptotic MSE of the interpolation in a 2-D network with that in an 1-D network. In the simulation, the parameters are the same as those in Fig. 3.7 except $\sigma_w^2 = 0$. In both 1-D and 2-D networks, it is observed that the interpolation MSE decreases monotonically with the temporal sampling rate. Again, the 1-D asymptotic MSE is smaller (better) than its 2-D counterpart for all temporal correlation coefficients ρ_t. The performance difference between the 1-D and 2-D networks increases as ρ_t increases.

If we just consider the spatial interpolation of the 2-D network, for the special case of uncorrelated data in the time domain, we can simplify the result as follows.

Corollary 3.8: If $\rho_t = 0$, the asymptotic MSE of the estimated data during the spatial interpolations of the 2-D network with AWGN channels is

$$\varphi_s^2 = \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} \left[\Lambda_{ss}(f_1, f_2) - \frac{|\Lambda_{ds}(f_1, f_2)|^2}{\Lambda_{ss}(f_1, f_2) + \sigma_w^2 + (1 + \sigma_w^2) \frac{\theta \delta}{\gamma_0}} \right] df_1 df_2,$$

(3.37)

where $\Lambda_{ss}(f_1, f_2)$ is given in (3.31), and $\Lambda_{ds}(f_1, f_2)$ is

71
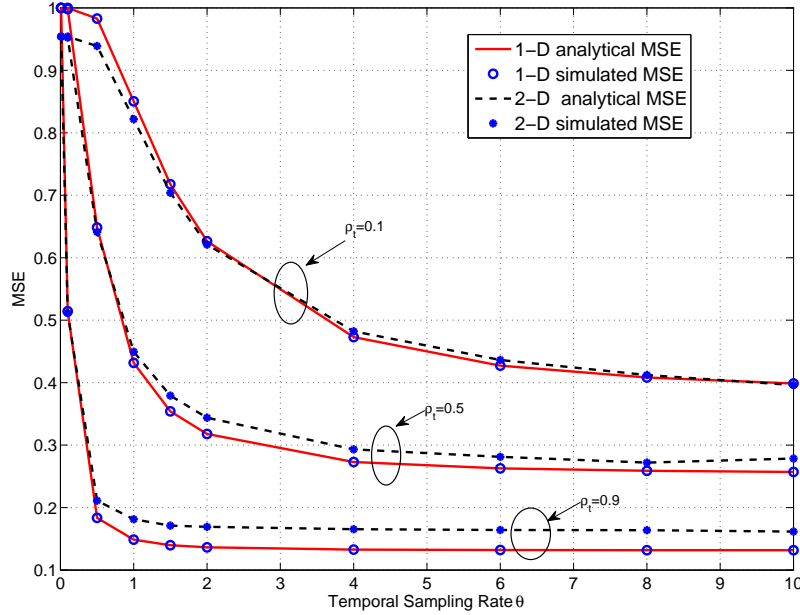
Figure 3.9: The asymptotic MSE of space-time interpolations in the 1-D and 2-D networks (AWGN, $\rho_s = 0.5$, $\sigma_w^2 = 0$, $\gamma_0 = 0$ dB).

$$\Lambda_{ds}(f_1, f_2) = \sum_{i=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \rho_s^{\sqrt{[(i+\frac{1}{2})^2 + (k+\frac{1}{2})^2]/\delta}} e^{-j2\pi(if_1 + kf_2)} \tag{3.38}$$

Proof: The proof is in Appendix 3.7.9. ∎

The result in Corollary 3.8 with $\sigma_w^2 = 0$ simplifies to [6, Proposition 4], where only the spatial data correlation is considered.

3.5.3 Optimum spatial-temporal sampling

The asymptotically optimum spatial and temporal sampling rates in a 2-D network can be obtained by numerically solving $|\frac{\partial \varphi_{st}^2}{\partial \delta}| = \varepsilon_s$ and $|\frac{\partial \varphi_{st}^2}{\partial \theta}| = \varepsilon_t$, with ε_s and ε_t being very small numbers. Fig. 3.10 shows the asymptotically optimum temporal sampling rate as a function of the temporal correlation coefficient in the 1-D and 2-D networks with AWGN channels. In the figure, $\rho_s = 0.5$,
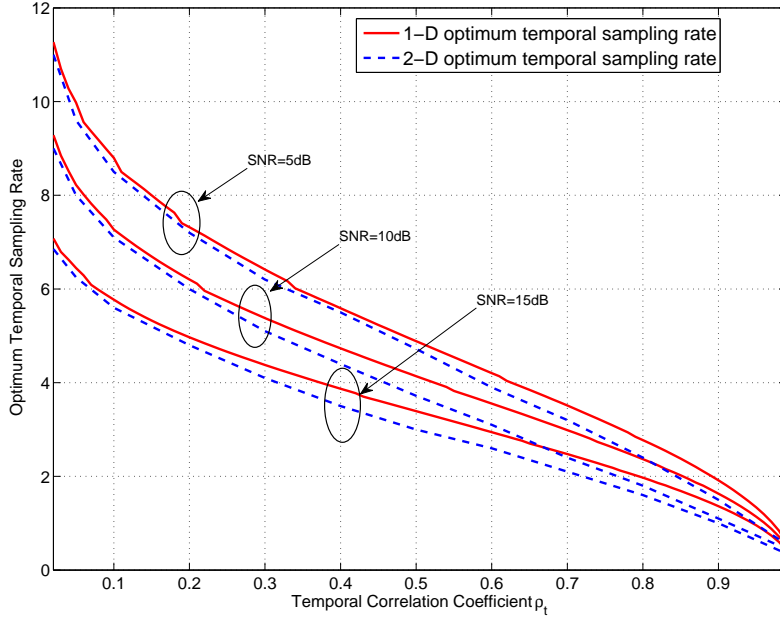
Figure 3.10: The asymptotically optimum sampling rate in the 1-D and 2-D networks (AWGN, $\rho_s = 0.5$, $\sigma_w^2 = 0$, $\varepsilon = 10^{-3}$).

$\sigma_w^2 = 0$, and $\varepsilon = 10^{-3}$ are used for both 1-D and 2-D networks. It is observed that the asymptotically optimum sampling rate for the 1-D and 2-D networks are almost identical, with the optimum sampling rate in the 1-D network slightly larger.

It should be noted that the analysis methods presented in this chapter can be extended to high dimensional networks by employing block multilevel Toeplitz matrix. In this chapter, the 1-D and 2-D networks are used as examples to investigate the interactions among the various network parameters and their impacts on the system performance. The results of high dimensional networks can be obtained in a similar manner.

3.5.4 Randomly distributed networks

So far all the studies are for networks with deterministic topologies. In this section, we will compare the MSE performance between networks with deterministic topology and random topology,

respectively. The random topology follows a Poisson point process, i.e., the number of nodes in a given area follows a Poisson distribution, and the coordinates of each node follows a uniform distribution in each dimension.

The MSE of the 1-D and 2-D networks with random topology can be evaluated numerically through a hybrid analytical-simulation method. The MSE conditioned on a particular deployment of the nodes can be calculated by using (3.12) for the 1-D network, or (3.29) for the 2-D network. The elements in the autocorrelation matrix, $\mathbf{R}_{xx}$ or $\mathbf{\Phi}_{xx}$, depends on the actual locations of the nodes. The unconditional MSE can then be calculated by averaging a large number of random deployments.

Fig. 3.11 compares the performance of networks with random topology and deterministic topology, operating in AWGN channels. The parameters are $\gamma_0 = 10\,\text{dB}$ and $\sigma_w^2 = 0$. For both 1-D and 2-D networks, networks with deterministic topology consistently outperform their random topology counterparts. The difference between the two types of networks becomes smaller as ρ_s and ρ_t increase. The topology of practical networks is usually a combination of the grid-based deterministic topology and random topology. Therefore the performance of practical networks will fall between the bounds delimited by the two types of networks.

3.6   Conclusions

In this chapter, the optimum sampling in the 1-D and 2-D WSNs with spatial-temporally correlated data was studied. The impacts of the spatial node density and the temporal sampling rate on the network performance were investigated through asymptotic analysis and numerical studies. Under the constraint of fixed power per unit area, the MSE performance of various networks were studied
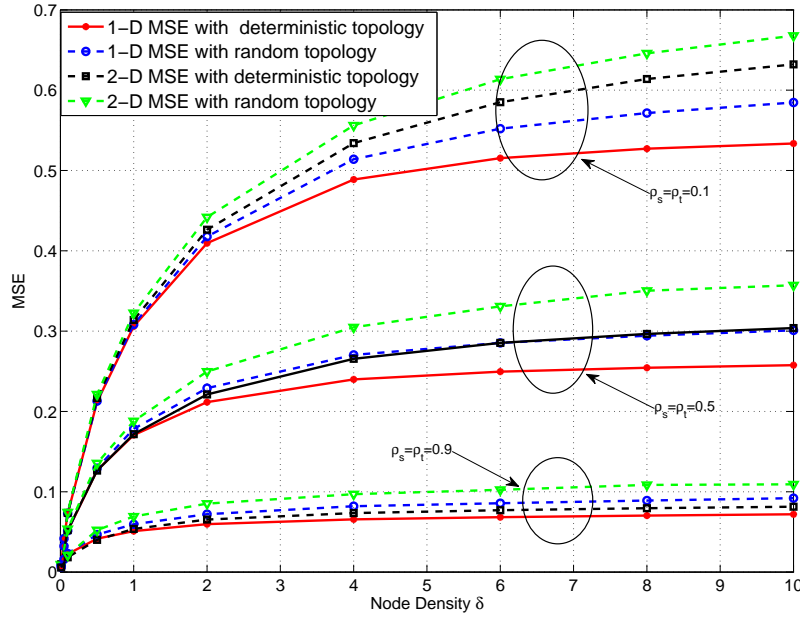
Figure 3.11: The MSE of the estimated discrete data samples in 1-D and 2-D Possion distributed networks (AWGN, $\gamma_0 = 10$ dB, $\theta = 10$ Hz, $\sigma_w^2 = 0$).

through a combination of analytical and simulation methods. The results quantitatively identified the interactions between the estimation fidelity and a large number of system parameters, such as node density, sampling rate, measurement noise, fading, and random topology, etc. It was observed that the network with a deterministic grid-based topology and operating in AWGN channels has the best performance, yet that with a random topology and operating in fading channels has the worst performance. Therefore, whenever possible, a grid-based deterministic topology is preferred over a random topology. The MSE performance of these two types of networks can serve as lower and upper bounds for practical networks, and their difference gradually diminish as the correlation coefficients increase.

3.7 Appendix of Proofs

3.7.1 Proof of Proposition 3.1

Setting $\mathbf{H} = \mathbf{I}_N$ in (3.12) and performing the eigenvalue decomposition of $\mathbf{R}_{xx}$, we can rewrite the

MSE as

$$\sigma_{st,N}^2 = \frac{1}{N} \sum_{m=1}^{N_s-1} \sum_{k=1}^{N_t-1} \left(\frac{1}{\lambda_{m,k}} + \frac{1}{\sigma_w^2 + (1+\sigma_w^2)\frac{\theta\delta}{\gamma_0}} \right)^{-1}, \tag{3.39}$$

where $\lambda_{m,k}$, for $m = 0, 1, \ldots, N_s - 1$, and $k = 0, 1, \ldots, N_t - 1$, are the eigenvalues of $\mathbf{R}_{xx}$. When $N_s \to$

∞ and $N_t \to \infty$, the 2-D discrete-time Fourier transform (DTFT) of the sequence, $\left\{ \rho_s^{|m|d} \rho_t^{|k|T_s} \right\}_{m,k}$,

which are elements of the TBT matrix $\mathbf{R}_{xx}$, can be calculated as $\Lambda_{xx}(f_1, f_2) = \Lambda(\rho_s^d, f_1) \times \Lambda(\rho_t^{T_s}, f_2)$,

where

$$\Lambda(a, f) = \sum_{m=-\infty}^{+\infty} a^{|m|} e^{-j2\pi m f} = \frac{1-a^2}{1+a^2 - 2a\cos(2\pi f_1)}, \tag{3.40}$$

Based on the extension of the Szego's theorem to TBT matrices [18, Theorem 1], when $N_s \to \infty$

and $N_t \to \infty$, the asymptotic MSE is

$$\sigma_{st}^2 = \lim_{N \to \infty} \sigma_{st,N}^2 = \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} \left[\frac{1}{\Lambda_{xx}(f_1, f_2)} + \frac{1}{\sigma_w^2 + (1+\sigma_w^2)\frac{\theta\delta}{\gamma_0}} \right]^{-1} df_1 df_2, \tag{3.41}$$

Substituting the result of $\Lambda_{xx}(f_1, f_2)$ into (3.41), and applying [19, Equation (2.553.3)], we can

solve the inner integral as

$$\sigma_{st}^2 = \frac{1-\rho_t^{2T_s}}{2\rho_t^{T_s}} \int_{-\frac{1}{2}}^{\frac{1}{2}} \left[\cos^2(2\pi f_2) - a_2 \cos(2\pi f_2) + b_2 \right]^{-\frac{1}{2}} f_2, \tag{3.42}$$

where

$$a_2 = \left[1 + \rho_t^{2T_s} + \frac{1}{\sigma_w^2 + (1+\sigma_w^2)\frac{\theta\delta}{\gamma_0}} \cdot \frac{1+\rho_s^{2d}}{1-\rho_s^{2d}} \cdot (1-\rho_t^{2T_s}) \right] \cdot \rho_t^{-T_s}, \tag{3.43a}$$

$$b_2 = \left[\frac{2(1+\rho_t^{2T_s})^2}{\sigma_w^2 + (1+\sigma_w^2)\frac{\theta\delta}{\gamma_0}} \cdot \frac{1+\rho_s^{2d}}{1-\rho_s^{2d}} \cdot (1-\rho_t^{4T_s}) + \left(\frac{1-\rho_t^{2T_s}}{\sigma_w^2 + (1+\sigma_w^2)\frac{\theta\delta}{\gamma_0}} \right)^2 \right] \cdot (4\rho_t^{-2T_s}) \tag{3.43b}$$

Based on [19, Equation (2.580.2)] and [19, Equation (3.152.2)], we can get the results in (3.13).

3.7.2 Proof of Corollary 3.2

The MSE in (3.13) can be alternatively written as

$$\sigma_{st}^2 = \frac{\sqrt{2}}{\pi} \cdot \int_0^{\frac{\pi}{2}} \frac{1}{\sqrt{\beta - \alpha \sin^2 x}} dx \tag{3.44}$$

Since integration is a linear operator, we can directly find the limit of the integrand, and the result

is

$$\lim_{\delta \to \infty} \left(\frac{1}{\sqrt{\beta - \alpha \sin^2 x}} \right) = \sqrt{2 \log(\rho_s)} \left[\log(\rho_s) - \frac{2\gamma_0}{\theta(1 + \sigma_w^2)} \cdot \frac{1 + \rho_t^{\frac{1}{\theta}}}{1 - \rho_t^{\frac{1}{\theta}}} - \frac{8\gamma_0}{\theta(1 + \sigma_w^2)} \cdot \frac{\rho_t^{\frac{1}{\theta}}}{1 - \rho_t^{\frac{2}{\theta}}} \sin^2 x \right]^{-\frac{1}{2}}. \tag{3.45}$$

Substituting (3.45) into (3.44) and simplifying lead to (3.16).

3.7.3 Proof of Corollary 3.4

The limit in (3.17) can be rewritten as

$$\frac{2}{\pi} \left(1 + \frac{4}{\varpi} \right)^{-\frac{1}{2}} \cdot K \left(\sqrt{\frac{4}{4 + \varpi}} \right) = \frac{2}{\pi} \int_0^{\frac{\pi}{2}} \frac{dw}{\sqrt{1 + \frac{4}{\varpi}(1 - \sin^2 w)}}. \tag{3.46}$$

Since $(1 - \sin^2 w)$ is a non-negative real number, the limit is an increasing function of ϖ, thus proportional to ρ_s and ρ_t, but inverse proportional to the SNR γ_0.

3.7.4 Proof of Proposition 3.2

The Toeplitz matrix, $\mathbf{R}'_s$, is uniquely determined by the sequence $\left\{ \rho_s^{|m+\frac{1}{2}|d} \right\}_{m=-(N_s-1)}^{(N_s-1)}$. Similarly, the Toeplitz matrix, $\mathbf{R}'_t$, is uniquely determined by the sequence, $\left\{ \rho_t^{|m+\frac{1}{2}|T_s} \right\}_{m=-(N_t-1)}^{(N_t-1)}$. Based on [21], when $N_s \to \infty$ and $N_t \to \infty$, the 2-D DTFT of the sequence, $\left\{ \rho_s^{|m+\frac{1}{2}|d} \rho_t^{|k+\frac{1}{2}|T_s} \right\}_{m,k}$, which

are elements of the TBT matrix $\mathbf{R}_{dx}$, can be calculated as $\Lambda_{dx}(f_1, f_2) = \Lambda'(\rho_s{}^d, f_1) \times \Lambda'(\rho_t{}^{T_s}, f_2)$,

where

$$\Lambda'(a, f) = a^{\frac{1}{2}} \frac{(1-a)(1+e^{j2\pi f})}{1+a^2 - 2a\cos(2\pi f)}. \tag{3.47}$$

Based on [18, Lemma 1], $\mathbf{R}_{dx}$ is asymptotically equivalent to a circulant-block-circulant (CBC)

matrix, $\mathbf{C}_{dx} = \mathbf{U}_N^H \mathbf{D}_{dx} \mathbf{U}_N$, where $\mathbf{U}_N^H$ is the unitary discrete Fourier transform (DFT) matrix and $\mathbf{D}_{dx}$

is a diagonal matrix with its kth diagonal element being

$$(\mathbf{D}_{dx})_{k,k} = \Lambda'\left(\rho_s^d, \frac{k-1}{N_s}\right) \cdot \Lambda'\left(\rho_t^{T_s}, \frac{k-1}{N_t}\right). \tag{3.48}$$

Similarly, the TBT matrix, $\mathbf{R}_{xx}$, is asymptotically equivalent to a CBC matrix, $\mathbf{C}_{xx} = \mathbf{U}_N^H \mathbf{D}_{xx} \mathbf{U}_N$,

where $\mathbf{D}_{xx}$ is a diagonal matrix with its kth diagonal element being $(\mathbf{D}_{xx})_{k,k} = \Lambda\left(\rho_s^d, \frac{k-1}{N_s}\right) \cdot$

$\Lambda\left(\rho_t^{T_s}, \frac{k-1}{N_t}\right)$, with $\Lambda(\rho^d, f)$ defined in (3.40).

In addition, the CBC matrices, $\mathbf{C}_{xx}$ and $\mathbf{C}_{dx}$, share the same orthonormal eigenvectors [22].

Based on [23, Theorem 2.1], the error correlation matrix, $\mathbf{R}_{ee}^{(d)}$, is asymptotically equivalent to a

CBC matrix, $\mathbf{C}_{ee}^{(d)} = \mathbf{C}_{xx} - \mathbf{C}_{dx}\left(\mathbf{C}_{xx} + (\sigma_w^2 + \frac{\theta\delta}{\gamma_0})\mathbf{I}\right)^{-1}\mathbf{C}_{dx}^H = \mathbf{U}_N^H \mathbf{D}_{ee}^{(d)} \mathbf{U}_N$, where the diagonal ma-

trix $\mathbf{D}_{ee}^{(d)} = \mathbf{D}_{xx} - \mathbf{D}_{dx}\left(\mathbf{D}_{xx} + (\sigma_w^2 + \frac{\theta\delta}{\gamma_0})\mathbf{I}_N\right)^{-1}\mathbf{D}_{dx}^H$.

Based on the extension of the Szego's theorem to TBT matrices [18, Theorem 1], we have

$$\vartheta_{st}^2 = \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} \left[\Lambda_{xx}(f_1, f_2) - \frac{|\Lambda_{dx}(f_1, f_2)|^2}{\Lambda_{xx}(f_1, f_2) + \sigma_w^2 + (1+\sigma_w^2)\frac{\theta\delta}{\gamma_0}} \right] df_1 df_2. \tag{3.49}$$

With [19, Equation (2.559.2)], we can solve the inner integral, and the result is

$$
\begin{aligned}
\vartheta_{st}^2 = \int_{-\frac{1}{2}}^{\frac{1}{2}} & \left[\Lambda(\rho_t^{T_s}, f) - \frac{|\Lambda'(\rho_t^{T_s}, f)|^2}{\Lambda(\rho_t^{T_s}, f)} + \frac{|\Lambda'(\rho_t^{T_s}, f)|^2}{\Lambda(\rho_t^{T_s}, f)} \left(\sigma_w^2 + (1+\sigma_w^2)\frac{\theta\delta}{\gamma_0} + \frac{1-\rho_s^{\frac{1}{\delta}}}{1+\rho_s^{\frac{1}{\delta}}}\Lambda(\rho_t^{T_s}, f) \right)^{\frac{1}{2}} \right. \\
& \left. \left(\sigma_w^2 + (1+\sigma_w^2)\frac{\theta\delta}{\gamma_0} + \frac{1+\rho_s^{\frac{1}{\delta}}}{1-\rho_s^{\frac{1}{\delta}}}\Lambda(\rho_t^{T_s}, f) \right)^{-\frac{1}{2}} \right] df
\end{aligned} \tag{3.50}
$$

From [19, Equation (2.558.2)], we get

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \left[\Lambda(\rho_t^{T_s}, f_2) - \frac{|\Lambda'(\rho_t^{T_s}, f_2)|^2}{\Lambda(\rho_t^{T_s}, f_2)} \right] df_2 = \frac{1 - \rho_t^{\frac{1}{\theta}}}{1 + \rho_t^{\frac{1}{\theta}}}. \tag{3.51}$$

Substituting (3.51) into (3.50) and simplifying lead to (3.22).

3.7.5   Proof of Proposition 3.3

The result in (3.24) can be proved by following a procedure that is similar to the proof of Proposition 3.2. Since the interpolation is performed in the space domain alone, we can replace $\Lambda'(\rho_t^{T_s}, f_2)$ with $\Lambda(\rho_t^{T_s}, f_2)$ in (3.50), and the result is

$$\begin{aligned}
\vartheta_s^2 &= \int_{-\frac{1}{2}}^{\frac{1}{2}} \left[\Lambda(\rho_t^{T_s}, f_2) \left(\sigma_w^2 + (1 + \sigma_w^2) \frac{\theta\delta}{\gamma_0} + \frac{1 - \rho_s^{\frac{1}{\delta}}}{1 + \rho_s^{\frac{1}{\delta}}} \Lambda(\rho_t^{T_s}, f_2) \right)^{\frac{1}{2}} \right. \\
&\quad \left. \left(\sigma_w^2 + (1 + \sigma_w^2) \frac{\theta\delta}{\gamma_0} + \frac{1 + \rho_s^{\frac{1}{\delta}}}{1 - \rho_s^{\frac{1}{\delta}}} \cdot \Lambda(\rho_t^{T_s}, f_2) \right)^{-\frac{1}{2}} \right] df_2
\end{aligned} \tag{3.52}$$

The above integral can be solved by using [19, Equation (3.147.2)], [19, Equation (3.151.2)], and the definition of $\Lambda(\rho_t^{T_s}, f_2)$ in (3.40), and the result is (3.24).

3.7.6   Proof of Corollary 3.6

Setting $\delta \to \infty$ in (3.52) leads to

$$\lim_{\delta \to \infty} \vartheta_s^2 = \int_{-\frac{1}{2}}^{\frac{1}{2}} \left[\Lambda(\rho_t^{T_s}, f_2) \cdot \left(1 - \frac{\Lambda(\rho_t^{T_s}, f_2)}{\log(\rho_s)} \right)^{-\frac{1}{2}} \right] df_2 \tag{3.53}$$

The above integral can be solved by using [19, Equation (3.147.2)].

3.7.7 Proof of Proposition 3.4

Setting $\mathbf{H} = \mathbf{I}_N$ in (3.29) and performing the eigenvalue decomposition of $\blacksquare_{xx}$ in (3.28), we have

$$\psi_{st,N}^2 = \frac{1}{N} \sum_{i=1}^{M_s} \sum_{k=1}^{M_s} \sum_{m=1}^{N_t} \left(\frac{1}{\lambda_{ikm}} + \frac{1}{\sigma_w^2 + (1+\sigma_w^2)\frac{\theta\delta}{\gamma_0}} \right)^{-1}, \tag{3.54}$$

where λ_{ikm}, for $i,k = 1, \ldots, M_s$, and $m = 1, \ldots, N_t$, are eigenvalues of $\blacksquare_{xx}$. When $N_s \to \infty$ and

$N_t \to \infty$, the 3-D DTFT of the sequence, $\left\{ \rho_s^{\sqrt{(i^2+k^2)/\delta}} \rho_t^{|m|T_s} \right\}_{ikm}$, which are elements of the 3-

level Toeplitz matrix $\blacksquare_{xx}$, can be calculated as

$$\Lambda'_{xx}(f_1, f_2, f_3) = \Lambda_{ss}(f_1, f_2) \times \Lambda(\rho_t^{T_s}, f_3). \tag{3.55}$$

The result in (3.30) follows immediately from (3.55) and [20, Theorem 1], which is the extension of the Szego's theorem to multilevel Toeplitz matrices.

3.7.8 Proof of Proposition 3.5

According to [20, Lemma 2], the multilevel Toeplitz matrices, $\blacksquare_{xx}$ and $\blacksquare_{dx}$, are asymptotically equivalent to multilevel circulant matrices, $\mathbf{B}_{xx}$ and $\mathbf{B}_{dx}$, respectively, where the eigenvalues of $\mathbf{B}_{xx}$ and $\mathbf{B}_{dx}$ are samples of $\Lambda'_{xx}(f_1, f_2, f_3)$ in (3.55) and $\Lambda'_{dx}(f_1, f_2, f_3)$ in (3.36), respectively. In addition, the multilevel circulant matrices, $\mathbf{B}_{xx}$ and $\mathbf{B}_{dx}$, share the same orthonormal eigenvectors [20]. Once the asymptotic equivalence is established, the rest of the proof follows the same procedure as described in Appendix E for the 1-D case.

3.7.9 Proof of Corollary 3.8

When $N_s \to \infty$ and $N_t \to \infty$, while keeping δ and θ finite, the asymptotic MSE of spatial interpolations in a 2-D network is

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} \left[\Lambda_{ss}(f_1, f_2)\Lambda_t(f_3) - \frac{|\Lambda_{ds}(f_1, f_2)\Lambda_t(f_3)|^2}{\Lambda_{ss}(f_1, f_2)\Lambda_t(f_3) + \sigma_w^2 + (1 + \sigma_w^2)\frac{\theta\delta}{\gamma_0}} \right] df_1 df_2 df_3. \quad (3.56)$$

where $\Lambda_{ss}(f_1, f_2)$ is defined in (3.31) and $\Lambda_{ds}(f_1, f_2)$ is computed as in (3.38). When $\rho_t = 0$, we

have $\Lambda_t(f_3) = 1$. Substituting $\Lambda_t(f_3) = 1$ into (3.56) directly leads to (3.37).

3.8 Appendix of the Copyright

3.8.1 Documentation of multi-authored chapter

UNIVERSITY OF ARKANSAS.

College of Engineering
Department of Electrical Engineering

July 31, 2013

To Whom It May Concern

This letter is to certify that Ms. Ning Sun, a Ph.D. candidate under my supervision at the Department of Electrical Engineering, has contributed more than 51% of the work for the following paper,

Ning Sun and Jingxian Wu, "Optimum sampling in spatial-temporally correlated wireless sensor networks," EURASIP Journal on Wireless Communication Networking, doi:10.1186/1687-1499-2013-5, January 2013.

Ms. Sun is the first author of this paper. The paper is included in her Ph.D. dissertation entitled "Distortion-Tolerant Communications with Correlated Information".

Sincerely Yours,

Jingxian Wu
Director, Wireless Information Network Lab
Associate Professor of Electrical Engineering
University of Arkansas
Email: wuj@uark.edu
Tel: (479) 575-6584

3217 Bell Engineering Center • Fayetteville, AR 72701 • 479-575-3005 • Fax: 479-575-7967 • www.uark.edu
The University of Arkansas is an equal opportunity/affirmative action institution.

82

3.8.2 Copyright Clearance

3.9 References

[1] J Wu, G Zhou, A new ultra-low power wireless sensor network with integrated energy harvesting, data sensing, and wireless communication, in *Proc. IEEE Intern. Conf. Commun ICC'11* (2011)

[2] JF Chamberland, VV Veeravalli, How dense should a sensor network be for detection with correlated observations? IEEE Trans. Inf. Theory **52**, 5099–5106 (2006)

[3] SD Servetto, On the feasibility of large scale wireless sensor networks, in *Proc. IEEE Int. Conf. Commu., Control, and Computing* (2002)

[4] Y Sung, HV Poor, H Yu, Optimal node density for two-dimensional sensor arrays, in *Proc. IEEE Sensor Array Multichannel Signal Processing Workshop* (2008) pp. 271–274

[5] J Wu, N Sun, Optimal sensor density in a distortion-tolerant linear wireless sensor network, in *Proc. IEEE Global Telecommun. Conf. GLOBECOM'10* (2010)

[6] J Wu, N Sun, Optimum sensor density in distortion tolerant wireless sensor networks. IEEE Trans. Wirel. Commun. **11**, 2056–2064 (2012)

[7] MC Vuran, OB Akan, IF Akyildiz, Spatio-temporal correlation: theory and applications for wireless sensor networks. Comput. Netw. J. (Elsevier) **45**(3), 245–261 (2004)

[8] R Cristescu, M Vetterli, On the optimal density for real-time data gathering of spatio-temporal processes in sensor networks. IEEE Trans. Sig. Proc. **51**, 159–164 (2003)

[9] X Zhang, H Wang, FN Abdesselam, AA Khokhar Distortion analysis for real-Time data collection of spatially temporally correlated data fields in wireless sensor networks. IEEE Trans. Veh. Technol. **58**, 1583–1594 (2009)

[10] S Bandyopadhyay, Q Tian, EJ Coyle, Spatio-temporal sampling rates and energy efficiency in wireless sensor networks. IEEE Trans. Network. **13**, 1339–1352 (2005)

[11] M Gastpar, M Vetterli, Power, spatio-temporal bandwidth, and distortion in large sensor networks. IEEE J. Sel. Areas Commun. **23**, 745–755 (2005)

[12] N Sun, J Wu, Optimum sampling in a spatial-temporally correlated wireless sensor network, in *Proc. IEEE Global Telecommun. Conf. Globecom'11* (2011) pp. 1–5

[13] Q Zhao, L Tong, Opportunistic carrier sensing for energy-efficient information retrieval in sensor networks. EURASIP J. Wirel. Commun. Network. 231–241 (2005)

[14] S Appadwedula, V Veeravalli, D Jones, Energy-efficient detection in sensor networks. IEEE J. Sel. Area Commun. **23**, 693–702 (2005)

[15] JF Chamberland, VV Veeravalli, Asymptotic results for decentralized detection in power constrained wireless sensor networks. IEEE J. Sel. Areas Commun. **22**, 1007–1015 (2004)

[16] J Wu, Ultra-low power compressive wireless sensing for distributed wireless networks, in *Proc. IEEE Mil. Commun. Conf. MILCOM'09* (2009) pp. 1–7

[17] SM Kay, *Fundamentals of Statistical Signal Processing, vol. I, Estimation Theory* (Prentice-Hall, NJ, 1993)

[18] PA Voois, A theorem on the asymtotic eigenvalue distribution of Toeplitz-block-Toeplitz matrix. IEEE Trans. Sig. Proc. **44**, 1837–1841 (1996)

[19] IS Gradshteyn, IM Ryzhik, *Table of Integrals, Series, and Products*, 6th edn. (Academic Press, 2000)

[20] M Oudin, JP Delmas, Asymptotic generalized eigenvalue distribution of block multilevel toeplitz matrices. IEEE Trans. Sig. Proc. **57**, 382–387 (2009)

[21] H Gazzah, PA Regalia, JP Delmas, Asymptotic eigenvalue distribution of block Toeplitz matrices and application to blind SIMO channel identification. IEEE Trans. Inf. Theory **47**, 1243–1251 (2001)

[22] T Mazancourt, D Gerlic, The inverse of a block-circulant matrix. IEEE Trans. Antennas Propag. **AP-31**, 808–810 (1983)

[23] RM Gray, Toeplitz and Circulant Matrices: A Review, NOW Publishers, 2006 (available online at http://ee.stanford.edu/ gray/)

Distributed Joint Source and Channel Code with Correlated Information Sources

Ning Sun , Jingxian Wu, and Hai Lin

4.1 Abstract

In this chapter, a new distributed joint source-channel code (DJSCC) is proposed for a communication network with multiple correlated information sources. The DJSCC is performed by puncturing the information bits of a linear block code but leaving the parity bits intact, given the fact that the correlation among the parity bits is usually much lower compared to the corresponding information bits. In recognition of the different roles of the information and parity bits in the DJSCC scheme, we propose to allocate unequal amounts of energy per bit to these two different types of bits. The unequal energy allocation leads to significant performance gains over conventional equal energy transmissions. At the receiver, the sources are jointly decoded with the iterative message passing algorithm. Simulation results demonstrate that the proposed scheme can achieve considerable performance gains over conventional schemes.

4.2 Introduction

The Slepian-Wolf (S-W) theorem [1] states that distributed sources with correlated information can perform encoding separately, yet achieve a code rate that is the same as when the information is encoded jointly. A wide range of applications can benefit from the S-W theorem. For example, in a wireless sensor network, data collected by the spatially distributed sensors are usually correlated due to the redundancy of the underlying monitored object; in a wireless relay network, the signal

transmitted by one source might be observed by multiple relays, the information to be transmitted by which is thus correlated. However, the S-W theorem is not constructive, *i.e.*, it provides no practical coding scheme to achieve the optimum performance.

There have been considerable works in the literature devoted to the design of practical distributed source codes (DSC) [2]-[11]. Many practical DSC schemes are designed by using the syndromes of channel codes, such as block and trellis codes [2], turbo codes [3], and low-density parity-check (LDPC) codes [4]. Many syndrome-based DSC designs focus on the asymmetric scenario, *i.e.*, the distributed coding is only applied to one of the sources, and the other source is used as side-information and assumed to be known perfectly at the decoder. Designs of symmetric DSCs are discussed in [5]–[8], with punctured linear block codes or LDPC codes. All of the above work assume distortion-free communications between the encoder and decoder, which rely on a separately designed ideal channel code to protect the signal from channel distortions. It is shown in [9] that the source-channel coding separation theorem does not hold for a multiuser network, thus necessitates the design of distributed joint source-channel code (DJSCC). In [10], a Raptor code is employed for an asymmetric DJSCC over a packet erasure channel, where a correlated video source is assumed to be available at the receiver for decoding. A symmetric DJSCC scheme over the additive white Gaussian channel is proposed in [11].

In this chapter, we propose a new symmetric DJSCC coding scheme for multiple correlated sources. The source correlation is utilized to both reduce the energy consumption and to protect the information from channel distortion. The DJSCC is performed by transmitting a subset of the information bits and all the parity bits of a linear block code over a noisy channel. The distortions from source coding and channel impairments can be partly recovered by using the source correlation and the parity bits. Compared to existing schemes in the literature, the newly proposed

87

DJSCC scheme has the following contributions. First, a new unequal energy allocation scheme is proposed for the delivery of the codeword from the transmitter to the receiver. The information bits and parity bits are transmitted with different energy per bit in recognition of their different roles in DJSCC, and significant performance gains are achieved over conventional schemes equal energy allocation. Second, unlike many of the symmetric DSC or DJSCC schemes that puncture both the information and parity bits [7] and [11], only the information bits are punctured in the proposed coding scheme. This is based on our observation that the correlation among the parity bits from different sources are relatively low even if the correlation of the information bits is strong. Therefore puncturing the parity bits with low mutual correlation might deteriorate the overall system performance. Simulations are performed by using the LDPC codes as the constituent code, and the results demonstrate significant performance gains of the newly proposed DJSCC scheme.

The remainder of the chapter is organized as follows. Section 4.3 introduces the proposed DJSCC scheme with unequal energy allocation. Section 4.4 presents the message passing decoding algorithm. Simulation results are given in Section 4.5, and Section 4.6 concludes the chapter.

4.3 Distributed Joint Source and Channel Code

Consider a network with N spatially distributed sources transmitting to an information sink. Denote $b_n(k) \in \mathscr{B}$ as the k-th information bit from the n-th source, where $\mathscr{B} = \{0, 1\}$. The binary information of the N sources are mutually correlated. Define the cross probability between users m and n as $p_{mn} = P\{b_m(k) \neq b_n(k)\}$. If the binary information is equal probable, *i.e.* $P(b_m(k) = 1) = P(b_m(k) = 0) = 0.5$, then the covariance coefficient between $b_m(k)$ and $b_n(k)$

is

$$\rho_{mn} = \frac{\mathbb{E}\left[(b_m - \mu_m)(b_n - \mu_n)\right]}{\sigma_m \sigma_n} = 2(1 - p_{mn}) - 1, \tag{4.1}$$

where μ_m and σ_m are the mean and standard deviation of $b_m(k)$, respectively.

4.3.1 Codeword Structure

Each source encodes its own information *without* the knowledge of the information from the other sources. The proposed DJSCC is a linear block code. Let $\mathbf{b}_n = [b_n(1), \cdots, b_n(M)]^T \in \mathscr{B}^{M \times 1}$ denote a block of M information bits to be encoded at the n-th source. In the proposed DJSCC scheme, M is chosen to be an integer multiple of the number of users N as $M = KN$ with K being an integer. The corresponding DJSCC codeword of the n-th source can then be represented as

$$\mathbf{c}_n^T = \mathbf{b}_n^T \left[\mathbf{T}_n, \mathbf{P}_n\right] = \left[\mathbf{s}_n^T, \mathbf{p}_n^T\right]^T \tag{4.2}$$

where $\mathbf{T}_n \in \mathscr{B}^{M \times K}$ is the information compression matrix with $K = \frac{M}{N}$, $\mathbf{P}_n \in \mathscr{B}^{M \times P}$ is the parity generation matrix, $\mathbf{s}_n = \mathbf{T}_n^T \mathbf{b}_n \in \mathscr{B}^{K \times 1}$ is the compressed information vector, $\mathbf{p}_n = \mathbf{P}_n^T \mathbf{b}_n \in \mathscr{B}^{P \times 1}$ is the parity vector, and the matrix operations in (4.2) are performed in the Galois field of two elements, GF(2). The parity generation matrix $\mathbf{P}_n$ will generate P parity bits from M information bits. The code rate of the DJSCC code is thus $r = \frac{M}{K+P}$.

The information compression matrix $\mathbf{T}_n$ is obtained by removing $M - K = K(N-1)$ columns of a size-M identity matrix $\mathbf{I}_M$. Denote $\mathscr{T}_n = \{n_1, \cdots, n_K\} \subseteq \{1, 2, \cdots, M\}$ as the set of the K indices corresponding to the columns *not* removed from $\mathbf{I}_M$ during the construction of $\mathbf{T}_n$, then $\mathbf{T}_n = [\mathbf{i}_{n_1}, \cdots, \mathbf{i}_{n_K}]$ with $\mathbf{i}_m$ being the m-th column of $\mathbf{I}_M$. In the proposed DJSCC, we have $\mathscr{T}_n \cap \mathscr{T}_m = \emptyset$, and $\bigcup_{n=1}^N \mathscr{T}_n = \{1, 2, \cdots, M\}$. When $N = 1$, we have $\mathbf{T}_n = \mathbf{I}_M$, and the DJSCC codeword in (4.2) degrades to a regular systematic linear block code of code rate $\frac{M}{M+P}$.

When $N > 1$, the codeword structure in (4.2) combines distributed source code and channel code in a unified structure. The channel code is performed with the combination of the punctured information and the parity vectors. The distributed source code is performed with the information compression matrix $\mathbf{T}_n$, which punctures the length-M information vector $\mathbf{b}_n$ into a length-K vector $\mathbf{s}_n$, with mutually exclusive puncture patterns defined by the index set $\{\mathscr{T}_n\}_{n=1}^N$. The information puncture operation deliberately adds distortion to the information to reduce the amount of information to be transmitted, thus reduce the overall energy requirement. With mutually exclusive puncture patterns, if $b_n(k)$ is punctured, it is guaranteed that there exists $m \neq n$ such that $b_m(k)$ on source m is transmitted. Then $b_n(k)$ can be partly recovered by using the correlation between $b_m(k)$ and $b_n(k)$, as well as the parity vector $\mathbf{p}_n(k)$.

In summary, for a system employing the DJSCC, the information distortion comes from two sources, the distortion deliberately added by the information puncture operation, and the channel distortion. At the decoder, the distortions are compensated from two aspects, the spatial correlation, and the parity vector. The information correlation is utilized to both reduce the energy consumption and to protect the information from channel distortion. Therefore, the distributed source code and channel code are jointly performed in a single step.

In the proposed DJSCC scheme, only the information vector is punctured and the parity vector is transmitted in its entirety. This is because the punctured information can be partly compensated by the information correlation, yet the correlation among the parity vectors from different sources are relatively low even if the information is strongly correlated. The correlation between two parity bits from different sources is stated in the following lemma.

Lemma 4.1: If the users use different parity generation matrix, then the probability that two

parity bits, p_{nk} and p_{mk} are different is $P(p_{nk} \neq p_{mk}) = 0.5$. If all users share the same parity generation matrix, and the Hamming weight of the k-th column of the parity generation matrix is L, then

$$P(p_{nk} \neq p_{mk}) = \sum_{u=1,u \text{ odd}}^{L} \binom{L}{u} p_{mn}^{u} (1 - p_{mn})^{L-u} \tag{4.3}$$

Proof: If the users use different parity generation matrix, then p_{nk} and p_{mk} are mutually independent because the information bits are assumed to be independent in the time domain. If all the users use the same parity generation matrix, the probability $P(p_{nk} \neq p_{mk})$ is equal to the probability that, for the L bits corresponding to the non-zero positions of the k-th column of $\mathbf{P}_n$, there are an odd number of bits from the m-th user that are not equal to their counterparts from the n-th user. The probability that u bits are not equal to each other follows a binomial distribution as $\binom{L}{u} p_{mn}^{u} (1 - p_{mn})^{L-u}$. The result in (4.3) follows immediately. ∎

As L becomes large, the binomial distribution $\binom{L}{u} p_{mn}^{u} (1 - p_{mn})^{L-u}$ can be accurately approximated by a normal distribution with mean Lp_{mn} and variance $Lp_{mn}(1 - p_{mn})$. Since the normal distribution is symmetric with respect to Lp_{mn}, the summation in (4.3) tends to $\frac{1}{2} \sum_{u=1}^{L} \binom{L}{u} p_{mn}^{u} (1 - p_{mn})^{L-u} = 0.5$ when L is large.

This is corroborated by the result in Fig. 4.1, which shows the cross probability of the parity bits, $P(p_{nk} \neq p_{mk})$, as a function of L with various values of the cross probability p_{mn}. The cross probability of the parity bits is less than the cross probability of the corresponding information bits under all configurations, and they tend to 0.5 as L becomes large.

Based on the above analysis, the parity bits across different users usually have very weak correlations, even if the information bits are strongly correlated. Therefore, puncturing the parity bits will have very little contribution to the distributed source code, yet it will sacrifice the performance
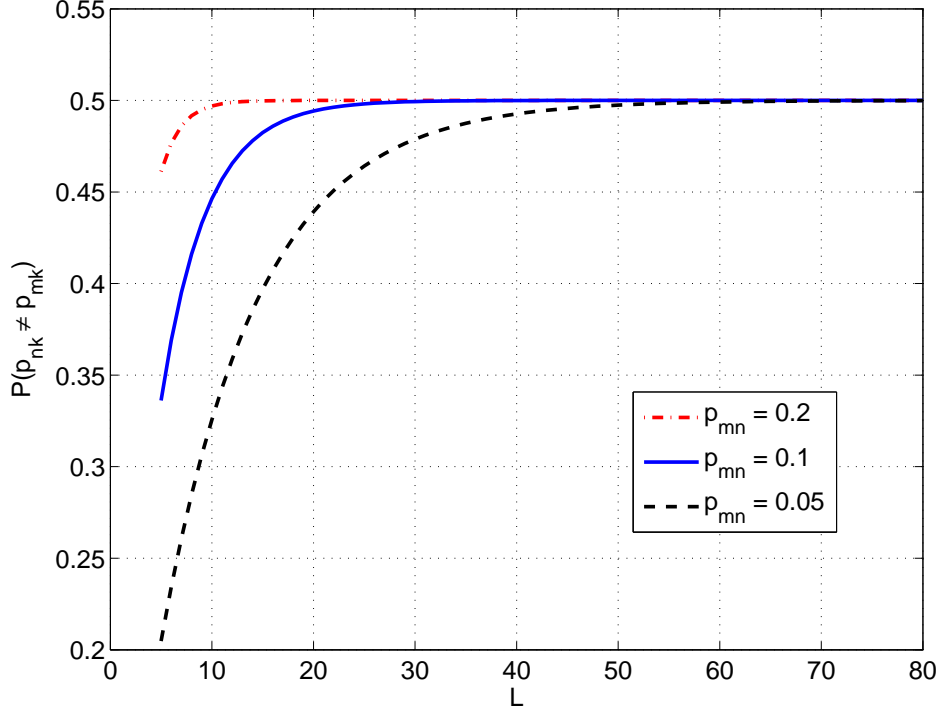
Figure 4.1: Cross probability of the parity bits.

of the channel code. Therefore, we propose to puncture only the information bits, and transmit the parity bits at its entirety.

4.3.2 Transmission with Unequal Energy Allocation

Since the information bits and the parity bits are treated differently during the encoding process, we propose to allocate different amounts of energy per bit to the information and parity bits during transmissions.

Denote $\alpha_n = \text{mod}[\mathbf{s}_n]$ and $\beta_n = \text{mod}[\mathbf{p}_n]$, where $\text{mod}[\mathbf{b}] \in \mathscr{S}$ maps the binary vector $\mathbf{b}$ to the modulation constellation set $\mathscr{S}$. The codeword after modulation and energy allocation is

$$\mathbf{x}_n = [\sqrt{E_s}\alpha_n^T, \sqrt{E_p}\beta_n^T]^T, \tag{4.4}$$

where E_s is the energy per punctured information symbol, and E_p is the energy per parity symbol. The average energy per information bit is thus $E_b = \frac{E_s K + E_p P}{M \log_2(S)}$, where $S = |\mathscr{S}|$ is the cardinality of

92

the constellation set $\mathscr{S}$.

Define the energy allocation factor as $\theta = \frac{E_p}{E_s}$. Intuitively, more energy per symbol should be allocated for the punctured information bits to compensate for the punctured bits. This intuition is supported by our simulation results. The energy allocation factor, θ, is used to adjust the energy allocation between the punctured information bits and the parity bits. When $\theta = 0$, no parity bits will be transmitted and the scheme degrades to a punctured transmission scheme without channel code.

The unequal energy allocation between the punctured information and parity symbols is motivated by the fact that the sink can have the entire parity vectors from all the users, but only a punctured version of the original information vector. The punctured information is recovered by using a combination of the parity bits and the information correlation. Therefore, the transmitted information bits from one user are used to recover the punctured bits from the other users. Therefore, more energy can be allocated to the unpunctured bits to compensate the extra distortions introduced by the puncturing operations.

The modulated codeword, $\mathbf{x}_n$, is transmitted to the sink through an orthogonal media access control (MAC) scheme. The signal received from the n-th user is

$$\mathbf{y}_n = \mathbf{x}_n + \mathbf{z}_n, \tag{4.5}$$

where $\mathbf{z}_n$ is the additive white Gaussian noise (AWGN) with a single-sided power spectral density N_0. The received signal vector at the sink can be expressed as $\mathbf{y}_n = [\mathbf{y}_{\alpha n}^T, \mathbf{y}_{\beta n}^T]^T$, where $\mathbf{y}_{\alpha n}$, $\mathbf{y}_{\beta n}$ are the received signal vectors corresponding to the coded sequence $\boldsymbol{\alpha}_n$ and $\boldsymbol{\beta}_n$, respectively.

4.4 DJSCC Decoding with the Message Passing Algorithm

The sink recovers the information vector by performing joint decoding with the message passing algorithm based on the received signals from all the N sources.

Before the decoding, we need to calculate the initial log-likelihood ratio (LLR) from the channel observations. The initial LLR of the k-th coded information bit $s_n(k) = b_n(n_k)$, from the channel observation, $\mathbf{y}_{\alpha n}$, can be calculated by

$$\lambda_n(n_k) = \log \frac{\sum_{s \in \mathscr{S}_{r_k}^-} \exp\left[-\frac{1}{N_0} \left|y_{\alpha n}(m_k) - \sqrt{E_s}s\right|^2\right]}{\sum_{s \in \mathscr{S}_{r_k}^+} \exp\left[-\frac{1}{N_0} \left|y_{\alpha n}(m_k) - \sqrt{E_s}s\right|^2\right]}, \tag{4.6}$$

where the k-th bit in a coded vector is mapped to the $m_k = \lfloor \frac{k}{\log_2 S} \rfloor$ modulated symbol, $r_k = k - m_k \log_2 S$, $\mathscr{S}_{r_k}^+ \subset \mathscr{S}$ is the set that contains all the symbols with the r_k-th bit in the demodulated vector being 1, and $\mathscr{S}_{r_k}^- = \mathscr{S} \backslash \mathscr{S}_{r_k}^+$.

Similarly, the initial LLR of the k-th parity bit, $p_n(k)$, can be calculated from $y_{\beta n}$ as

$$\lambda_n(M+k) = \log \frac{\sum_{s \in \mathscr{S}_{r_k}^-} \exp\left[-\frac{1}{N_0} \left|y_{\beta n}(m_k) - \sqrt{E_p}s\right|^2\right]}{\sum_{s \in \mathscr{S}_{r_k}^+} \exp\left[-\frac{1}{N_0} \left|y_{\beta n}(m_k) - \sqrt{E_p}s\right|^2\right]}. \tag{4.7}$$

The LLRs of the punctured bits of one user can be calculated from their correlated counterparts transmitted by a different user. Assume $k \in \mathscr{T}_n$, $i.e.$, the bit b_{nk} is transmitted by the n-th user, and b_{mk} is punctured at the m-th user, $\forall m \neq n$. Then the LLRs of the punctured bits can be calculated as

$$\hat{\lambda}_m(k) = \log \frac{(1 - p_{mn})P(b_n(k) = 1) + p_{mn}P(b_n(k) = 0)}{p_{mn}P(b_n(k) = 1) + (1 - p_{mn})P(b_n(k) = 0)}, \tag{4.8}$$

where $P(b_n(k) = 1) = \frac{1}{1 + \exp[\lambda_n(k)]}$ and $P(b_n(k) = 0) = 1 - P(b_n(k) = 1)$.

The initial LLR of the information and parity bits can then be expressed by

$$\delta_n(k) = \begin{cases} \lambda_n(k), & k \in \mathscr{T}_n \text{ or } M < k \leq M + P, \\ \hat{\lambda}_n(k), & \text{otherwise,} \end{cases} \tag{4.9}$$

The message passing will be performed on a bipartite graph defined by an extended parity check matrix $\mathbf{H} = [\mathbf{P}_n^T, \mathbf{I}_P]^T \in \mathscr{B}^{(M+P) \times P}$. It should be noted that $\mathbf{H}$ is not the parity check matrix for the codeword defined in (4.2). It is the parity check matrix of the unpunctured codeword with the generation matrix $[\mathbf{I}_M, \mathbf{P}_n]$.

The tanner graph corresponding to $\mathbf{H}$ has $(M+P)$ variable nodes and P check nodes. The k-th variable node is connected to the p-th check node if the (k,p)-th element of $\mathbf{H}$ is 1. Denote $\mathscr{V}_p$ as the set of variable nodes that are connected to the p-th check node, and $\mathscr{C}_k$ as the set of check nodes that are connected to the k-th variable nodes.

For the information from the n-th user, the message from the k-th variable node to the p-th check node during the i-th iteration is

$$\eta_{kp}^{(i)}(n) = \delta_n(k) + \sum_{p' \in \mathscr{C}_k \backslash p} \mu_{p'k}^{(i-1)}(n), \tag{4.10}$$

where $\mu_{pk}^{(i)}(n)$ is the message from the p-th check node to the k-th variable node during the i-th iteration,and it can be calculated as

$$\mu_{pk}^{(i)}(n) = 2\mathrm{atanh} \prod_{k' \in \mathscr{V}_p \backslash k} \tanh \frac{\eta_{k'p}^{(i)}(n)}{2} \tag{4.11}$$

During the first iteration, $\mu_{pk}^{(0)}(n) = 0$.

For the message passing algorithm [12], eqns. (4.10) and (4.11) are performed iteratively for a single codeword, and the values of $\delta_n(k)$ are the same for all the iterations. At the i-th iteration, the n-th user will output a soft decision for its information bits, as

$$l_n^{(i)}(k) = \delta_n(k) + \sum_{p' \in \mathscr{C}_k} \mu_{p'k}^{(i)}(n). \tag{4.12}$$

At the final iteration, the hard decision with I iterations is obtained as $\hat{b}_n(k) = 1$ if $l_n^{(I)}(k) > 0$ and $\hat{b}_n(k) = 0$ otherwise.

The iterative message passing algorithm is summarized as follows.

I) **Initialization**

    i) Calculate the initial LLRs of the transmitted information bits, $\lambda_n(k)$, with (4.6), for

    $k \in \mathcal{T}_n$, $n = 1, \cdots, N$.

    ii) Calculate the initial LLRs of the parity bits, $\lambda_n(M + k)$, with (4.7), for $k = 1, \cdots, P$,

    $n = 1, \cdots, N$.

    iii) Calculate the LLRs of the punctured information bits, $\hat{\lambda}_m(k)$, with (4.8).

    iv) Set $i = 1$.

II) **Iterations**

    i) Calculate the message from the variable node to the check node, $\eta_{kp}^{(i)}(n)$, with (4.10),

    for $k = 1, \cdots, N$, $p = 1, \cdots, P$, and $n = 1, \cdots, N$.

    ii) Calculate the message from the check node to the variable node, $\mu_{pk}^{(i)}(n)$, with (4.11),

    for $k = 1, \cdots, N$, $p = 1, \cdots, P$, and $n = 1, \cdots, N$.

    iii) Calculate the soft decisions, $l_n^{(i)}(k)$, with (4.12).

    iv) If $i < I$, go back to step II.i); otherwise go to the next step.

III) **Detection**

Make hard decision as $\hat{b}_n(k) = 1$ if $l_n^{(I)}(k) > 0$ and $\hat{b}_n(k) = 0$ otherwise.

4.5   Simulation Results

Simulation results are presented in this section to demonstrate the performance of the proposed

DJSCC scheme. In the simulation, an irregular LDPC with a 32400-by-64800 parity-check matrix
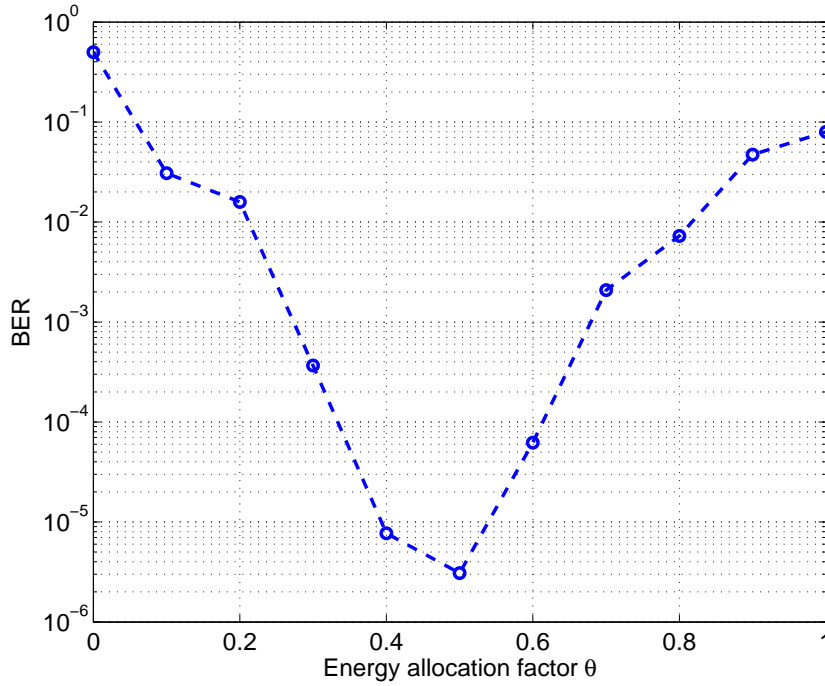
Figure 4.2: BER as a function of the energy allocation factor θ (E_b/N_0 of the source-sink link is -0.65 dB).

is used to generate the DJSCC codeword. In the simulation, the pairwise covariance coefficient between any pair of users is the same, *i.e.*, $\rho_{mn} = \rho$, or $p = p_{mn}$, $\forall m \neq n$. The mutually correlated information of the N users is generated by passing the a length-M binary vector through N independent and identically distributed (i.i.d.) binary symmetric channel (BSC) channels. The cross probability of the BSC channel is $p_0 = \frac{1}{2} - \frac{1}{2}\sqrt{1 - 2p}$. It can be easily shown that the cross probability between the output of any pair of BSC channels is p. It should be noted that the proposed DJSCC scheme can be applied to sources with arbitrary correlations.

We first study the impact of the unequal energy allocation on the performance of the DJSCC scheme in Fig. 4.2, where the bit error rate (BER) is shown as a function of the energy allocation factor, θ, for a network with two users. The covariance coefficient between the two users is $\rho = 0.9$. The source-sink communication links are AWGN channels with $E_b/N_0 = -0.65$ dB. It can be seen
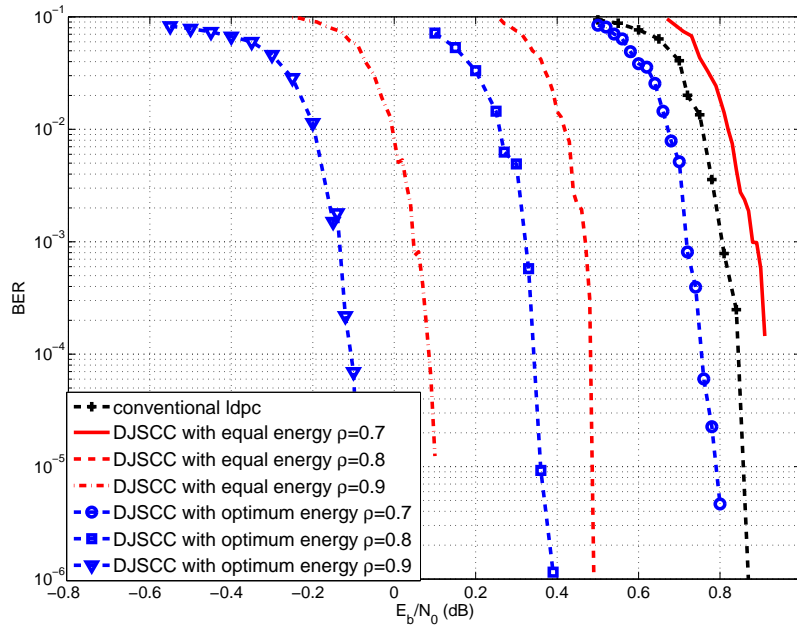
Figure 4.3: The performance of equal and optimum energy for two users

that the optimum performance is achieved at $\theta = 0.5$, e.g., the energy of one parity bit is half of that of one transmitted information bit. The performance degrades significantly with equal energy allocation at $\theta = 1$.

Fig. 4.3 shows the BER performance of the proposed DJSCC scheme under various values of ρ. There are two users in the system. The energy allocation factor is $\theta = 0.6$. The performances of systems with equal energy allocation are also shown for comparison. The curve labeled as conventional LDPC is obtained without DJSCC, thus its performance is independent of the number of users. As expected, the performance improves as ρ increases. With unequal energy allocation and at a BER $= 10^{-4}$, the DJSCC obtains 0.1 dB, 0.5 dB, and 1 dB performance gains over conventional LDPC coded system at $\rho = 0.7, 0.8$, and 0.9, respectively. The DJSCC with unequal energy allocation outperform their equal energy counterparts by about 0.18 dB.

The impact of the number of users on the performance is shown in Fig. 4.4. The covariance
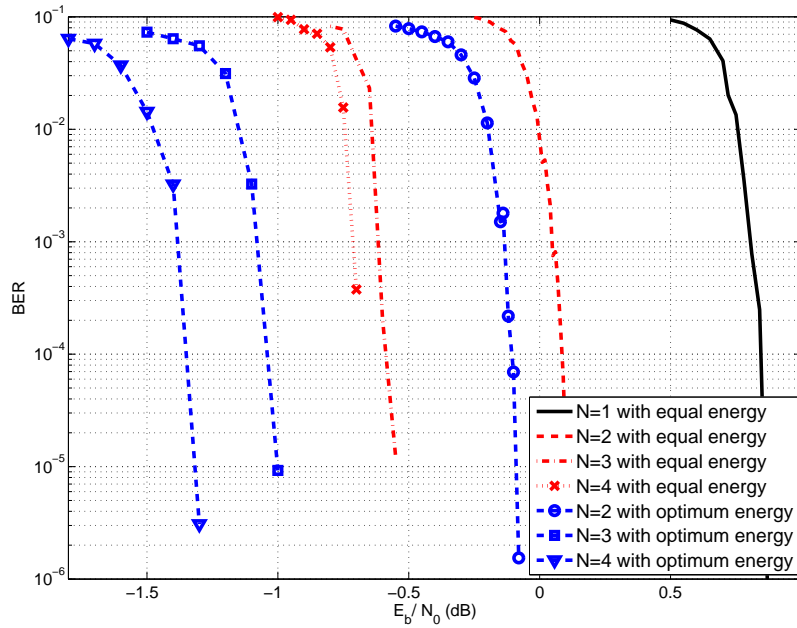
Figure 4.4: The BER under varies number of users N (covariance coefficient $\rho = 0.9$)

coefficient is $\rho = 0.9$. No puncture operation is employed in the $N = 1$ case and it is the same as

the conventional LDPC code. The proposed DJSCC scheme benefits from the presence of more

users, due to the better compression ratio of the distributed source code. The DJSCC systems with

$N = 2$, 3, and 4 outperform the conventional LDPC coded system by 1 dB, 1.9 dB, and 2.3 dB,

respectively.

4.6 Conclusion

A new DJSCC scheme based on linear block code for a communication network with correlated

information sources and operating over noisy channels was proposed in this chapter. It was demon-

strated that the correlation among the parity bits of a linear block code was usually very low even

when the correlation among the information bits is high. Therefore, the DJSCC was performed by

puncturing the information bits but transmitting the parity bits in its entirety. The information and

parity bits were transmitted with unequal energy per bit to achieve additional performance gain. The message passing algorithm was used at the receiver to jointly recover the information from all the sources. Simulation results demonstrated that the proposed DJSCC scheme with unequal energy allocation can achieve significant performance gains over conventional schemes, and the performance improves as the number of users increases.

4.7 Appendix of documentation of multi-authored chapter

UNIVERSITY OF ARKANSAS.

College of Engineering
Department of Electrical Engineering

July 31, 2013

To Whom It May Concern

This letter is to certify that Ms. Ning Sun, a Ph.D. candidate under my supervision at the Department of Electrical Engineering, has contributed more than 51% of the work for the following paper,

Ning Sun, Jingxian Wu, and Hai Lin, "Distributed joint source and channel code with correlated information sources," in Proceedings of IEEE International Conference on Communications in China (ICCC'12), August 2012.

Ms. Sun is the first author of this paper. The paper is included in her Ph.D. dissertation entitled "Distortion-Tolerant Communications with Correlated Information".

Sincerely Yours,

Jingxian Wu
Director, Wireless Information Network Lab
Associate Professor of Electrical Engineering
University of Arkansas
Email: wuj@uark.edu
Tel: (479) 575-6584

3217 Bell Engineering Center • Fayetteville, AR 72701 • 479-575-3005 • Fax: 479-575-7967 • www.uark.edu
The University of Arkansas is an equal opportunity/affirmative action institution.

4.8   References

[1] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Info. Theory*, vol. 19, pp. 471-480, July 1973.

[2] S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): Design and construction," *IEEE Trans. Info. Theory*, pp. 626-643, Mar. 2003.

[3] J. Bajcsy and P. Mitran, "Coding for the SlepianWolf problem with turbo codes," in *Proc. IEEE Globecom'01*, vol. 2, pp. 1400-1404, Nov. 2001.

[4] A. D. Liveris, Z. Xiong, and C. N. Georghiades, "Compression of binary sources with side information at the decoder using LDPC codes," *IEEE Commun. Lett.*, vol. 6, pp. 440-442, Oct. 2002.

[5] J. Garcia-Frias, "Compression of correlated binary sources using turbo codes," *IEEE Commun. Lett.*, vol. 5, pp. 417-419, Oct. 2001.

[6] M. Sartipi and F. Fekri, "Distributed source coding using short to moderate length rate-compatible LDPC codes: The entire Slepian-Wolf rate region," *IEEE Trans. Commun.*, vol. 56, pp. 400-411, 2008.

[7] I. Shahid and P. Yahampath, "Distributed Joint source-channel coding of correlated binary sources in wireless sensor networks," in *Proc. IEEE Intern. Symp. Wireless Commun. Syst.*, pp. 236-240, 2011.

[8] N. Gehrig and P. L. Dragotti, "Symmetric and a-symmetric Slepian-Wolf codes with systematic and non-systematic linear codes," *IEEE Commun. Lett.*, Vol. 9, pp.61 - 63, Jan.,2005.

[9] R. Rajesh, V. K. Varshneya, and V. Sharma, "Distributed joint source channel coding on a multiple access channel with side information," *Proc. IEEE ISIT*, pp. 2707-2711, July 2008.

[10] Q. Xu, V. Stankovic, and Z. Xiong, "Distributed joint source-channel coding of video using raptor codes," *IEEE J. Sel. Areas Commun.*, vol. 25, pp. 851-861, 2007.

[11] X. Zhu, L. Zhang, and Y. Liu, "A distributed joint source-channel coding scheme for multiple correlated sources," in *Proc. IEEE Commun. and Networking in China*, pp. 26-28, Aug. 2009

[12] A. Goldsmith, *Wireless Communications*, 2nd Ed., Cambridge University Press, 2005.

Distributed Joint Source-Channel Code for Spatially Correlated Markov Sources

Ning Sun , Jingxian Wu, and Guoqing Zhou

5.1 Abstract

In this chapter, a new distributed joint source-channel code (DJSCC) is proposed for a communication network with spatially correlated Markov sources. The DJSCC is performed by puncturing the information bits of a systematic linear block code but leaving the parity bits intact. Due to the different roles of the information and parity bits in the DJSCC scheme, unequal energy is allocated to these two different types of bits during transmission. At the receiver, the spatial data correlation is exploited with a new multi-codeword message passing (MCMP) decoding algorithm. The MCMP decoder performs both intra- and inter-codeword soft information exchange, whereas conventional message passing (MP) algorithms exchanges soft information only inside a codeword. The inter-code soft information exchange of MCMP leads to additional performance gain over the MP algorithm. In recognition that the signals at the receiver are distorted observations of the Markov source and thus can be modeled by a hidden Markov model (HMM), we propose to add a HMM decoding module to the MCMP decoder to exploit the temporal data correlation. The HMM decoder iteratively exchanges soft information with the MCMP decoder, and this leads to significant performance gains over conventional systems.

5.2 Introduction

In a wireless sensor network (WSN), data collected by the spatially distributed sensors are often correlated in both the space and time domains due to the inherent space redundancy and time variation of the monitored entities or phenomena. The spatial-temporal data correlation can be exploited to significantly improve the performance of the wireless network. The Slepian-Wolf (S-W) theorem [1] states that distributed sources with correlated information can perform encoding separately, yet achieve a code rate that is the same as when the information is encoded jointly. However, the S-W theorem is not constructive, i.e., it provides no practical coding scheme to achieve the optimum performance.

There have been considerable works in the literature devoted to the design of practical distributed source codes (DSC) [2]–[6]. Many practical DSC schemes are designed by using the syndromes of channel codes [2], [3]. Many syndrome-based DSC designs focus on the asymmetric scenario, *i.e.*, the distributed coding is only applied to one of the sources, and the other sources are used as side-information and assumed to be known perfectly at the decoder. Designs of symmetric DSCs are discussed in [4]–[7], with punctured linear block codes. However, All of the above work focus only on the spatial correlation among the sensors, and they do not consider the variation of the data in the time domain. In reality, the physical phenomenon under monitoring changes with respect to time, and the consecutive observations of a sensor node are often correlated in the time domain.

The temporal data correlation can be utilized during the detection process at the receiver to improve the system performance [8]–[11]. In [8], a joint turbo decoding and estimation scheme is proposed for the estimation of a temporally correlated source modeled by a hidden Markov

model (HMM). It treats the trellis describing the HMM as a constituent code, which exchanges soft information with other constituent codes during turbo detection. The turbo code in [8] is replaced by a low-density parity-check (LDPC) code in [9], where the random nature of the LDPC code eliminates the need of an interleaver between the HMM trellis and channel decoder, thus it reduces the decode delay. In [10], the HMM decoding is performed by modeling the Markov source to a bipartite graph. All of the above works [8]–[10] consider only decoding algorithm at the receiver. None of them considers spatial source correlation or source code at the transmitter. An asymmetric distributed source code for HMM source is presented in [11] by using a punctured Raptor code, but the correlated source is used as a perfectly known side information at the decoder.

In this chapter, we propose a new symmetric distributed joint source and channel code (DJSCC) for correlated Markov sources. The data at the receiver are distorted observations of the Markov sources, and they can be modeled as HMM. At the transmitter, each source performs the DJSCC encoding by puncturing the information bits of the codewords of a systematic channel code. The punctured bits can be partly recovered by using the spatial-temporal data correlation and the parity bits of the channel code. In addition, the information and parity bits are transmitted over a noisy channel with different energy to achieve additional performance gain. Two new decoders, one for spatially correlated memoryless sources, and one for sources correlated in both the space and time domains, are proposed. The first decoder employs a new iterative multi-codeword message passing (MCMP) algorithm. The iterative MCMP algorithm performs decoding on multiple codewords simultaneously, and soft information are exchanged between the codewords throughout the iterations. This is different from the conventional message passing (MP) algorithm, where the soft information is only updated inside one codeword. The second decoder exploits the temporal data correlation by adding a HMM decoding module to the MCMP decoder. The HMM decoding

106

module extracts the *a postorior* probability of the data by utilizing the hidden Markovian property of the data at the receiver, and it iteratively exchanges soft information with the MCMP decoder to improve the decoding performance. Simulation results demonstrate that both of two new decoders outperform the conventional MP decoder, and significant performance gain is achieved by the newly proposed DJSCC scheme.

5.3 Distributed Joint Source-Channel Code

5.3.1 System Model

Consider a system with two correlated binary sources, $b_n(k) \in \mathscr{B}$, for $n = 1$ and 2, transmitting to one information sink, where k is the time index and $\mathscr{B} = \{0, 1\}$. The correlation between the two sources can be modeled as $b_2(k) = b_1(k) \oplus e(k)$, where $e(k) \in \mathscr{B}$ is an identically and independently distributed (i.i.d.) random process with $P(e(k) = 1) = p_0$. The cross probability between the two sources is $P(b_1(k) \neq b_2(k)) = p_0$. Define the correlation coefficient between $b_1(k)$ and $b_2(k)$ as

$$\rho_{mn} = \mathbb{E}\left[\frac{(b_m - \mu_m)(b_n - \mu_n)}{\sigma_m \sigma_n}\right] = 1 - 2p_0, \tag{5.1}$$

where μ_m and σ_m are the mean and standard deviation of $b_m(k)$, respectively.

The information of each source is correlated in the time domain. Without loss of generality, it is assumed that the time correlation of source 1 can be modeled as a first-order Markov chain, with the following parameters

- States: $S_0 = 0$ and $S_1 = 1$.

- Transition probability: $a_{ij} = P(S_k = j | S_{k-1} = i), i, j \in \{0, 1\}$.

- Initial state distribution: $P(S_0)$ and $P(S_1)$.

The information from source 2 is a distorted observation of the Markov process from source 1, thus it can be modeled as a HMM source. The data from the two sources will be separately encoded, modulated, and transmitted to the information sink through a noisy channel.

5.3.2 Codeword Structure

Each source encodes its own information *without* the knowledge of the information from the other users. The proposed DJSCC is a linear block code. Let $\mathbf{b}_n = [b_n(1), \cdots, b_n(M)]^T \in \mathscr{B}^{M \times 1}$ denote a block of M information bits to be encoded at the n-th source user. In the proposed DJSCC scheme, M is chosen to be an integer multiple of the number of users as $M = 2K$ with K being an integer. The corresponding DJSCC codeword of the n-th source can then be represented as

$$\mathbf{c}_n^T = \mathbf{b}_n^T [\mathbf{T}_n, \mathbf{P}_n] = \left[\mathbf{s}_n^T, \mathbf{p}_n^T \right]^T \tag{5.2}$$

where $\mathbf{T}_n \in \mathscr{B}^{M \times K}$ is the information compression matrix with $K = \frac{M}{2}$, $\mathbf{P}_n \in \mathscr{B}^{M \times P}$ is the parity generation matrix, $\mathbf{s}_n = \mathbf{T}_n^T \mathbf{b}_n \in \mathscr{B}^{K \times 1}$ is the compressed information vector, $\mathbf{p}_n = \mathbf{T}_n^T \mathbf{b}_n \in \mathscr{B}^{P \times 1}$ is the parity vector, and the matrix operations in (5.2) are performed in the Galois field of two elements, GF(2). The parity generation matrix $\mathbf{P}_n$ will generate P parity bits from M information bits. The code rate of the DJSCC code is thus $r = \frac{2M}{M+2P}$.

The information compression matrix $\mathbf{T}_n$ is obtained by removing $K = \frac{M}{2}$ columns of a size-M identity matrix $\mathbf{I}_M$. Denote $\mathscr{T}_n \subset \{1, 2, \cdots, M\}$ as the set of the K indices corresponding to the columns *not* removed from $\mathbf{I}_M$ during the construction of $\mathbf{T}_n$, then $\mathbf{T}_n = [\mathbf{i}_{n_1}, \cdots, \mathbf{i}_{n_K}]$ with $n_k \in \mathscr{T}_n$, where $\mathbf{i}_m$ is the m-th column of $\mathbf{I}_M$. In the proposed DJSCC, we have $\mathscr{T}_1 \cap \mathscr{T}_2 = \emptyset$, and $\mathscr{T}_1 \cup \mathscr{T}_2 = \{1, 2, \cdots, M\}$.

The distributed source code is performed with the information compression matrix $\mathbf{T}_n$, which punctures the length-M information vector $\mathbf{b}_n$ into a length-K vector $\mathbf{s}_n$, with mutually exclusive puncture patterns defined by the index sets $\{\mathscr{T}_n\}_{n=1}^2$. With mutually exclusive puncture patterns, if $b_n(k)$ is punctured, then $b_m(k)$ on source $m \neq n$ is transmitted. Then $b_n(k)$ can be partly recovered by using the correlation between $b_m(k)$ and $b_n(k)$, as well as the parity vector $\mathbf{p}_n(k)$.

Based on [7, Lemma 1], the parity bits across different users usually have very weak correlations, even if the information bits are strongly correlated. Therefore, puncturing the parity bits will have very little contribution to the distributed source code, yet it will sacrifice the performance of the channel code. Therefore, we propose to puncture only the information bits, and transmit the parity bits at its entirety.

5.3.3 Transmission with Unequal Energy Allocation

Since the information bits and the parity bits are treated differently during the encoding process, we propose to allocate different amounts of energy per bit to the information and parity bits during transmission.

Denote the modulated symbols as $\boldsymbol{\alpha}_n=\text{mod}\,[\mathbf{s}_n]$ and $\boldsymbol{\beta}_n=\text{mod}\,[\mathbf{p}_n]$, respectively, where $\text{mod}\,[\mathbf{b}] \in \mathscr{S}$ maps the binary vector $\mathbf{b}$ to the modulation constellation set $\mathscr{S}$. The codeword after modulation and energy allocation is

$$\mathbf{x}_n = [\sqrt{E_s}\boldsymbol{\alpha}_n^T, \sqrt{E_p}\boldsymbol{\beta}_n^T]^T, \tag{5.3}$$

where E_s is the energy per information symbol, and E_p is the energy per parity symbol. The average energy per information bit is thus $E_b = \frac{E_s K + E_p P}{M \log_2(S)}$, where $S = |\mathscr{S}|$ is the cardinality of the constellation set $\mathscr{S}$.

Define the energy allocation factor as $\theta = \frac{E_p}{E_s}$. Intuitively, more energy per symbol should be allocated for the punctured information bits to compensate for the punctured bits. This intuition is supported by our simulation results. The energy allocation factor, θ, is used to adjust the energy allocation between the punctured information bits and the parity bits. When $\theta = 0$, no parity bits will be transmitted and the scheme degrades to a punctured transmission scheme without channel code.

The modulated codeword, $\mathbf{x}_n$, is transmitted to the sink user through an orthogonal media access control (MAC) scheme. The signal received from the n-th user is

$$\mathbf{y}_n = \mathbf{x}_n + \mathbf{z}_n, \text{ for } n = 1, 2, \tag{5.4}$$

where $\mathbf{z}_n$ is the additive white Gaussian noise (AWGN) with a single-sided power spectral density N_0. The received signal vector at the sink can be expressed as $\mathbf{y}_n = [\mathbf{y}_{\alpha n}^T, \mathbf{y}_{\beta n}^T]^T$, where $\mathbf{y}_{\alpha n}, \mathbf{y}_{\beta n}$ are the received signal vectors corresponding to the coded sequence $\boldsymbol{\alpha}_n$ and $\boldsymbol{\beta}_n$, respectively.

5.4 DJSCC Decoding with a Multi-Codeword Message Passing Algorithm

The information sink recovers the information vector by performing joint decoding based on the received signals from the sources. We propose two new joint decoding algorithms, which can be utilized for correlated memoryless sources and correlated Markovian sources, respectively. The first algorithm performs decoding by iteratively exchanging information between the two codewords without utilizing the time correlation, and we denote it as the multi-codeword message passing algorithm (MCMP). The second algorithm extends the proposed MCMP by exploiting the Markov property of the information sources in each iteration.

5.4.1 Joint Decoding for Correlated Memoryless Sources

We present a MCMP decoding algorithm in this subsection, which can effectively improve the performance of correlated memoryless sources.

Different from the conventional message passing (MP) algorithm where the LLR is exchanged within the structure of one codeword, the proposed MCMP algorithm will iteratively exchange information between the codewords from the two sources.

Before decoding, we need to calculate the initial log-likelihood ratio (LLR) from the channel observations. Assume the k-th bit is mapped to the m_k-th modulated symbol, and it corresponds to the r_k-th bit in the demodulated vector of the modulated symbol. If the k-th information bit is not punctured during the encoding, then the initial LLR of $b_n(k)$ from the channel observation, $\mathbf{y}_{\alpha n}$, can be calculated by

$$\lambda_n^{(0)}(k) = \log \frac{\sum_{s \in \mathscr{S}_{r_k}^-} \exp\left(-\frac{1}{N_0}\left|y_{\alpha n}(m_k) - \sqrt{E_s}s\right|^2\right)}{\sum_{s \in \mathscr{S}_{r_k}^+} \exp\left(-\frac{1}{N_0}\left|y_{\alpha n}(m_k) - \sqrt{E_s}s\right|^2\right)}, \tag{5.5}$$

where $\mathscr{S}_{r_k}^+ \subset \mathscr{S}$ is the set that contains all the symbols with the r_k-th bit in the demodulated vector being 1, and $\mathscr{S}_{r_k}^- = \mathscr{S} \backslash \mathscr{S}_{r_k}^+$.

Similarly, the initial LLR of the k-th parity bit, $p_n(k)$, can be calculated from $y_{\beta n}$ as

$$\lambda_n^{(0)}(M+k) = \log \frac{\sum_{s \in \mathscr{S}_{r_k}^-} \exp\left(-\frac{1}{N_0}\left|y_{\beta n}(m_k) - \sqrt{E_p}s\right|^2\right)}{\sum_{s \in \mathscr{S}_{r_k}^+} \exp\left(-\frac{1}{N_0}\left|y_{\beta n}(m_k) - \sqrt{E_p}s\right|^2\right)}. \tag{5.6}$$

The LLRs of the punctured bits of one user can be calculated from their correlated counterparts transmitted by the other user. Assume $k \in \mathscr{T}_n$, *i.e.*, the bit b_{nk} is transmitted by the n-th user, and b_{mk} is punctured at the m-th user, for $m \neq n$. Then the LLRs of the punctured bits at the i-th iteration can be calculated as

$$\hat{\lambda}_m^{(i)}(k) = \log \frac{(1-p_0)P_{n0}^{(i)}(k) + p_0[1 - P_{n0}^{(i)}(k)]}{p_0 P_{n0}^{(i)}(k) + (1-p_0)[1 - P_{n0}^{(i)}(k)]}, \tag{5.7}$$

where $P_{n0}^{(i)}(k) = \frac{\exp[\lambda_n^{(i)}(k)]}{1+\exp[\lambda_n^{(i)}(k)]}$. The calculation of $\lambda_n^{(i)}(k)$ for $i > 0$ will be discussed later in this subsection.

Define the initial LLR for the i-th iteration as

$$
\delta_n^{(i)}(k) = \begin{cases} \lambda_n^{(0)}(k), & k \in \mathscr{T}_n \text{ or } M < k \le M+P, \\ \hat{\lambda}_n^{(i-1)}(k), & \text{otherwise,} \end{cases} \tag{5.8}
$$

In (5.8), for a given user, the initial LLRs of the transmitted information and parity bits are obtained from their respective channel measurements, and they keep unchanged throughout the iterations. The initial LLRs for the punctured information bits are updated as the iterations progress.

The MCMP will be performed on a bipartite graph defined by an extended parity check matrix $\mathbf{H} = [\mathbf{P}_n^T, \mathbf{I}_P]^T \in \mathscr{B}^{(M+P) \times P}$. It should be noted that $\mathbf{H}$ is not the parity check matrix for the codeword defined in (5.2). It is the parity check matrix of the unpunctured codeword with the generation matrix $[\mathbf{I}_M, \mathbf{P}_n]$, because the LLRs of the punctured bits can be calculated from their correlated counterparts transmitted by the other user.

The tanner graph corresponding to $\mathbf{H}$ has $(M+P)$ variable nodes and P check nodes. The k-th variable node is connected to the p-th check node if the (k,p)-th element of $\mathbf{H}$ is 1. Denote $\mathscr{V}_p$ as the set of variable nodes that are connected to the p-th check node, and $\mathscr{C}_k$ as the set of check nodes that are connected to the k-th variable node.

For the information from the n-th user, the message from the k-th variable node to the p-th check node during the i-th iteration is

$$
\eta_{kp}^{(i)}(n) = \delta_n^{(i)}(k) + \sum_{p' \in \mathscr{C}_k \setminus p} \mu_{p'k}^{(i-1)}(n), \tag{5.9}
$$

where $\mu_{pk}^{(i)}(n)$ is the message from the p-th check node to the k-th variable node during the i-th iteration, and it can be calculated as

$$\mu_{pk}^{(i)}(n) = 2\mathrm{atanh}\left(\prod_{k' \in \mathcal{V}_p \setminus k} \tanh \frac{\eta_{k'p}^{(i)}(n)}{2}\right). \tag{5.10}$$

During the first iteration, $\mu_{pk}^{(0)}(n) = 0$.

For a conventional MP algorithm, eqns. (5.9) and (5.10) are performed iteratively for a single codeword, and the values of $\delta_n^{(i)}(k)$ are the same for all the iterations. In the proposed MCMP algorithm, the iterations will be performed across the two codewords through the update of $\delta_n^{(i)}(k)$ as in (5.7) and (5.8).

At the end of the i-th iteration, the n-th user will output a soft decision for its un-punctured information bits, as

$$\lambda_n^{(i)}(k) = \delta_n^{(i)}(k) + \sum_{p' \in \mathcal{C}_k} \mu_{p'k}^{(i)}(n). \tag{5.11}$$

This soft decision from the n-th user will be used to update the LLRs of the punctured bits for the m-th user $(m \neq n)$ as described in (5.8).

At the final iteration, the hard decision is obtained as $\hat{b}_n(k) = 1$ if $\lambda_n^{(I)}(k) > 0$ and $\hat{b}_n(k) = 0$ otherwise. Simulation results demonstrate that the proposed MCMP achieves significant performance gain over conventional MP algorithm that performs iteration for a single codeword.

5.4.2 Joint Decoding for Correlated Markovian Sources

The decoder that utilizes the spatial and temporal correlation of the information sources is discussed in this subsection.

The joint decoder contains a MCMP decoder and a HMM decoder [12]. The MCMP decoder is used to exploit the spatial correlation among the sources, and the HMM decoder can take advantage of the temporal correlation of the data from the same source. The MCMP decoder and the

HMM decoder iteratively exchange soft information through the iterations. Details of the decoding process is described as follows.

The initial values for the joint decoder are $\lambda_n^{(0)}(k)$, for $k \in \mathcal{T}_n$ or $M < k \leq M+P$ which are the LLRs of the unpunctured information and parity bits calculated from (5.5) and (5.6).

5.4.2.1 MCMP Decoding

At the $(i+1)$-th iteration, the soft information available at the input of the MCMP decoder include: $\lambda_n^{(0)}(k)$, for $k \in \mathcal{T}_n$ or $M < k \leq M+P$, which are the LLRs of the unpunctured information and parity bits as calculated from (5.5) and (5.6); $\hat{\lambda}_n^{(i)}(k)$, for $k \notin \mathcal{T}_n$, which are the estimated LLRs of the punctured information bits as calculated from (5.7); $\zeta_j^{(i)}(k) \triangleq P^{(i)}(b_1(k) = j|\mathbf{y}_n)$, for $j = 0, 1$ and $k = 1, \cdots, M$, which are the *a posteriori* probability (APP) of the data from source 1 at the output of the HMM decoder. When $i = 0$, $\zeta_j^{(0)}(k) = \pi_j(k)$, where $\pi_j(k)$ is the stable probability of the Markov process. The calculation of $\zeta_j^{(i)}(k)$, for $i \geq 1$, will be discussed later in this subsection.

It should be noted that only the APP from source 1 is available at the output of the HMM decoder, because the data from source 2 is not a Markov process. Define the APP ratio of the data from sources 1 and 2 as

$$\Delta_1^{(i)}(k) = \log \frac{\zeta_0^{(i)}(k)}{\zeta_1^{(i)}(k)}, \tag{5.12a}$$

$$\Delta_2^{(i)}(k) = \log \frac{(1-p_0)\zeta_0^{(i)}(k) + p_0\zeta_1^{(i)}(k)}{p_0\zeta_0^{(i)}(k) + (1-p_0)\zeta_1^{(i)}(k)} \tag{5.12b}$$

Based on the inputs, the initial LLRs for the i-th iteration can be calculated as

$$\delta_n^{(i)}(k) = \begin{cases} \lambda_n^{(0)}(k), & M < k \leq M+P, \\ \Delta_n^{(0)}(k) + \lambda_n^{(0)}(k), & k \in \mathcal{T}_n \\ \Delta_n^{(i-1)}(k) + \hat{\lambda}_n^{(i-1)}(k), & \text{otherwise.} \end{cases} \tag{5.13}$$

114

Please note that the parity bits do not form a Markov process so the outputs of the HMM decoder are not used in the calculation of the LLRs of the parity bits.

With the initial LLRs given in (5.13), the MCMP algorithm described in the previous subsection can be performed. Specifically, the message from the variable node to the check node is first calculated with (5.9), then the message from the check node to the variable node is calculated with (5.10). The output of the MCMP at the i-th iteration is calculated from (5.11), which will be used as the input for the MCMP decoder for the next iteration as in (5.13). In addition, the soft output of the MCMP will also be used as the input for the HMM decoder of the current iteration.

5.4.2.2 HMM Decoding

Since the data from source 1 is a Markov process and the data from source 2 is a hidden Markov process, the corresponding observations at the receiver, $\mathbf{y}_{\alpha n}$, are hidden Markov processes.

Define the observation likelihood function of the data from source 1 at the i-th iteration as

$$\gamma_j^{(i)}(k) = \begin{cases} p^{(i)}(y_{\alpha 1}(m_k)|b_1(k) = j), & k \in \mathcal{T}_1, \\ p^{(i)}(y_{\alpha 2}(m_k)|b_1(k) = j), & k \in \mathcal{T}_2. \end{cases} \tag{5.14}$$

We propose to update the observation likelihood function with the soft output of the MCMP decoder, as

$$\gamma_j^{(i)}(k) = P_{1j}^{(i)}(k), \qquad\qquad\qquad \text{if } k \in \mathcal{T}_1,$$

$$\gamma_0^{(i)}(k) = (1 - p_0)P_{20}^{(i)}(k) + p_0 P_{21}^{(i)}(k), \quad \text{if } k \in \mathcal{T}_2, \tag{5.15}$$

$$\gamma_1^{(i)}(k) = p_0 P_{20}^{(i)}(k) + (1 - p_0)P_{21}^{(i)}(k), \quad \text{if } k \in \mathcal{T}_2,$$

where $P_{n0}^{(i)}(k) = \frac{\exp[\lambda_n^{(i)}(k)]}{1+\exp[\lambda_n^{(i)}(k)]}$ and $P_{n1}^{(i)}(k) = 1 - P_{n0}^{(i)}(k)$.

The APP, $P^{(i)}(b_1(k) = j|\mathbf{y}_{\alpha 1}, \mathbf{y}_{\alpha 2})$, can be efficiently calculated by using a foward-backward

recursion as [14]

$$\psi_j^{(i)}(k) = \left[\sum_{l=0}^{1}\psi_l^{(i)}(k-1)a_{lj}\right]\gamma_j^{(i)}(k), k = 2,\cdots,M, \tag{5.16}$$

$$\varphi_j^{(i)}(k) = \sum_{l=0}^{1}\varphi_l^{(i)}(k+1)a_{jl}\gamma_l^{(i)}(k+1), k = 1,\cdots,M-1, \tag{5.17}$$

where $\psi_j^{(i)}(k) = p(\mathbf{y}_{\alpha 1}(1:k),\mathbf{y}_{\alpha 2}(1:k),b_1(k) = j)$ is the forward message and initialized by

$\psi_j^{(i)}(1) = \pi_i \times \gamma_j^{(i)}(1)$, with $\mathbf{a}(k_1 : k_2)$ representing a vector formulated by using the k_1-th to k_2-th

elements of the vector $\mathbf{a}$, and $\varphi_j^{(i)}(k) = p(\mathbf{y}_{\alpha 1}(k+1:M),\mathbf{y}_{\alpha 2}(k+1:M)|b_1(k) = j)$ is the backward

message and initialized by $\varphi_j^{(i)}(M) = 0.5$.

After the forward-backward recursion, the APP output of the HMM decoder is calculated by

$$\zeta_j^{(i)}(k) = C_k^{(i)} \times \psi_j^{(i)}(k) \times \varphi_j^{(i)}(k), \tag{5.18}$$

where $C_k^{(i)}$ is a normalizing constant such that $\zeta_0^{(i)}(k) + \zeta_1^{(i)}(k) = 1$. The APP output is then used

as the input to the MCMP decoder in the next iteration as described in (5.12) and (5.13).

The iterations will be performed until convergence or the maximum number of iterations is

reached. The joint MCMP-HMM decoding algorithm is summarized as follows.

I) **Initialization**

  i) Calculate the initial LLRs of the transmitted information bits and parity bits, $\lambda_n^{(0)}(k)$

   with (5.5) and (5.6 for $k \in \mathscr{T}_n$ or $k = M+1,\cdots N$, and $n = 1,2$.

  ii) Calculate the estimated initial LLRs of the punctured information bits, $\hat{\lambda}_n^{(0)}(k)$, with

   (5.7) for $k \in \mathscr{T}_m$, where $m \neq n$.

  iii) Calculate the stable probability of the Markov process $\pi_j(k)$, for $k = 1,\cdots,M$ and

   $j = 0,1$.

iv) Calculate $\Delta_n^{(0)}(k)$ with (5.12), for $k = 1, \cdots, M$ and $n = 1, 2$.

v) Set $i = 1$.

II) MCMP Decoding

i) Calculate the initial LLRs, $\delta_n^{(i)}(k)$, for $k = 1, \cdots, N$, and $n = 1, 2$.

ii) Calculate the message from the variable node to the check node, $\eta_{kp}^{(i)}(n)$, with (5.9), for

$k = 1, \cdots, N$, $p = 1, \cdots, P$, and $N = 1, \cdots, N$.

iii) Calculate the message from the check node to the variable node, $\mu_{pk}^{(i)}(n)$, with (5.10),

for $k = 1, \cdots, N$, $p = 1, \cdots, P$, and $N = 1, \cdots, N$.

iv) Calculate the soft decisions, $\lambda_n^{(i)}(k)$, with (5.11).

v) Update the estimated LLR of the punctured information bits, $\hat{\lambda}_n^{(0)}(k)$, with (5.7) for

$k \in \mathscr{T}_m$, where $m \neq n$.

III) HMM Decoding

i) Calculate the observation likelihood function, $\gamma_j^{(i)}(k)$, with (5.15), for $k = 1, \cdots, N$ and

$j = 0, 1$.

ii) Perform the forward algorithm described in (5.16) to obtain $\psi_j^{(i)}(k)$, for $k = 2, \cdots, M$

and $j = 0, 1$.

iii) Perform the backward algorithm described in (5.17) to obtain $\varphi_j^{(i)}(k)$, for $k = 1, \cdots, M -$

1, and $j = 0, 1$.

iv) Calculate the APP $\zeta_j^{(i)}(k)$ with (5.18), for $k = 1, \cdots, N$ and $j = 0, 1$.

v) Update $\Delta_n^{(i)}(k)$ with (5.12), for $k \in \mathscr{T}_m$ with $n \neq m$.
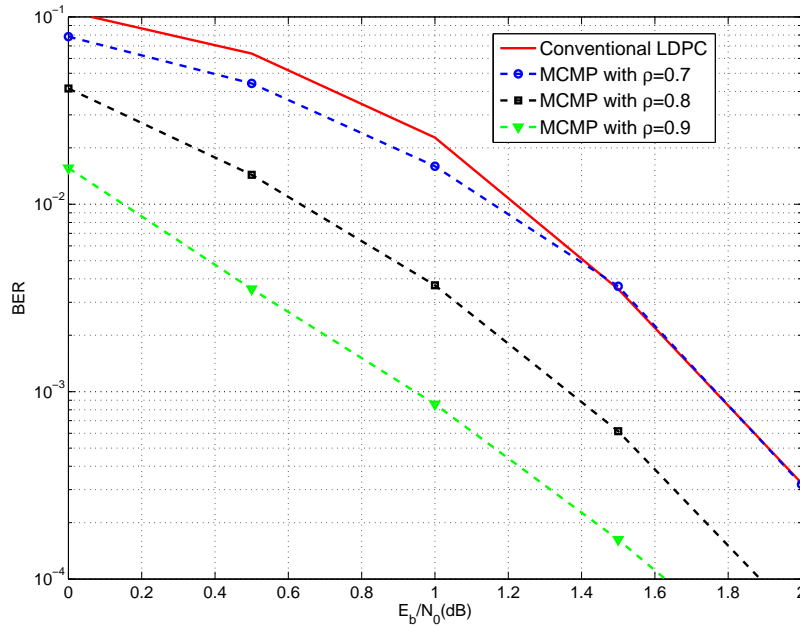
Figure 5.1: The performance of MCMP for memoryless sources under various values of ρ.

vi) If $i \leq I$, go bakc to step II.1); otherwise go to the next step.

III) **Detection**

Make hard decision as $\hat{b}_n(k) = 1$ if $\Delta_n^{(I)}(k) > 0$ and $\hat{b}_n(k) = 0$ otherwise.

5.5 Simulation Results

In this section, the performance of proposed DJSCC scheme with the MCMP and MCMP-HMM algorithms are demonstrated with simulation results.

In the simulation, an irregular LDPC with a 508-by-1016 parity-check matrix is used to generate the DJSCC codeword. We would like to stress that the parity check matrix of the irregular LDPC code has been chosen without any optimization. It is expected that more performance gains can be achieved with a carefully designed parity check matrix. In all the examples, the initial probability of the Markov process is 0.5 for both states.

We first study the bit error rate (BER) performance of the proposed MCMP decoding algorithm for the network with two correlated sources, without considering the Markov property in the time domain. Fig. 5.1 plots the BER of the DJSCC with MCMP decoding algorithm under various values of the spatial correlation coefficient, ρ. The performance of a system with the two sources encoded separately with conventional LDPC code and the MP decoder is also shown for comparison. In the simulation, the transition probabilities $a_{00} = a_{11} = 0.5$, which means both source 1 and source 2 are memoryless. The energy allocation factor is $\theta = 0.5$ for all cases. The iteration number of the MCMP decoder is 20. It is observed that, with unequal energy allocation, the BER performance of the proposed DJSCC with MCMP decoder is better than the conventional LDPC code even at $\rho = 0.7$, and performance improves as ρ increases. At $\rho = 0.9$, the DJSCC with MCMP decoder outperforms the conventional system by 1.8 dB at BER=10^{-3}

The performances of the various decoding algorithms for spatial-temporally correlated sources are compared in Fig. 5.2. The data from source 1 follows a Markov process with transition probabilities $a_{00} = 0.8$ and $a_{11} = 0.7$. The correlation coefficient between sources 1 and 2 is $\rho = 0.8$. As expected, the MCMP-HMM algorithm has the best performance because it exploits both the spatial and temporal data correlations, compared by MCMP and the conventional MP algorithm. At BER=10^{-3}, the MCMP algorithm is about 0.2 dB better than the conventional MP algorithm due to the extra information exchange during the decoding process, and the MCMP-HMM algorithm is about 0.4 dB superior than the DJSCC with the MP algorithm.

Fig. 5.3 demonstrates the performance of the proposed MCMP-HMM decoding algorithm under various spatial and temporal correlations. For comparison, the simulations are performed by considering different HMM parameters as follows: a) $a_{00} = a_{11} = 0.5$, b) $a_{00} = a_{11} = 0.7$, c) $a_{00} = a_{11} = 0.8$. The proposed DJSCC with the MCMP-HMM decoding can effectively exploit
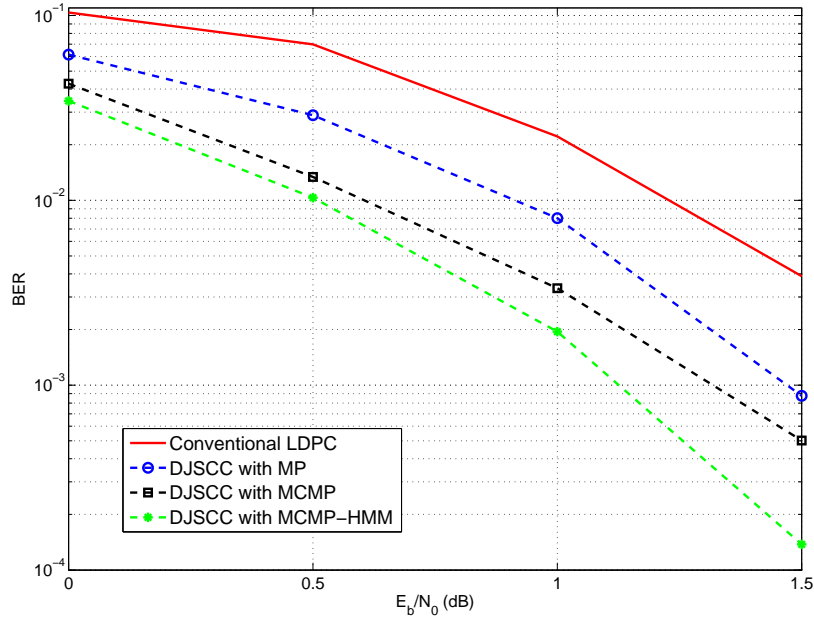
Figure 5.2: Comparison of different iterative decoding algorithms.

the correlation in both the space and time domains, thus increasing the spatial and/or temporal correlation yields better performance. As can be seen from the figure, when $\rho = 0.9$ and BER $= 10^{-3}$, increasing a_{00} and a_{11} from 0.5 to 0.8 leads to a 0.38 dB performance gain.

5.6 Conclusion

In this chapter, we propose a new DJSCC coding scheme for sources with spatially and temporally correlated data. The DJSCC encoding is performed by puncturing the information bits in the encoder and allocating unequal energy for the information and parity bits in the transmission. Two decoding algorithms are proposed. The first algorithm, the MCMP algorithm, is designed for spatially correlated memoryless sources. The second algorithm, the MCMP-HMM algorithm, can effectively exploit the data correlation in both the space and time domains. The MCMP algorithm
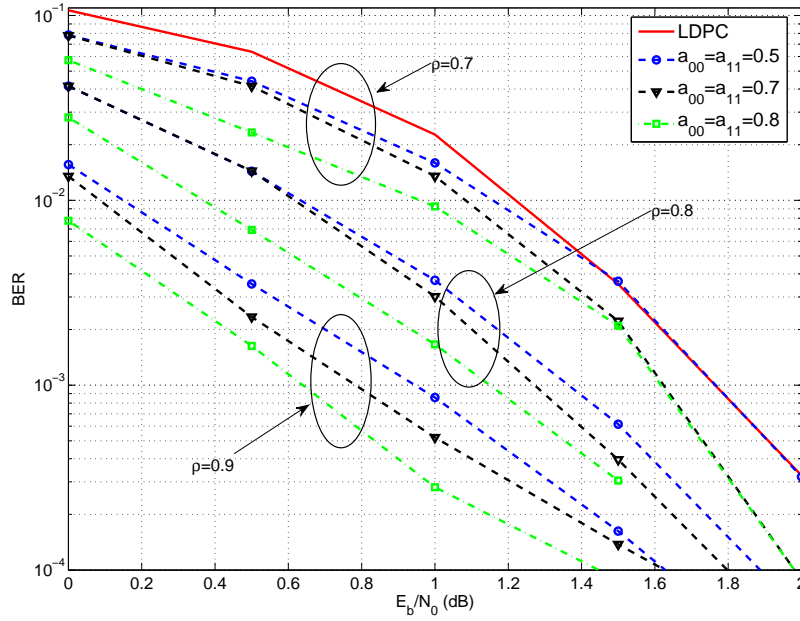
Figure 5.3: The performance of MCMP-HMM for HMM sources under various values of ρ.

performs iterative joint decoding over two correlated sources simultaneously, such that the soft information between the two codewords are exchanged at each iteration to obtain extra performance gains over conventional MP algorithm, where the soft information is only exchanged inside of one codeword. The MCMP-HMM algorithm exchanges soft information between the MCMP decoder an a HMM decoder, which exploit the temporal data correlation. Simulation results demonstrate that the proposed new DJSCC coding schemes achieve considerable performance gains over conventional systems.

5.7 Appendix of documentation of multi-authored chapter

UNIVERSITY OF ARKANSAS.

College of Engineering
Department of Electrical Engineering

July 31, 2013

To Whom It May Concern

This letter is to certify that Ms. Ning Sun, a Ph.D. candidate under my supervision at the Department of Electrical Engineering, has contributed more than 51% of the work for the following paper,

Ning Sun, Jingxian Wu, and Guoqing Zhou, "Distributed joint source-channel code for spatial-temporally correlated Markov sources," in Proceedings of IEEE International Conference on Communications (ICC'13), June 2013.

Ms. Sun is the first author of this paper. The paper is included in her Ph.D. dissertation entitled "Distortion-Tolerant Communications with Correlated Information".

Sincerely Yours,

Jingxian Wu
Director, Wireless Information Network Lab
Associate Professor of Electrical Engineering
University of Arkansas
Email: wuj@uark.edu
Tel: (479) 575-6584

5.8 References

[1] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Information theory*, vol. 19, no. 4, pp. 471-480, Jul., 1973.

[2] S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): Design and construction," *IEEE Trans. Info. Theory*, pp. 626-643, Mar., 2003.

[3] A. D. Liveris, Z. Xiong, and C. N. Georghiades, "Compression of binary sources with side information at the decoder using LDPC codes," *IEEE Communications Letters*, vol. 6, no. 10, pp. 440-442, Oct., 2002.

[4] I. Shahid and P. Yahampath, "Distributed joint source-channel coding of correlated binary sources in wireless sensor networks," in *Proc. IEEE Intern. Symposium on Wireless Commun. Systems*, pp. 236-240, 2011.

[5] N. Gehrig and P. L. Dragotti, "Symmetric and a-symmetric Slepian-Wolf codes with systematic and non-systematic linear codes," *IEEE Communications Letters*, Vol. 9, pp.61 - 63, Jan., 2005.

[6] X. Zhu, L. Zhang, and Y. Liu, "A distributed joint source-channel coding scheme for multiple correlated sources," in *Proc. IEEE Commun. and Networking in China*, pp. 26-28, Aug., 2009

[7] Ning Sun, Jingxian Wu, and Hai Lin, "Distributed joint source and channel code with correlated information sources," in *Proc. IEEE Intern. Conf. Commun. China ICCC'12*, Aug., 2012

[8] J. Garcia-Frias and J. D. Villasenor, "Joint Turbo decoding and estimation of hidden Markov sources," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 9, pp. 1671-1679, Oct., 2001.

[9] L. Yin,J. Lu, and Y. Wu, "Combined hidden Markov source estimation and low-density parity-check coding: a novel joint source-channel coding scheme for multimedia communication," *Wirel. Commun. Mob. Comput.*, pp. 643-650, July, 2002.

[10] S. Majumder and S. Verma, "Joint source-channel decoding of IRA code for hidden Markov source," in *Proc. IEEE Conf. on Recent Advances in Information Technology*, pp. 220-223, Mar., 2012.

[11] M. Fresia and H. V. Poor, "Distributed source coding using Raptor codes for hidden Markov sources," *IEEE Trans. Sig. Processing*, vol. 57, no. 7, pp. 2868-2875, July, 2009.

[12] F. R. Kschischang, B. J. Frey, and H. A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498C519, Feb., 2001.

[13] J. L. Smith, "Approximate stationary probability vectors of a finite Markov chain," *SIAM Journal on Applied Mathematics*, Vol. 20, No. 4, pp. 612-618, Jun., 1971.

[14] Mark Stamp, "A revealing introduction to hidden Markov models," pp. 1-20, April, 2012.

Maximizing Spectral Efficiency for High Mobility Systems with Imperfect Channel State

Information

Ning Sun and Jingxian Wu

6.1 Abstract

This chapter studies the optimum system design that can maximize the spectral efficiency of high mobility wireless communication systems with imperfect channel state information (CSI). High mobility of the wireless terminals results in fast time-varying fading, which can be tracked at the receiver by employing pilot-assisted channel estimation. The percentage of pilot symbols in the transmitted symbols plays a critical role on the system performance: a higher pilot percentage yields a more accurate channel estimation, but also more overhead. The effects of pilot percentage are quantified through the derivation of the channel estimation mean square error (MSE), which is expressed as a closed-form expression of various system parameters through asymptotic analysis. It is discovered that the channel interpolation at non-pilot locations can yield the same asymptotic MSE as the channel estimation at pilot locations if the pilots sample the channel above its Nyquist rate. Based on the statistical properties of the channel estimation error, we quantify the impacts of imperfect CSI on system performance by developing the analytical symbol error rate (SER) and a spectral efficiency lower bound of the communication system. The optimum pilot percentage that can maximize the spectral efficiency lower bound is identified through both analytical and simulation results.

6.2 Introduction

High mobility wireless communications have received increasing attentions recently with the growing demands for applications such as high speed railways and aircraft communications. One of the main challenges faced by high mobility communications is the fast time-varying fading caused by the Doppler shift, which could be as high as 1,000 Hz for a 2.4 GHz system operating at a speed of 450 km/hr. In a high mobility system, the accurate estimation and tracking of the fast time-varying fading are critical to reliable system operations. Channel estimation can be performed either through the direct estimation of the fading coefficients [1]-[7], or through basis expansion models (BEMs) that transform the fading coefficients to low-dimensional transform domains [8, 9].

Many channel estimation related works focus on the design of optimum pilot patterns that can minimize the channel estimation mean square error (MSE) [1]-[3], [9]. In [1], the optimum pilot design for an orthogonal frequency division multiplexing (OFDM) system employing the minimum mean square error (MMSE) channel estimation is discussed, and it is shown that the MSE can be minimized by using identical equally-spaced frequency domain pilot clusters. The MMSE estimator requires the priori knowledge of channel statistics, yet such information is not needed by a least squares (LS) channel estimator. In [2], the optimum pilot pattern for the LS estimation of quasi-static channel in OFDM systems is obtained through numerical convex optimizations. The LS estimation of doubly selective channels are discussed in [3] for a multiple-input multiple-output (MIMO) OFDM system, where the pilot matrix is designed as a unitary matrix to avoid matrix inversions during the LS channel estimation. A windowed LS (WLS) channel estimation with a BEM channel model is proposed in [9], and it is shown that the estimation accuracy of WLS can approach that of MMSE-based estimators. All above methods are designed by using

the MSE as a metric under the constraint of fixed pilot power and/or pilot numbers. They do not consider how the pilot patterns or imperfect channel state information (CSI) will impact the overall communication performance, such as the bit error rate (BER), spectral efficiency, or energy efficiency.

In high mobility systems, channel estimation errors are non-negligible and they might have significant impacts on the system performance and designs [4]-[6]. In [4], it is discovered that systems employing LS or MMSE channel estimations can achieve the same symbol error rate (SER) performance if the optimum receivers are designed by considering the statistics of the channel estimation errors. In [5], the impacts of channel estimation error on the BER of an ultra-wide band (UWB) system are studied. Both [4] and [5] use system error probability as the design metric. An information theoretic metric, a sum-rate lower bound of a two-way relay network, is used in [6] to evaluate the system performance in the presence of imperfect CSI. The sum-rate lower bound is numerically maximized by considering parameters such as training vector structures and the number of training symbols. A quasi-static block fading model is assumed in [4]-[6], thus the results are not applicable to high mobility systems. In [7], the tracking of a time-varying channel is achieved by using polynomial interpolations, and the results are used to quantify the BER of a two-way relay system with analog network coding. It is demonstrated that polynomial interpolations might not be sufficient to track the channel variation in high mobility systems.

In this chapter, the optimum pilot design that maximizes the spectral efficiency of high mobility wireless communication systems is studied. The fast time-varying fading coefficients are estimated and tracked through the MMSE estimation and interpolation. The MSE of both channel estimation at pilot locations and channel interpolation at non-pilot locations are studied through asymptotic analysis, and the results are expressed as closed-form expressions of parameters such

127

as the Doppler spread, the signal-to-noise ratio (SNR) at pilot locations, and the percentage of pilot symbols in the transmitted symbols. It is discovered that, if the pilots sample the channel at or above the Nyquist rate of the time-varying fading, then the MMSE interpolation at non-pilot locations yields the same asymptotic MSE as channel estimation at pilot locations. The statistical properties of the estimated channel coefficients are studied, and the results are used develop an analytical SER and a spectral efficiency lower bound for systems operating with imperfect CSI. A higher pilot percentage yields a better SER. However, a lower SER does not necessarily mean a better overall performance, considering the fact that the excessive use of pilot symbols means more overhead. Such a tradeoff relationship is revealed in the system spectral efficiency. The optimum pilot percentage that can maximize the spectral efficiency lower bound is analytically identified. The impacts of imperfect CSI on system performance are studied through both analytical and simulation results.

The remainder of this chapter is organized as follows. The system model and the MMSE channel estimation are presented in Section 6.3. Section 6.4 studies the analytical asymptotic MSE for both channel estimation at pilot locations and channel interpolation at non-pilot locations. The impacts of the imperfect CSI on the SER are analyzed in Section 6.5 by analyzing the statistical properties of the estimated channel coefficients. In Section 6.6, a spectral efficiency lower bound is developed for systems with imperfect CSI, and the optimum pilot density that maximizes the spectral efficiency is identified. Numerical results are given in Section 6.7, and Section 6.8 concludes the chapter.

6.3 Problem Formulation

6.3.1 System Model

Consider a system that employs pilot-assisted channel estimation and experiences fast time-varying fading. At the transmitter, the data to be transmitted are divided into slots, and each slot has N_s modulated data symbols and $N_p \leq N_s$ pilot symbols. The values of N_s and N_p can be chosen such that $K = \frac{N_s}{N_p}$ is an integer. The pilot symbols are equally spaced such that each pair of adjacent pilot symbols are separated by K data symbols. Denote the symbol vector as $\mathbf{x} = [x_1, \cdots, x_N]^T \in \mathscr{S}^{N \times 1}$, where $N = N_s + N_p$ is the total number of symbols per slot, $\mathscr{S}$ is the modulation alphabet set, and $\mathbf{A}^T$ represents the matrix transpose. Denote the k-th pilot symbol as $x_{i_k} = p_k$, where $i_k = kK$ is the index of the k-th pilot symbol, for $k = 1, \cdots, N_p$. The average energy of the symbols is normalized to 1, $\mathbb{E}(|x_n|^2) = 1$, where $\mathbb{E}$ is the mathematical expectation operator. Define the percentage of the pilot symbols as $\delta = \frac{N_p}{N} = \frac{1}{K+1}$.

The data and pilot symbols are transmitted over the fast time-varying fading channel with additive white Gaussian noise (AWGN). The signals observed at the receiver is

$$\mathbf{y} = \sqrt{E_0} \cdot \mathbf{X} \cdot \mathbf{h} + \mathbf{z}, \tag{6.1}$$

where $\mathbf{y} = [y_1, \cdots, y_N]^T \in \mathscr{C}^{N \times 1}$ is the received signal, $\mathbf{z} = [z_1, \cdots, z_N]^T \in \mathscr{C}^{N \times 1}$ is the AWGN with covariance matrix $\mathbf{R}_z = \sigma_z^2 \mathbf{I}_N$, $\mathbf{I}_N$ is a size-N identity matrix, E_0 is the average transmission energy of a symbol, $\mathbf{h} = [h_1, \cdots, h_N]^T \in \mathscr{C}^{N \times 1}$ is the channel fading coefficient vector, and $\mathbf{X} = \text{diag}(\mathbf{x})$ is a size-N diagonal matrix with the transmitted signal vector $\mathbf{x}$ on its diagonal. The time-varying fading coefficients are correlated with cross-correlation being

$$\rho(m - n) = \mathbb{E}[h_m h_n^*] = J_0(2\pi f_{\mathrm{D}} |m - n| T_s), \tag{6.2}$$

129

where f_{D} is the maximum Doppler spread of the fading channel, T_s is the symbol period, and $J_0(x)$ is the zero-order Bessel function of the first kind.

It is assumed that the energy per symbol E_0 is fixed for both data and pilot symbols. Therefore, the average transmission power is $P_0 = \frac{E_0}{T_s}$.

6.3.2 MMSE Channel Estimation

Before detection, the receiver first performs the channel estimation to obtain an estimate of the fading channel, $\mathbf{h}$, over the entire slot based on distorted observations of pilot symbols, $\mathbf{y}_p = [y_{i_1}, y_{i_2}, \cdots, y_{i_{N_p}}]^T \in \mathscr{C}^{N_p \times 1}$. The channel estimation is performed to minimize the mean square error (MSE) between the estimated and the actual channels. The MSE of the n-th channel coefficient is

$$\sigma_n^2 = \mathbb{E}\left[|\hat{h}_n - h_n|^2\right], \text{ for } n = 1, \cdots, N \tag{6.3}$$

where $\hat{h}_n$ is the n-th estimated channel fading coefficient.

The optimum linear MMSE estimator of $\hat{h}_n$ is

$$\hat{h}_n = \sqrt{E_0}\mathbf{r}_n^H \mathbf{P}^H \left(E_0 \mathbf{P}\mathbf{R}_{hh}\mathbf{P}^H + \sigma_z^2 \mathbf{I}_{N_p}\right)^{-1} \mathbf{y}_p, \tag{6.4}$$

where $\mathbf{r}_n = \mathbb{E}\left[\mathbf{h}_p h_n^*\right] \in \mathscr{C}^{N_p \times 1}$, with $\mathbf{h}_p = [h(i_1), \cdots, h(i_{N_p})]^T \in \mathscr{C}^{N_p \times 1}$ being the fading coefficients at pilot locations, $\mathbf{R}_{hh} = \mathbb{E}\left[\mathbf{h}_p \mathbf{h}_p^H\right] \in \mathscr{C}^{N_p \times N_p}$ with its elements defined in (6.2), $\mathbf{A}^H$ denotes the matrix Hermitian operation, and the diagonal matrix $\mathbf{P} = \mathrm{diag}\{[p_1, \cdots, p_{N_p}]\} \in \mathscr{S}^{N_p \times N_p}$ contains the transmitted pilot symbols. The channel auto-correlation matrix $\mathbf{R}_{hh}$ is a Toeplitz matrix with the (m,n)-th element being $J_0\left(2\pi f_{\mathrm{D}}|m-n|\frac{T_s}{\delta}\right)$ as defined in (6.2), where δ is the percentage of pilot symbols.

With the optimum MMSE estimator given in (6.4), the MSE σ_n^2 can then be calculated as

$$\sigma_n^2 = 1 - \mathbf{r}_n^H \left(\mathbf{R}_{hh} + \frac{1}{\gamma_0} \mathbf{I}_{N_p} \right)^{-1} \mathbf{r}_n, \tag{6.5}$$

where $\gamma_0 = \frac{E_0}{\sigma_z^2}$ is the signal-to-noise ratio (SNR) without fading, and the assumption $|p_n|^2 = 1$ is used in the above equation. This assumption can be easily met by choosing only constant amplitude symbols, such as phase shift keying symbols, as the pilot symbols. It should be noted that the data symbols do not need to be constant amplitude.

The MSE σ_n^2 given in (6.5) is a function of the symbol index n, the SNR γ_0, the data rate R_s, the maximum Doppler frequency f_D, and the pilot percentage δ. Intuitively, given a fixed transmission power, the pilot percentage, δ, plays a critical role on the MSE σ_n^2 and the overall system performance. A smaller pilot percentage means less overhead, thus a higher spectral efficiency. On the other hand, a smaller pilot percentage might not be sufficient to track the fast time-varying fading, and this will degrade the channel estimation accuracy. In the next section, we will study the impact of pilot percentage on the channel estimation MSE.

6.4 Impacts of Pilot Percentage on Channel Estimation

In this section, the impacts of pilot percentage on the channel estimation MSE are analytically studied through asymptotic analysis. The channel estimation is performed in two steps: the receiver first obtains an estimate of the channel coefficients at pilot locations, then the channel coefficients at non-pilot locations are obtained by performing MMSE interpolations over the estimated CSI at pilot locations. It has been shown in [10] that such a two-step MMSE estimation yields the same performance as (6.4).

6.4.1 MMSE Channel Estimation at Pilot Locations

The receiver first obtains an estimate of the channel fading at the pilot locations $\mathbf{h}_p$ by minimizing the average MSE, $\sigma^2_{p,N_p} = \frac{1}{N_p}\mathbb{E}(\|\hat{\mathbf{h}}_p - \mathbf{h}_p\|^2)$, as

$$\hat{\mathbf{h}}_p = \mathbf{W}_p^H \mathbf{y}_p, \tag{6.6}$$

where $\hat{\mathbf{h}}_p = [\hat{h}_{i_1}, \cdots, \hat{h}_{i_{N_p}}]^T \in \mathscr{C}^{N_p \times 1}$ is an estimate of $\mathbf{h}_p$, and $\mathbf{W}_p = \sqrt{E_0}\left(E_0\mathbf{P}\mathbf{R}_{hh}\mathbf{P}^H + \sigma_z^2\mathbf{I}_{N_p}\right)^{-1}\mathbf{P}\mathbf{R}_{hh}$ is the MMSE estimation matrix.

The error correlation matrix, $\mathbf{R}_{ee} = \mathbb{E}\left[\mathbf{e}_p\mathbf{e}_p^T\right]$, with $\mathbf{e}_p = \hat{\mathbf{h}}_p - \mathbf{h}_p$, can be calculated as

$$\mathbf{R}_{ee} = \mathbf{R}_{hh} - \mathbf{R}_{hh}\left(\mathbf{R}_{hh} + \frac{1}{\gamma_0}\mathbf{I}_{N_p}\right)^{-1}\mathbf{R}_{hh} \tag{6.7}$$

where the orthogonal principal, $\mathbb{E}\left[(\hat{\mathbf{h}}_p - \mathbf{h}_p)\mathbf{y}_p^H\right] = 0$, is used in the above equality.

The average MSE can then be calculated as

$$\sigma^2_{p,N_p} = \frac{1}{N_p}\text{trace}(\mathbf{R}_{ee}). \tag{6.8}$$

From (6.7) and (6.8), the calculation of the MSE involves matrix inversion and the trace operation. In order to explicit identify the impacts of pilot percentage on the MSE, we resort to the asymptotic analysis by letting $N_p \to \infty$ and $N_s \to \infty$ while keeping a finite pilot percentage δ and data rate $R_s = \frac{1}{T_s}$. The results are presented as follows.

Proposition 6.1: When $N_p \to \infty$ while keeping a finite δ and R_s, the asymptotic MSE, $\sigma_p^2 = \lim_{N_p \to \infty} \sigma^2_{p,N_p}$, of the estimated channel coefficient at the pilot locations is

$$\sigma_p^2 = 1 - \frac{8\gamma_0 \arctan\left(\sqrt{\frac{2\gamma_0 - \frac{\alpha}{\delta}}{2\gamma_0 + \frac{\alpha}{\delta}}}\right)}{\pi\sqrt{(2\gamma_0)^2 - (\frac{\alpha}{\delta})^2}}, \quad \text{for } \delta \geq \frac{\alpha}{\pi}, \tag{6.9}$$

where $\alpha = 2\pi f_D T_s$, γ_0 is the SNR without fading, and δ is the pilot percentage.

Proof: The proof is in Appendix 6.9.1. ∎

Intuitively, the asymptotic MSE should be a decreasing function in the pilot percentage δ, because a larger δ means a higher sampling rate of the time-varying channel. This intuition is corroborated by the following corollary.

Corollary 6.1: The asymptotic MSE given in (6.9) is a monotonic decreasing function in the pilot percentage δ and an increasing function in the maximum Doppler spread f_{D}, for $\delta \geq \frac{\alpha}{\pi}$.

Proof: The proof is in Appendix 6.9.2. ∎

6.4.2 MMSE Channel Interpolation

Once the estimates of the channel information at the pilot locations are obtained, they can be interpolated to obtain the channel estimations of the entire slot.

Consider the estimation of the fading coefficients with symbol indices $\{i'_k = (k-1) + u\}_{k=1}^{N_p}$, where $u = 2, \cdots, K-1$ correspond to the indices of the non-pilot data symbols. Define the fading vector to be estimated through interpolation as $\mathbf{h}_d = [h(i'_1), \cdots, h(i'_{N_p})]^T \in \mathscr{C}^{N_p \times 1}$.

Following the orthogonal principal, $\mathbb{E}\left[(\hat{\mathbf{h}}_d - \mathbf{h}_d)\hat{\mathbf{h}}_p^T\right] = \mathbf{0}$, where $\hat{\mathbf{h}}_d$ is an estimate of $\mathbf{h}_d$, the MMSE spatial interpolation can be expressed by

$$\hat{\mathbf{h}}_d = \mathbf{R}_{d\hat{h}}\mathbf{R}_{\hat{h}\hat{h}}^{-1}\hat{\mathbf{h}}_p, \tag{6.10}$$

where

$$\mathbf{R}_{d\hat{h}} \triangleq \mathbb{E}(\mathbf{h}_d\hat{\mathbf{h}}_p^H) = \sqrt{E_0}\mathbf{R}_{dh}\mathbf{P}^H\mathbf{W}_p, \tag{6.11a}$$

$$\mathbf{R}_{\hat{h}\hat{h}} \triangleq \mathbb{E}(\hat{\mathbf{h}}_p\hat{\mathbf{h}}_p^H) = \mathbf{W}_p^H(E_0\mathbf{P}\mathbf{R}_{hh}\mathbf{P}^H + \sigma_z^2\mathbf{I}_{N_p})\mathbf{W}_p. \tag{6.11b}$$

The cross-correlation matrix, $\mathbf{R}_{dh} = \mathbb{E}(\mathbf{h}_d \hat{\mathbf{h}}_p^H) \in \mathscr{R}^{N_p \times N_p}$, is a Toeplitz matrix with its first row being $[\rho(-K+u), \rho(-2K+u), \cdots, \rho(-N_pK+u)]$ and the first column $[\rho(-K+u), \rho(u), \cdots, \rho((N_p - 2)K+u)]^T$.

Combining (6.6), (6.10) and (6.11) yields

$$\hat{\mathbf{h}}_d = \sqrt{E_0}\mathbf{R}_{dh}\mathbf{P}^H \left(E_0 \mathbf{P}\mathbf{R}_{hh}\mathbf{P}^H + \sigma_z^2 \mathbf{I}_{N_p}\right)^{-1} \mathbf{y}_p. \tag{6.12}$$

The corresponding error correlation matrix, $\blacksquare_{ee} \triangleq \mathbb{E}\left[(\hat{\mathbf{h}}_d - \mathbf{h}_d)(\hat{\mathbf{h}}_d - \mathbf{h}_d)^T\right]$, can then be calculated by

$$\blacksquare_{ee} = \mathbf{R}_{hh} - \mathbf{R}_{dh}\left(\mathbf{R}_{hh} + \frac{1}{\gamma_0}\mathbf{I}_{N_p}\right)^{-1}\mathbf{R}_{hd}, \tag{6.13}$$

where $\mathbf{R}_{dd} = \mathbb{E}(\mathbf{h}_d \mathbf{h}_d^H) = \mathbf{R}_{hh}$ is used in the above equation, and $\mathbf{R}_{hd} = \mathbf{R}_{dh}^H$. The average MSE for spatial interpolation is $\sigma_{e,N_p}^2 = \frac{1}{N_p}\text{trace}(\blacksquare_{ee})$. The asymptotic average MSE for channel estimation through temporal interpolation is given in the following proposition.

Proposition 6.2: When $N_p \to \infty$ while keeping a finite δ, if $\delta \geq \frac{\alpha}{\pi}$, then channel estimations through temporal interpolation yields the same asymptotic MSE as channel estimations at pilot locations, i.e., $\sigma_e^2 = \lim_{N_p \to \infty} \sigma_{e,N_p}^2 = \sigma_p^2$, with σ_p^2 defined in (6.9).

Proof: The proof is in Appendix 6.9.3. $\blacksquare$

The results in Proposition 6.2 state that the temporal interpolation will not degrade the channel estimation performance, as long as the channel coefficients are sampled by the pilots at a rate no less than the Nyquist rate, $\frac{1}{T_p} \geq 2f_D$, or equivalently $\delta \geq \frac{\alpha}{\pi}$. The temporal interpolation introduces a time shift in the correlation between $\mathbf{h}_d$ and $\mathbf{h}_p$. A shift in the time domain corresponds to a phase shift in the frequency domain. The asymptotic MSE is only related to the squared amplitude of the frequency domain representation of the channel correlation. If there is no spectrum aliasing, then the phase shift does not have any impact on the asymptotic MSE.

6.5 Impacts of Pilot Percentage on Symbol Error Probability

In this section, the statistical properties of the estimated channel are studied, and the results are used to derive the SER in the presence of imperfect CSI.

6.5.1 Statistical Properties of the Estimated Channel

To build an explicit relationship between the channel estimation MSE and the SER, we first study the statistical properties of the estimated channel in this subsection. To simplify notation, the data symbol index is dropped in the subsequent analysis.

Proposition 6.3: For a system operating in a Rayleigh fading channel, the estimated channel coefficient, $\hat{h}$, is a complex Gaussian random variable (CGRV) with zero mean and variance $\sigma_{\hat{h}}^2 = 1 - \sigma_e^2$, i.e., $\hat{h} \sim \mathcal{N}(0, 1 - \sigma_e^2)$, where σ_e^2 is the channel estimation MSE.

Proof: The proof is in Appendix 6.9.4. ∎

Corollary 6.2: Consider a system operating in a Rayleigh fading channel. Conditioned on the estimated channel coefficient $\hat{h}$, the true channel coefficient h is Gaussian distributed with mean $\hat{h}$ and variance σ_e^2, i.e., $h|\hat{h} \sim \mathcal{N}(\hat{h}, \sigma_e^2)$.

Proof: The proof is in Appendix 6.9.5. ∎

The receiver performs detection based on the knowledge of the received sample y and the estimated channel coefficient $\hat{h}$. We have the following corollary regarding the likelihood function, $p(y|\hat{h}, x)$, in the presence of imperfect CSI.

Corollary 6.3: Consider a system operating in a Rayleigh fading channel. If the channel coefficient is obtained through MMSE channel estimation, then the likelihood function, $p(y|\hat{h}, x)$, is

a Gaussian probability density function (pdf), with the conditional mean, $u_{y|x,\hat{h}}$, and conditional

variance, $\sigma^2_{y|x,\hat{h}}$, given by

$$u_{y|x,\hat{h}} = \sqrt{E_0}\hat{h}x, \tag{6.14}$$

$$\sigma^2_{y|x,\hat{h}} = E_0\sigma^2_e|x|^2 + \sigma^2_z. \tag{6.15}$$

where σ^2_e is the channel estimation MSE.

Proof: The proof is in Appendix 6.9.6. ∎

6.5.2 SER in the Presence of Imperfect CSI

The SER performance of systems with imperfect CSI is studied in this subsection by utilizing the

statistical properties of the channel estimation error.

For systems with equiprobable transmitted symbols and imperfect CSI, the SER can be min-

imized by maximizing the likelihood function, $p(y|\hat{h},x)$, which is a Gaussian pdf with the con-

ditional mean and variance given in Corollary 6.3. From Corollary 6.3, the maximum likelihood

decision rule for system with M-ary phase shift keying (MPSK) can be expressed as

$$\hat{x} = \operatorname*{argmin}_{x \in \mathscr{S}} \left\{ \frac{|y - u_{y|x,\hat{h}}|^2}{\sigma^2_{y|x,\hat{h}}} \right\} = \operatorname*{argmin}_{x \in \mathscr{S}} \left\{ |\mu - x|^2 \right\} \tag{6.16}$$

where $\mathscr{S}$ is the MPSK modulation alphabet set, and $\mu = \frac{1}{\sqrt{E_0}}\hat{h}^*y$ is the decision variable for MMSE

channel estimation, with a^* being the complex conjugate operator.

From eqn. (6.16), the SER can be calculated by finding the probability that the decision variable

μ is outside the decision region of the transmitted symbol, thus the SER depends on the statistical

properties of μ. Given $\hat{h}$ and the transmitted symbol x, the decision variable μ is also Gaussian

distributed with the conditional mean and variance given by

$$u_{\mu|x,\hat{h}} = |\hat{h}|^2 x, \tag{6.17a}$$

$$\sigma^2_{\mu|x,\hat{h}} = |\hat{h}|^2(\sigma_e^2 + \frac{1}{\gamma_0}). \tag{6.17b}$$

Note that the identity $|x|^2 = 1$ is used in the above derivation for the MPSK modulated system.

With the statistical properties of the decision variable given in (6.17), the SER of MPSK modulated systems with the imperfect CSI is given in the following proposition.

Proposition 6.4: For an MPSK modulated system operating in fast time-varying Raleigh fading channels, if the channel is estimated with an MMSE estimator, then the SER is

$$P(E) = \frac{1}{\pi}\int_0^{\pi-\frac{\pi}{M}}\left[1 + \zeta \cdot \frac{\sin^2(\frac{\pi}{M})}{\sin^2(\phi)}\right]^{-1} d\phi, \tag{6.18}$$

where $\zeta = \frac{1-\sigma_e^2}{\sigma_e^2 + \frac{1}{\gamma_0}}$, and σ_e^2 is the asymptotic MSE of the channel estimation given in Proposition 2.

Proof: The proof is in Appendix 6.9.9. ∎

In Proposition 6.4, the SER is expressed as a function of the channel estimation MSE σ_e^2, and the SNR γ_0. Since the asymptotic MSE σ_e^2 is a function of δ and f_D, the SER can be expressed as an explicit function in δ, f_D, and γ_0.

Corollary 6.4: The SER given in (6.18) is a monotonically decreasing function in the pilot percentage δ and an increasing function in the maximum Doppler spread f_D, for $\delta \geq \frac{\alpha}{\pi}$.

Proof: It is easy to show that $P(E)$ in (6.18) is an increasing function in σ_e^2. The relationship between $P(E)$ and δ or f_D can then be obtained by using the results of corollary 6.1. ∎

Based on the results in corollary 6.4, a higher pilot percentage yields a better SER performance. However, more pilots means more overhead, and this might negatively affect the system spectral

efficiency. This tradeoff relationship is studied through spectral efficiency analysis in the next section.

6.6 Maximizing Spectral Efficiency with Imperfect Channel Information

In this section, we study the optimum pilot design by maximizing a lower bound of the spectral efficiency in the presence of imperfect CSI. A higher pilot percentage yields a better channel estimation, thus less detection errors at the receiver. On the other hand, increasing pilot percentage will decrease the number of data bits transmitted per unit time per unit bandwidth.

Considering the presence of both pilot symbols and channel estimation error, we can calculate the effective system spectral efficiency as

$$\eta = \mathbb{E}_{\hat{h}}\left[\frac{N_s}{N}C(\hat{h})\right] = (1-\delta)\mathbb{E}_{\hat{h}}\left[C(\hat{h})\right], \tag{6.19}$$

where the expectation is performed with respect to $\hat{h}$, $C(\hat{h}) = \max_{p(x)} \mathrm{I}(y;x|\hat{h}_d)$ is the maximum mutual information between y and x given the knowledge of the estimated channel coefficient $\hat{h}$, with $p(x)$ being the pdf of the input x. $C(\hat{h})$ can be considered as the channel capacity in the presence of imperfect CSI, and it quantifies the impact of channel estimation error on the channel capacity.

It is difficult to obtain the exact expression of the conditional channel capacity $C(\hat{h})$. A lower bound on $C(\hat{h})$ is given as follows.

Lemma 6.1: For a system operating in a Rayleigh fading channel with pilot-assisted MMSE channel estimation, the channel capacity conditioned on the estimated channel coefficient is lower

bounded by

$$C_{\text{low}}(\hat{h}) = \log\left(1 + |\hat{h}|^2 \frac{1}{\sigma_e^2 + \frac{1}{\gamma_0}}\right) \tag{6.20}$$

Proof: The proof is in Appendix 6.9.8. ∎

Based on the results in Lemma 6.1, a lower bound on the effective spectral efficiency is given by the following proposition.

Proposition 6.5: For a system that employs MMSE channel estimation and experiences Rayleigh fading, the average spectral efficiency is lower bounded by

$$\eta_{\text{low}} = (1 - \delta)\exp\left(\frac{\sigma_e^2 + \frac{1}{\gamma_0}}{1 - \sigma_e^2}\right)\Gamma\left(0, \frac{\sigma_e^2 + \frac{1}{\gamma_0}}{1 - \sigma_e^2}\right), \tag{6.21}$$

where $\Gamma(s,x) = \int_x^\infty t^{s-1}e^{-t}dt$ is the incomplete Gamma function.

Proof: The proof is in Appendix 6.9.9. ∎

From Corollary 6.1, σ_e^2 is a decreasing function in δ and an increasing function in f_{D}. It is straightforward to show that $\xi(\delta) = \frac{1 - \sigma_e^2}{\sigma_e^2 + \frac{1}{\gamma_0}}$ is an increasing function in δ. The spectral efficiency lower bound in (6.45) can thus be decomposed into two components, $g_1(\delta) = (1 - \delta)$ and $g_2(\delta) = \int_0^\infty \exp(-v)\log\left(1 + v\frac{1 - \sigma_e^2}{\sigma_e^2 + \frac{1}{\gamma_0}}\right)dv$. The linear function $g_1(\delta)$ is strictly decreasing in δ, and it contributes to the spectral efficiency loss due to a higher pilot percentage. On the other hand, $g_2(\delta)$ is an increasing function in δ, and it contributes to the spectral efficiency gain due to a more accurate channel estimation at a higher δ. Therefore, $g_1(\delta)$ and $g_2(\delta)$ reveal the two opposite effects of the pilot percentage δ on the spectral efficiency.

The spectral efficiency lower bound is shown as a function of δ in Fig. 6.1 under various values of the normalized Doppler spread $f_{\text{D}}T_s$, where T_s is the symbol period. For a system with symbol rate 1 Msym/s and operating at 1.9 GHz, the range of the Doppler spread considered in the figure

is between 100 Hz ($f_D T_s = 10^{-3}$) to 1 KHz ($f_D T_s = 10^{-2}$), which correspond to a mobile speed in the range between 56.8 km/hr and 568.4 km/hr. It can be seen that η_{low} is a concave function of δ due to the tradeoff relationship between $g_1(\delta)$ and $g_2(\delta)$. When δ is small, *e.g.*, $\delta < 0.095$ for $f_D T_s = 0.01$, η_{low} increases in δ. This indicates that, when δ is small enough, the impacts of channel estimation error dominates the effective spectral efficiency. On the other hand, when δ becomes large enough, *e.g.*, $\delta > 0.095$ for $f_D T_s = 0.01$, increasing δ further will degrade the spectral efficiency because of the excessive overhead caused by the high percentage of the pilot symbols. The optimum pilot percentage that maximize the spectral efficiency lower bound can be obtained by solving the equation $\frac{\partial \eta_{\text{low}}}{\partial \delta} = 0$, which can be expressed as

$$
\exp\left(\frac{\sigma_e^2 + \frac{1}{\gamma_0}}{1 - \sigma_e^2}\right)\Gamma\left(0, \frac{\sigma_e^2 + \frac{1}{\gamma_0}}{1 - \sigma_e^2}\right)\left[(1-\delta)\frac{1 + \frac{1}{\gamma_0}}{(1 - \sigma_e^2)^2}(\sigma_e^2)' - 1\right]
$$
$$
- (1-\delta)\frac{1 + \frac{1}{\gamma_0}}{\sigma_e^2 + \frac{1}{\gamma_0}}\frac{1}{1 - \sigma_e^2}(\sigma_e^2)', \tag{6.22}
$$

where $(\sigma_e^2)'$ is the first derivative of σ_e^2 given in eqn. (6.30). The above equation in δ can be solved numerically by using standard software packages, such as fsolve in Matlab.

6.7   Numerical Results

Numerical and simulation results are provided in this section to verify the results obtained in this chapter, and to demonstrate the impacts of imperfect CSI on the system performance.

Fig. 6.2 shows the asymptotic MSE in Proposition 6.2 as a function of pilot percentage, under various values of the normalized Doppler spread $f_D T_s$. The SNR is $\gamma_0 = 10$ dB. The MSE obtained from simulations is also shown in the figure. In the simulations, a frame length of $N = 3,000$ is used to approximate the infinite frame length. Both the MSE for channel estimation at pilot positions and the MSE for channel interpolation at non-pilot locations with temporal interpolation are shown in
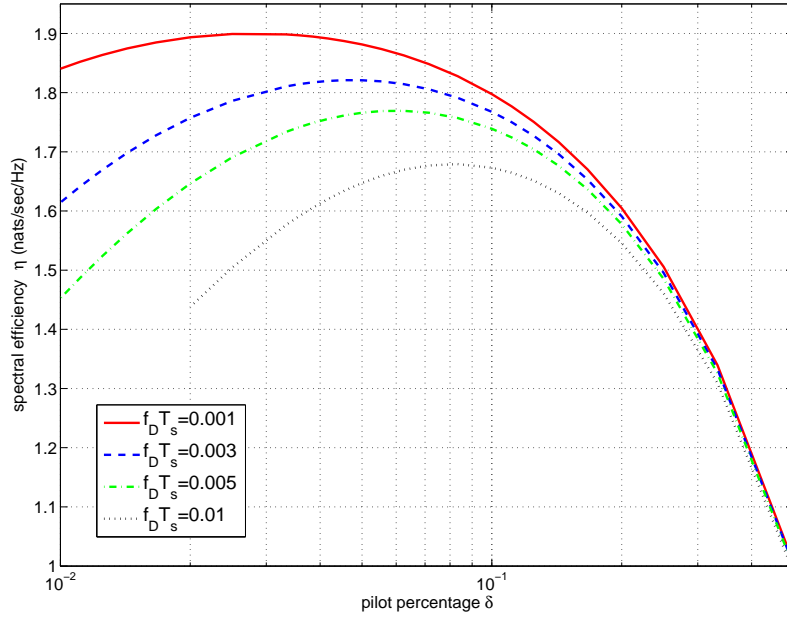
Figure 6.1: The spectral efficiency as a function of the pilot percentage.

the figure, and they are the same as predicted by Proposition 6.2. Excellent agreement is observed between the analytical MSE obtained with infinite frame length and the simulation results with finite frame length. As expected, the MSE is a decreasing function in δ, and an increasing function in $f_{\mathrm{D}}T_s$.

The analytical and simulated SER of the data symbols are shown in Fig. 6.3 under various values of the normalized Doppler spread $f_{\mathrm{D}}T_s$. The SNR is $\gamma_0 = 10$ dB, and the modulation is BPSK. The analytical results can accurately predict the actual SER in the presence of imperfect CSI. Similar to the MSE, the SER is a decreasing function in δ and an increasing function in $f_{\mathrm{D}}T_s$. In addition, it is also observed from the Fig. 6.3 that, when δ is small, the SER decreases dramatically as δ increases. When δ reaches a certain threshold, increasing δ further only yields very small performance gains, i.e., the slope of the SER-δ curve approaches zero as δ increases. Therefore, the desired pilot percentage can be chosen as the point such that $\left| \frac{\partial P(E)}{\partial \delta} \right| = \varepsilon$, with ε
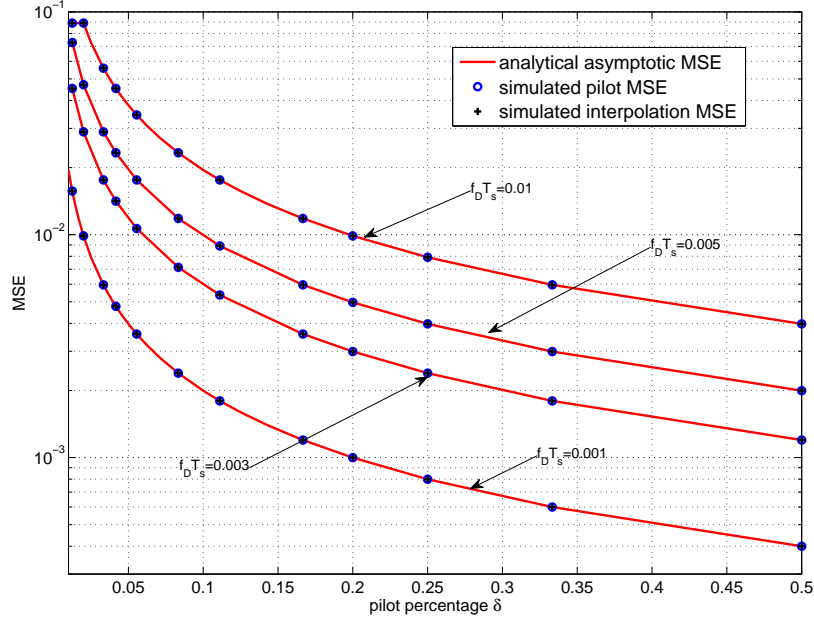
Figure 6.2: The asymptotic MSE as a function of the pilot percentage.

being a small number. The slope of the SER-δ curve can be calculated as

$$\frac{\partial P(E)}{\partial \delta} = \frac{1}{\pi} \int_0^{\pi - \frac{\pi}{M}} \left\{ \frac{(1 + \frac{1}{\gamma_0})(\sigma_e^2)' \frac{\sin^2(\frac{\pi}{M})}{\sin^2(\phi)}}{\left[\sigma_e^2 + \frac{1}{\gamma_0} + (1 - \sigma_e^2) \frac{\sin^2(\frac{\pi}{M})}{\sin^2(\phi)} \right]^2} \right\} d\phi \qquad (6.23)$$

where $(\sigma_e^2)'$ is the first derivative of σ_e^2 with respect to δ, and it is given in eqn. (6.30).

In Fig. 6.4, the desired pilot percentage is solved by choosing $\varepsilon = 10^{-5}$ and shown as functions of the the normalized Doppler spread $f_D T_s$ under different SNRs. The desired pilot percentage increases as $f_D T_s$ increases. This is intuitive because a channel that changes faster needs a higher percentage of pilots.

Fig. 6.5 shows the spectral efficiency maximizing pilot percentage as a function of the normalized Doppler spread $f_D T_s$, under various values of SNR. The optimum pilot percentage is calculated by solving (6.22). The optimum pilot percentage is a monotonically increasing function in $f_D T_s$, because more pilots per unit time are required to compensate the faster channel variation at higher
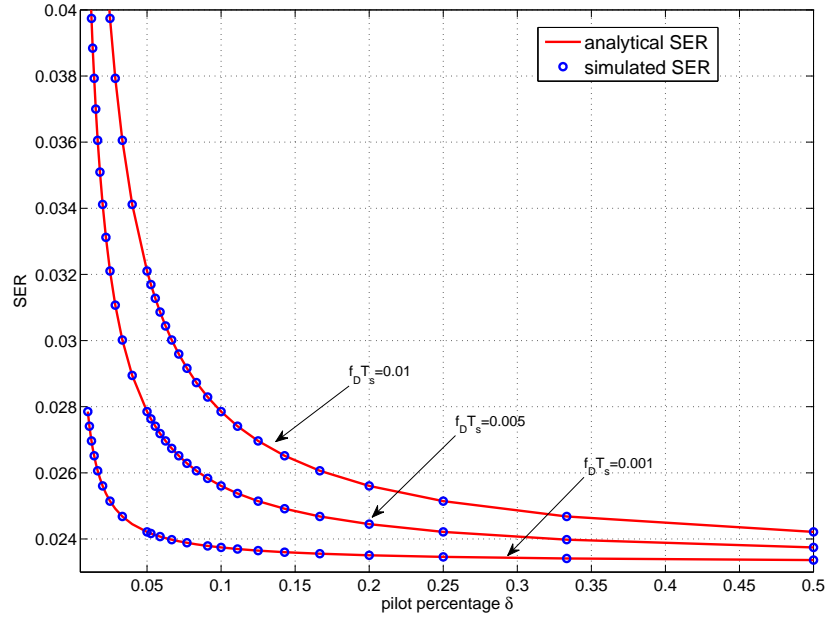
142

Figure 6.3: The SER as a function of the pilot percentage δ.

Doppler spread. At SNR = 10 dB, increasing $f_{\mathrm{D}}T_s$ from 2×10^{-3} to 10^{-2} will double the optimum pilot percentage from 4% to 8%. In addition, a lower pilot percentage is required for systems with higher SNR due to the better channel estimation quality when the SNR is high.

6.8   Conclusions

The optimum system design for high mobility wireless communication systems with imperfect CSI has been studied in this chapter. The asymptotic channel estimation MSE has been quantified as a closed-form expression of the percentage of pilots used for MMSE channel estimation. By analyzing the statistical properties of the estimated channel coefficients, we derive the explicit SER and a spectral efficiency lower bound of a communication system operating with imperfect CSI. It has been shown through theoretical study that, if the pilot samples the channel at a rate no less than the Nyquist rate of the time-varying channel, MMSE channel estimation at pilot locations or
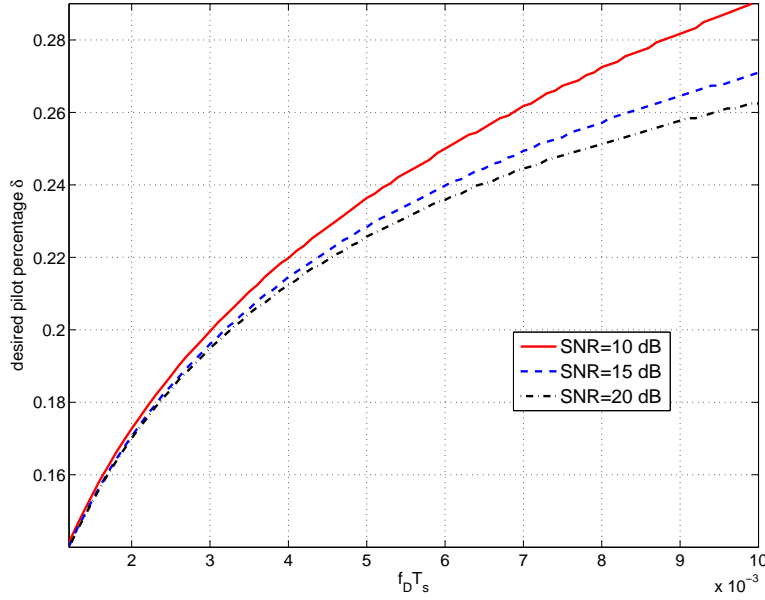
Figure 6.4: The desired pilot percentage obtained by solving (6.23) with $\varepsilon = 10^{-5}$.

MMSE channel interpolation at non-pilot locations yield the same MSE. In addition, the SER is a monotonic decreasing function in pilot density, yet the spectral efficiency is concave in the pilot density.

6.9 Appendix of Proofs

6.9.1 Proof of Proposition 6.1

Performing eigenvalue decomposition of $\mathbf{R}_{hh}$ in (6.7), we can rewrite the MSE as

$$
\sigma_{p,N_p}^2 = \frac{1}{N_p} \sum_{n=1}^{N_p} \left[\lambda_n - \left(\lambda_n + \frac{1}{\gamma_0} \right)^{-1} \lambda_n^2 \right]
$$
$$
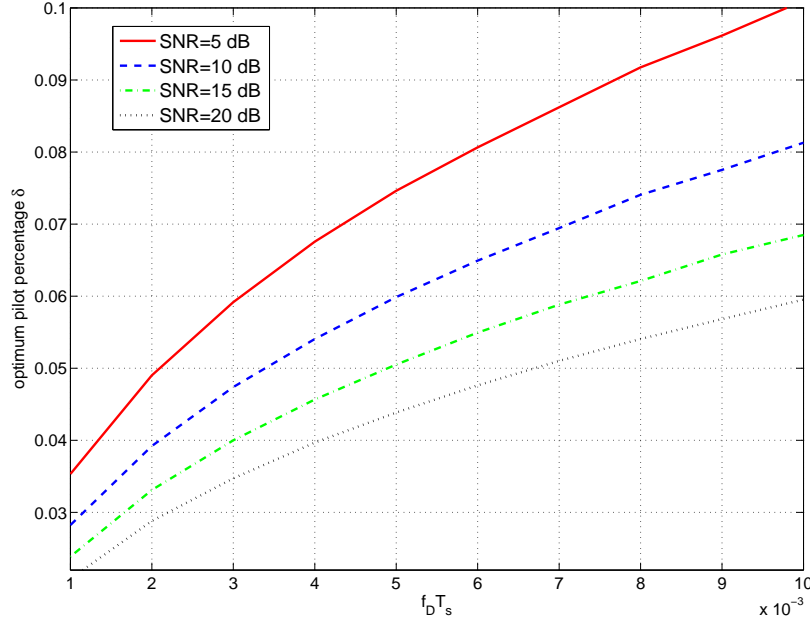= \frac{1}{N_p} \sum_{n=1}^{N_p} \left(\frac{\lambda_n}{\lambda_n \gamma_0 + 1} \right), \tag{6.24}
$$

144

Figure 6.5: The optimum pilot percentage as a function of the normalized Doppler spread.

where λ_n is the n-th eigenvalue of $\mathbf{R}_{hh}$. Based on Szego's Theorem [12], when $N_p \to \infty$, (6.24) can

be rewritten as

$$\sigma_p^2 = \lim_{N_p \to \infty} \sigma_{p,N_p}^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[\frac{\Lambda(\Omega)}{\Lambda(\Omega)\gamma_0 + 1} \right] d\Omega, \tag{6.25}$$

where $\Lambda(\Omega) = \sum_{k=-\infty}^{\infty} J_0\left(2\pi f_{\mathrm{D}}|k|T_p\right) e^{-jk\Omega}$ is the discrete-time Fourier transform (DTFT) of $\{J_0\left(2\pi f_{\mathrm{D}}|k|\frac{T_s}{\delta}\right)\}_k$,

with $T_p = \frac{T_s}{\delta}$ being the space between two pilot symbols.

The Fourier transform (FT) of the continuous-time function $J_0(2\pi f_{\mathrm{D}}t)$ is [14]

$$\Lambda_c(\omega) = \frac{2\mathrm{rect}(\frac{\omega}{4\pi f_{\mathrm{D}}})}{\sqrt{(2\pi f_{\mathrm{D}})^2 - \omega^2}}, \tag{6.26}$$

where $\omega = \frac{\Omega}{T_p}$ is the analog angular frequency with unit radians per second. Based on the sampling

theorem and (6.26), the DTFT $\Lambda(\Omega)$ can be written as

$$\Lambda(\Omega) = \frac{1}{T_p} \sum_{n=-\infty}^{\infty} \Lambda\left(\omega - n\frac{2\pi}{T_p}\right) = \sum_{n=-\infty}^{\infty} \frac{2\mathrm{rect}(\frac{\Omega-2\pi n}{2\beta})}{\sqrt{\beta^2 - (\Omega - 2\pi n)^2}}, \tag{6.27}$$

where $\beta = 2\pi f_{\mathrm{D}} T_p$.

145

Based on the sampling theorem, when $\delta \geq 2f_D T_s$, there is no spectrum aliasing in (6.27), thus the MSE in (6.25) can be simplified to

$$\sigma_p^2 = \frac{1}{2\pi} \int_{-\beta}^{\beta} \left[\frac{2}{2\gamma_0 + \sqrt{\beta^2 - \Omega^2}} \right] d\Omega \tag{6.28}$$

It should be noted that $\beta \leq \pi$ when $\delta \geq 2f_D T_s$.

By changing the integration variable $\Omega = \beta \sin(x)$, we can solve the above integral with the following integration

$$\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} [a + b\cos(x)]^{-1} dx = \frac{4\arctan\left(\sqrt{\frac{a-b}{a+b}}\right)}{\sqrt{a^2 - b^2}}, \tag{6.29}$$

where the equation is derived by combining [11, eqn. (2.553.3)] with the identity $\arctan(jx) = j\mathrm{arctanh}(x)$ for $x \in \mathscr{R}$ and $j^2 = -1$. The results in (6.9) can then be obtained by applying (6.29) to (6.28).

6.9.2   Proof of Corollary 6.1

We will consider two cases: $\delta \geq \frac{\alpha}{2\gamma_0}$, and $\frac{\alpha}{\pi} \leq \delta < \frac{\alpha}{2\gamma_0}$ with $\gamma_0 < \frac{\pi}{2}$. When $\delta > \frac{\alpha}{2\gamma_0}$, the first derivative of σ_p^2 with respect to δ can be written as

$$\frac{\partial \sigma_p^2}{\partial \delta} = \frac{-4\gamma_0\alpha\left[\delta\sqrt{(2\gamma_0)^2 - (\frac{\alpha}{\delta})^2} - 2\alpha\arctan\left(\sqrt{\frac{2\gamma_0 - \frac{\alpha}{\delta}}{2\gamma_0 + \frac{\alpha}{\delta}}}\right)\right]}{\pi\delta^3\left[(2\gamma_0)^2 - (\frac{\alpha}{\delta})^2\right]^{\frac{3}{2}}}. \tag{6.30}$$

Decompose (6.30) into two parts as $f_1(\delta) = \frac{-4\gamma_0\alpha}{\pi\delta^3[(2\gamma_0)^2 - (\frac{\alpha}{\delta})^2]^{\frac{3}{2}}}$ and $f_2(\delta) = \delta\sqrt{(2\gamma_0)^2 - (\frac{\alpha}{\delta})^2} - 2\alpha\arctan\left(\sqrt{\frac{2\gamma_0 - \frac{\alpha}{\delta}}{2\gamma_0 + \frac{\alpha}{\delta}}}\right)$. It is obvious $f_1(\delta) \leq 0$ with $\delta \geq \frac{\alpha}{2\gamma_0}$. Since $f_2'(\delta) = \sqrt{(2\gamma_0)^2 - (\frac{\alpha}{\delta})^2} \geq 0$, $f_2(\delta)$ is an increasing function. As a result $f_2(\delta) \geq f_2(\frac{\alpha}{2\gamma_0}) = 0$. Therefore, $\frac{\partial \sigma_p^2}{\partial \delta} = f_1(\delta)f_2(\delta) \leq 0$ when $\delta \geq \frac{\alpha}{2\gamma_0}$.

When $\frac{\alpha}{\pi} \leq \delta < \frac{\alpha}{2\gamma_0}$ and $\gamma_0 < \frac{\pi}{2}$, with the identity $\arctan(jx) = j\mathrm{arctanh}(x)$, the asymptotic MSE

146

in (6.9) can be alternatively expressed as

$$\sigma_p^2 = 1 - \frac{8\gamma_0 \mathrm{arctanh}\left(\sqrt{\frac{\frac{\alpha}{\delta}-2\gamma_0}{\frac{\alpha}{\delta}+2\gamma_0}}\right)}{\pi\sqrt{(\frac{\alpha}{\delta})^2 - (2\gamma_0)^2}}$$

$$\text{for } \frac{\alpha}{\pi} \leq \delta \leq \frac{\alpha}{2\gamma_0}, \gamma_0 < \frac{\pi}{2}. \tag{6.31}$$

The first derivative of σ_p^2 in (6.31) is $\frac{\partial \sigma_p^2}{\partial \delta} = f_1(\delta)f_3(\delta)$, where $f_3(\delta) = \delta\sqrt{(\frac{\alpha}{\delta})^2 - (2\gamma_0)^2} - 2\alpha \mathrm{arctanh}\left(\sqrt{\frac{\frac{\alpha}{\delta}-2\gamma_0}{\frac{\alpha}{\delta}+2\gamma_0}}\right)$.

We have $f_1(\delta) > 0$ when $\frac{\alpha}{\pi} \leq \delta < \frac{\alpha}{2\gamma_0}$. Since $f_3'(\delta) = \sqrt{(2\gamma_0)^2 - (\frac{\alpha}{\delta})^2} > 0$, $f_3(\delta)$ is an increasing

function in δ. Thus $f_3(\delta) < f_3(\frac{\alpha}{2\gamma_0}) = 0$. Therefore, $\frac{\partial \sigma_p^2}{\partial \delta} = f_1(\delta)f_3(\delta) < 0$.

Similarly, we can show that σ_p^2 is an increasing function in α, which is proportional to f_D.

6.9.3 Proof of Proposition 6.2

The Toeplitz matrix, $\mathbf{R}_{dh}$, is uniquely determined by the sequence $\mathbf{t}_{dh} = [t_{-N_p}, \cdots, t_0, \cdots, t_{N_p-2}]^T$,

where $t_k = \rho(kK + u) = J_0\left(2\pi f_\mathrm{D} T_p |k + \frac{u}{K}|\right)$. When $N_p \to \infty$, the DTFT of the sequence $\mathbf{t}_{dh}$ can be

calculated as

$$\Lambda_{dh}(\Omega) = \Lambda(\Omega) \times \exp\left(j\frac{u}{K}\Omega\right), \tag{6.32}$$

where $\Lambda(\Omega)$ is defined in (6.27).

Based on [12, Lemma 2], $\mathbf{R}_{dh}$ is asymptotically equivalent to a circulant matrix, $\mathbf{C}_{dh} = \mathbf{U}_N^H \mathbf{D}_{dh} \mathbf{U}_N$,

where $\mathbf{U}_N^H$ is the unitary discrete Fourier transform (DFT) matrix with the $(m+1, n+1)$-th element

being $(\mathbf{U}_N)_{m+1,n+1} = \frac{1}{\sqrt{N}}\exp\left[-j2\pi\frac{mn}{N}\right]$, and $\mathbf{D}_{dh}$ is a diagonal matrix with its k-th diagonal ele-

ment being $(\mathbf{D}_{dh})_{k,k} = \Lambda_{dh}\left(\frac{k-1}{N}\right)$.

Similarly, the Toeplitz matrix, $\mathbf{R}_{hh}$, is asymptotically equivalent to a circulant matrix, $\mathbf{C}_{hh} = $

$\mathbf{U}_N^H \mathbf{D}_{hh} \mathbf{U}_N$, where $\mathbf{D}_{hh}$ is a diagonal matrix with its k-th diagonal element being $(\mathbf{D}_{hh})_{kk} = \Lambda\left(\frac{k-1}{N}\right)$,

with $\Lambda(\Omega)$ defined in (6.27).

Based on [13, Theorem 2.1], the error correlation matrix, $\blacksquare_{ee}$, is asymptotically equivalent to a circulant matrix, $\mathbf{C}_{ee} = \mathbf{C}_{hh} - \mathbf{C}_{dh}\left(\mathbf{C}_{hh} + \frac{1}{\gamma_0}\mathbf{I}_{N_P}\right)^{-1}\mathbf{C}_{dh}^H = \mathbf{U}_N^H\mathbf{D}_{ee}\mathbf{U}_N$, where $\mathbf{D}_{ee} = \mathbf{D}_{hh} - \mathbf{D}_{dh}\left(\mathbf{D}_{hh} + \frac{1}{\gamma_0}\mathbf{I}_{N_P}\right)^{-1}\mathbf{D}_{dh}^H$. It is apparent that $\mathbf{D}_{ee}$ is diagonal given that $\mathbf{D}_{hh}$ and $\mathbf{D}_{dh}$ are diagonal.

Based on Szego's Theorem, we have

$$\sigma_e^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left[\Lambda(\Omega) - \frac{|\Lambda_{dh}(\Omega)|^2}{\Lambda(\Omega) + \frac{1}{\gamma_0}}\right]df. \tag{6.33}$$

Since $\delta \geq \frac{\alpha}{\pi}$, there is no spectrum aliasing for $\Lambda(\Omega)$ and $\Lambda_{dh}(\Omega)$ when $-\pi \leq \Omega \leq \pi$. As a result, $\Lambda_{dh}(\Omega)|^2 = |\Lambda(\Omega)|^2$ when $-\pi \leq \Omega \leq \pi$. Therefore (6.33) can be simplified to (6.25), and this completes the proof.

6.9.4   Proof of Proposition 6.3

Since $\mathbf{h}_p$ and $\mathbf{z}_p$ are zero mean Gaussian distributed, the received vector corresponding to the pilot symbols, $\mathbf{y}_p$, is zero mean Gaussian distributed with auto-correlation matrix $\mathbf{R}_{yy} = E_0\mathbf{P}\mathbf{R}_{hh}\mathbf{P}^H + \sigma_z^2\mathbf{I}_{N_P}$. From (6.12), the estimated channel vector $\hat{\mathbf{h}}_d$ is a linear transformation of $\mathbf{y}_p$, thus $\hat{\mathbf{h}}_d$ is zero mean Gaussian distributed with auto-correlation matrix given by

$$\mathbf{R}_{\hat{d}\hat{d}} = \mathbf{R}_{dh}\left(\mathbf{R}_{hh} + \frac{1}{\gamma_0}\mathbf{I}_{N_P}\right)^{-1}\mathbf{R}_{dh}^H. \tag{6.34}$$

Combining (6.13) with (6.34) yields $\mathbf{R}_{\hat{d}\hat{d}} = \mathbf{R}_{hh} - \Psi_{ee}$. Therefore, $\sigma_{\hat{h}}^2 = \frac{1}{N}\text{trace}(\mathbf{R}_{\hat{d}\hat{d}}) = 1 - \sigma_e^2$.

6.9.5   Proof of Corollary 6.2

Denote the estimation error vector $\mathbf{e}_d = \hat{\mathbf{h}}_d - \mathbf{h}_d$. Since both $\hat{\mathbf{h}}_d$ and $\mathbf{h}_d$ are zero-mean Gaussian distributed, $\mathbf{e}_d$ is zero-mean Gaussian distributed. The cross-covariance matrix between $\mathbf{e}_d$ and $\hat{\mathbf{h}}_d$ is $\mathbb{E}(\mathbf{e}_d\hat{\mathbf{h}}_d^H) = 0$ by following the orthogonal principal. Therefore, $\mathbf{e}_d$ and $\hat{\mathbf{h}}_d$ are uncorrelated.

The conditional mean can then be calculated as $\mathbf{u}_{h|\hat{h}} = \mathbb{E}(\mathbf{h}_d|\hat{\mathbf{h}}_d) = \hat{\mathbf{h}}_d - \mathbb{E}(\mathbf{e}_d|\hat{\mathbf{h}}_d) = \hat{\mathbf{h}}_d$. The auto-covariance matrix is, $\mathbb{E}[(\mathbf{h}_d - \mathbf{u}_{h|\hat{h}})(\mathbf{h}_d - \mathbf{u}_{h|\hat{h}})^H] = \mathbb{E}[(\mathbf{h}_d - \hat{\mathbf{h}}_d)(\mathbf{h}_d - \hat{\mathbf{h}}_d)^H] = \boldsymbol{\Psi}_{ee}$. The conditional variance is thus $\sigma_{h|\hat{h}}^2 = \frac{1}{N}\text{trace}\boldsymbol{\Psi}_{ee} = \sigma_e^2$.

6.9.6   Proof of Corollary 6.3

Since h conditioned on $\hat{h}$ is Gaussian distributed, it is straightforward that $y = \sqrt{E_0}hx + z$ conditioned on $\hat{h}$ and x is Gaussian distributed. The conditional mean and variance can be directly calculated by using the result from Corollary (6.2).

6.9.7   Proof of Proposition 6.4

Given the estimated CSI $\hat{h}$ and the transmitted symbol x, the conditional SER equals to the probability that the decision variable μ is outside of the decision region of x. Since μ conditioned on $\hat{h}$ and x is Gaussian distributed, the conditional error probability can be written as [4] and [15]

$$P(E|\hat{h}) = \frac{1}{\pi}\int_0^{\pi-\frac{\pi}{M}} \exp\left\{-\frac{|u_{\mu|x,\hat{h}}|^2 \sin^2(\frac{\pi}{M})}{\sigma_{u|x,\hat{h}}^2 \sin^2(\phi)}\right\} d\phi \qquad (6.35)$$

Substituting the values of $u_{\mu|x,\hat{h}}$ and $\sigma_{u|x,\hat{h}}^2$ from (6.17) into (6.35) yields

$$P(E|\hat{h}) = \frac{1}{\pi}\int_0^{\pi-\frac{\pi}{M}} \exp\left[-\frac{|\hat{h}|^2 \sin^2(\frac{\pi}{M})}{(\sigma_e^2 + \frac{1}{\gamma_0})\sin^2(\phi)}\right] d\phi \qquad (6.36)$$

The unconditional error probability $P(E) = \mathbb{E}[P(E|\hat{h})]$ can then be calculated by

$$P(E) = \frac{1}{\pi}\int_0^{\pi-\frac{\pi}{M}}\int_0^{+\infty} \exp\left[-\frac{\theta \sin^2(\frac{\pi}{M})}{(\sigma_e^2 + \frac{1}{\gamma_0})\sin^2(\phi)}\right] p_{|\hat{h}|^2}(\theta) d\theta d\phi,$$

where $p_{|\hat{h}|^2}(\theta)$ is the pdf of $|\hat{h}|^2$. From Proposition 6.3, $\hat{h} \sim \mathscr{CN}(0, 1 - \sigma_e^2)$, thus $|\hat{h}|^2$ is an exponentially distributed random variable with mean $1 - \sigma_e^2$. Changing the integration variable with $v = \frac{\alpha}{1-\sigma_e^2}$ and solving the integration of v results in (6.18).

6.9.8 Proof of Lemma 6.1

The conditional mutual information is defined as

$$I(y;x|\hat{h}) = \mathbb{E}_{x,y}\left[\log p(y|x,\hat{h})\right] - \mathbb{E}_{x,y}\left[\log p(y|\hat{h})\right]. \tag{6.37}$$

From Corollary 6.3, $p(y|x,\hat{h})$ is a Gaussian pdf with the conditional mean and variance given in (6.14). Then

$$\mathbb{E}_{x,y}\left[\log p(y|x,\hat{h})\right] = \mathbb{E}_x\left[\log \frac{1}{\pi(E_0\sigma_e^2|x|^2 + \sigma_z^2)e}\right]. \tag{6.38}$$

It can be easily shown that (6.38) is convex in $|x|^2$. Based on Jensen's inequality, we have

$$\mathbb{E}_x\left[\log \frac{1}{\pi(E_0\sigma_e^2|x|^2 + \sigma_z^2)e}\right] \geq \log \frac{1}{\pi(E_0\sigma_e^2 + \sigma_z^2)e}, \tag{6.39}$$

where $\mathbb{E}(|x|^2) = 1$ is used in the above inequality.

Define

$$I_{\text{low}}(y;x|\hat{h}) = \log \frac{1}{\pi(E_0\sigma_e^2 + \sigma_z^2)e} + \mathbb{E}\left[\log \frac{1}{p(y|\hat{h}_d)}\right]. \tag{6.40}$$

Thus $I_{\text{low}}(y;x|\hat{h}) \leq I(y;x|\hat{h})$.

The second term in (6.40) is the conditional differential entropy of y given $\hat{h}$. Conditioned on $\hat{h}$, the conditional mean and variance of y are given by

$$u_{y|\hat{h}} = 0, \tag{6.41a}$$

$$\sigma_{y|\hat{h}}^2 = E_0(|\hat{h}|^2 + \sigma_e^2) + \sigma_z^2. \tag{6.41b}$$

Given variance $\sigma_{y|\hat{h}}^2$, it is well known that the entropy of $y|\hat{h}$ is maximized if $y|\hat{h} \sim \mathcal{N}\left(0, \sigma_{y|\hat{h}}^2\right)$. In this case,

$$\max \mathbb{E}\left[\log \frac{1}{p(y|\hat{h}_d)}\right] = \log(E_0|\hat{h}|^2 + E_0\sigma_e^2 + \sigma_z^2) \tag{6.42}$$

Thus, (6.40) can be maximized by

$$C_{\text{low}}(\hat{h}) = \max I_{\text{low}}(y;x|\hat{h}) = \log\left(1 + |\hat{h}|^2 \frac{1}{\sigma_e^2 + \frac{1}{\gamma_0}}\right). \tag{6.43}$$

Since $I_{\text{low}}(y;x|\hat{h}) \leq I(y;x|\hat{h})$, we have $C_{\text{low}}(\hat{h}) \leq C(\hat{h})$.

6.9.9 Proof of Proposition 6.5

Based on (6.19) and (6.20), a lower bound on the effective spectral efficiency can be obtained as

$$\eta_{\text{low}} = (1 - \delta)\int_0^{+\infty} \log\left(1 + x\frac{1}{\sigma_e^2 + \frac{1}{\gamma_0}}\right) p_{|\hat{h}|^2}(x)dx, \tag{6.44}$$

where $p_{|\hat{h}|^2}(x)$ is the pdf of the exponentially distributed random variable $x = |\hat{h}|^2$ with mean $1 - \sigma_e^2$.

With the change of integration variable $v = \frac{x}{1-\sigma_e^2}$, (6.44) can be alternatively represented as

$$\eta = (1 - \delta)\int_0^\infty \exp(-v)\log\left(1 + v\frac{1 - \sigma_e^2}{\sigma_e^2 + \frac{1}{\gamma_0}}\right)dv \tag{6.45}$$

Solving the above integral based on the definition of the incomplete Gamma function yields (6.21).

6.10 Appendix of documentation of multi-authored chapter

UNIVERSITY OF ARKANSAS.

College of Engineering
Department of Electrical Engineering

July 31, 2013

To Whom It May Concern

This letter is to certify that Ms. Ning Sun, a Ph.D. candidate under my supervision at the Department of Electrical Engineering, has contributed more than 51% of the work for the following paper,

Ning Sun and Jingxian Wu, "Maximizing spectral efficiency for high mobility systems with imperfect channel state information," submitted to IEEE Transactions on Wireless Communications, 2013.

Ms. Sun is the first author of this paper. The paper is included in her Ph.D. dissertation entitled "Distortion-Tolerant Communications with Correlated Information".

Sincerely Yours,

Jingxian Wu
Director, Wireless Information Network Lab
Associate Professor of Electrical Engineering
University of Arkansas
Email: wuj@uark.edu
Tel: (479) 575-6584

3217 Bell Engineering Center • Fayetteville, AR 72701 • 479-575-3005 • Fax: 479-575-7967 • www.uark.edu
The University of Arkansas is an equal opportunity/affirmative action institution.

152

6.11 References

[1] K. M. Z. Islam, T. Y. Al-Naffouri, and N. Al-Dhahir, "On optimum pilot design for comb-type OFDM transmission over doubly-selective channels," *IEEE Trans. Commun.*, vol. 59, pp. 930-935, Apr. 2011.

[2] R. J. Baxley, J. E. Kleider, and G. T. Zhou, "Pilot design for OFDM with null edge subcarriers," *IEEE Trans. Wireless Commun.*, vol. 8, pp. 396-405, Jan. 2009.

[3] Z. Wu, J. He, and G. Gu, "Design of optimal pilot-tones for channel estimation in MIMO-OFDM systems," in *IEEE Wireless Commun. and Network. Conf.*, vol. 1, pp. 12-17, Mar. 2005.

[4] J. Wu and C. Xiao, "Optimal diversity combining based on linear estimation of Rician fading channels," *IEEE Trans. Commun.*, vol. 56, pp. 1612-1615, Oct. 2008.

[5] H. Sheng and A. M. Haimovich, "Impact of channel estimation on ultra-wideband system design," *IEEE J. Selected Topics in Signal Process.*, vol. 1, pp. 498-507, Oct. 2007.

[6] A. Vosoughi and Y. Jia, "How does channel estimation error affect average sum-rate in two-Way amplify-and-forward relay networks?" *IEEE Trans. Wireless Commun.*, vol. 11, pp. 1676-1687, May 2012.

[7] H. Gacanin, M. Salmela, and F. Adachi, "Performance analysis of analog network coding with imperfect channel estimation in a frequency-selective fading channel," *IEEE Trans. Wireless Commun.*, vol. 11, pp. 742-750, Feb. 2012.

[8] T. Whitworth, M. Ghogho, and D. McLernon, "Optimized training and basis expansion model parameters for doubly-selective channel estimation," *IEEE Trans. Wireless Commun.*, vol. 8, pp. 1490-1498, Mar. 2009.

[9] F. Qu and L. Yang, "On the estimation of doubly-selective fading channels," *IEEE Trans. Wireless Commun.*, vol. 9, pp. 1261-1265, Apr. 2010.

[10] J. Wu and N. Sun, "Optimum sensor density in distortion tolerant wireless sensor networks," *IEEE Trans. Wireless Commun.*, vol. 11, pp. 2056-2064, June 2012.

[11] I. S. Gradshteyn, and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 6nd ed. Academic Press, July 2000.

[12] H. Gazzah, P. A. Regalia, and J.P. Delmas, "Asymptotic eigenvalue distribution of block toeplitz matrices and application to blind SIMO channel identification," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1243-1251, Mar. 2001.

[13] R. M. Gray. Toeplitz and Circulant Matrices: a Review, NOW Publishers, 2006 (available online at http://ee.stanford.edu/~gray/).

[14] H. O. Beca, "An orthogonal set based on Bessel functions of the first kind," *Univ. Beograd. Public. Elektrotehnicki. Fak. Ser. Mat.*, pp. 85-90, Mar. 1980.

[15] J. Wu, C. Xiao, and N. C. Beaulieu, "Optimal diversity combining based on noisy channel estimation," in *Proc. IEEE Intern. Conf. Commun. ICC'04*, vol. 1, pp. 214-218, June 2004.

Chapter 7

Conclusions

This chapter summarizes the main contributions of this dissertation and lists some possible directions for the future research.

7.1 Contributions

The contents presented in this dissertation focus on the practical design of distortion-tolerant wireless communication systems by exploiting spatial and/or temporal correlation and the main contributions are summarized as follows.

First, we present the optimum sensor node density for 1-D and 2-D WSNs with spatial source correlation. The WSN is designed to minimize the MSE distortion between the original and the reconstructed signals under the constraint of a fixed power per unit area. It is observed that, for the network only needs to estimate spatially discrete data, placing exactly one sensor at the desired measurement locations will generate the optimum performance. For the estimation of the data at arbitrary locations in the measurement field, the optimum node density can be found when the MSE-density slope is close to zero.

Second, the analysis of the optimum sampling is extended into 1-D and 2-D WSNs with spatial-temporally correlated data. The impacts of the node density in the space domain, the sampling rate in the time domain, and the space-time data correlation on the network performance are investigated asymptotically by considering a large network with infinite area but finite node density and finite temporal sampling rate, under the constraint of fixed power per unit area. The impact of

155

space-time sampling on network performances is investigated in two cases. The first case studies the estimations of the space-time samples collected by the sensors, and the samples are discrete in both the space and time domains. The second case estimates an arbitrary data point on the space-time hyperplane by interpolating the discrete samples collected by the sensors. Optimum space-time sampling is obtained by minimizing the mean square error distortion at the network fusion center.

Third, a new DJSCC is proposed for a communication network with multiple correlated information sources. The DJSCC is performed by puncturing the information bits of a linear block code but leaving the parity bits intact. Unequal amounts of energy per bit is allocated to the information and parity bits. At the receiver, the sources are jointly decoded with the iterative message passing algorithm. Simulation results demonstrate that the proposed scheme can achieve considerable performance gains over conventional schemes.

Fourth, a new DJSCC coding scheme is designed for sources with spatially and temporally correlated data. At the receiver, two decoding algorithms are proposed. The fist MCMP decoding algorithm can perform both intra- and inter-codeword soft information exchange by using the spatial source correlation. The second one adds a hidden Markov model decoding module to the MCMP decoder to exploit the temporal data correlation. Both of these two decoding algorithms can lead to significant performance gains.

Finally, the optimum system design for high mobility distortion-tolerant wireless communication systems with imperfect CSI has been studied. The asymptotic channel estimation MSE has been quantified as a closed-form expression of the percentage of pilots used for MMSE channel estimation. Based on the statistical properties of the estimated channel coefficients, we derive the

explicit SER and a spectral efficiency lower bound of a communication system operating with imperfect CSI. It is discovered that the pilot samples the channel at a rate no less than the Nyquist rate of the time-varying channel, MMSE channel estimation at pilot locations or MMSE channel interpolation at non-pilot locations yield the same MSE. In addition, the SER is a monotonic decreasing function in pilot density, yet the spectral efficiency is concave in the pilot density.

7.2   Future Works

We listed several possible directions for the future works.

First, in the DJSCC proposed in Chapters 4 and 5, the simulation results show that the performance of the puncture operations is affected by the channel conditions. If the CSI is available at the transmitter, then the transmitter can adjust the puncturing rate to obtain extra performance gains. The CSI can be made available to the transmitter through a feedback channel. The design of adaptive DJSCC based on CSI is one of the future research directions worth exploring.

Second, in the design of distortion-tolerant high mobility systems introduced in Chapter 6, the role of the pilot is more important than data symbols, thus we can allocate unequal amounts of energy to the pilot and data symbols. Then the pilot information, *i.e.*, the pilot percentage and the energy allocation factor, can be jointly optimized to maximize the spectral efficiency or minimize the SER.

Finally, the channel model considered in Chapter 6 is time-varying flat fading. In reality, the channel in the high mobility case also suffers frequency selective fading, and this determines an operation environment of doubly-selective fading. The optimum designs of high mobility systems in doubly-selective fading with imperfect CSI is another important research topic.

Chapter 8

Vita

Ning Sun received the B.S. and M.S. degrees in Electrical Engineering from Shandong University of Science and Technology, Qingdao, China, in 2005 and 2008, respectively. From 2008 to 2009, she was a research assistant at the Underwater Acoustic Communication Institute, Soongsil University, Seoul, South Korea. She is currently a PhD candidate in the Department of Electrical Engineering, University of Arkansas, Fayetteville, USA. She is a student member of IEEE. Her research interests include wireless communications, wireless sensor networks and digital image processing.