ON MULTIPLE-ANTENNA COMMUNICATIONS: SIGNAL DETECTION, ERROR EXPONENT AND QUALITY OF SERVICE

A Dissertation

by

QIANG LI

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2007

Major Subject: Electrical Engineering

ON MULTIPLE-ANTENNA COMMUNICATIONS: SIGNAL DETECTION, ERROR EXPONENT AND QUALITY OF SERVICE

A Dissertation

by

QIANG LI

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Costas N. Georghiades
Committee Members,	Krishna R. Narayanan
	Takis Zourntos
	Prabir Daripa
Head of Department,	Costas N. Georghiades

December 2007

Major Subject: Electrical Engineering

ABSTRACT

On Multiple-Antenna Communications: Signal Detection,
Error Exponent and Quality of Service. (December 2007)
Qiang Li, B.E., Shanghai Jiaotong University;
M.E., Shanghai Jiaotong University
Chair of Advisory Committee: Dr. Costas N. Georghiades

Motivated by the demand of increasing data rate in wireless communication, multiple-antenna communication is becoming a key technology in the next generation wireless system. This dissertation considers three different aspects of multiple-antenna communication.

The first part is signal detection in the multiple-input multiple-output (MIMO) communication. Some low complexity near optimal detectors are designed based on an improved version of Bell Laboratories Layered Space-Time (BLAST) architecture detection and an iterative space alternating generalized expectation-maximization (SAGE) algorithm. The proposed algorithms can almost achieve the performance of optimal maximum likelihood detection. Signal detections without channel knowledge (noncoherent) and with co-channel interference are also investigated. Novel solutions are proposed with near optimal performance.

Secondly, the error exponent of the distributed multiple-antenna communication (relay) in the windband regime is computed. Optimal power allocation between the source and relay node, and geometrical relay node placement are investigated based on the error exponent analysis.

Lastly, the quality of service (QoS) of MIMO/single-input single- output(SISO) communication is studied. The tradeoff of the end-to-end distortion and transmission buffer delay is derived. Also, the SNR exponent of the distortion is computed for

MIMO communication, which can provide some insights of the interplay among time diversity, space diversity and the spatial multiplex gain.

To Mom and Dad

ACKNOWLEDGMENTS

First and foremost, I would deeply like to thank my doctoral advisor Prof. Georghiades. Five years ago, I was almost a blank sheet in wireless communication. He has taught me everything from conducting research to technical writing. He always encouraged me to develop a broader view of my research that included more than just the technical details. He is very knowledgeable and insightful. His patience and understanding helped me to pass the hardest time during my Ph.D study. He has also provided me generous financial support for my research.

I am very thankful to Prof. Krishna R. Narayanan for his excellent coding courses. I learned a lot from these courses, and they are very beneficial to my research. Also, the idea for last part of this dissertation is inspired from his research work. I would like to express my gratitude to Prof. Daripa and Prof. Zourntos for serving on my doctoral committee. I would also give thanks for Prof. Miller for his enlightening wireless communication courses. I am grateful to Dr. Yongzhe Xie for helping in the early stage of my Ph.D research and for his contribution to the first chapter of this dissertation. He is a very smart and nice guy, I enjoyed discussing with him. My thanks also go to my colleagues and friends in the Intel communication technology lab: Jing Zhu, Guoqing Li, Qinghua Li, Chunmei Liu and Palaskas Yorgos. During my six months stay in the Intel lab, I learned a lot from them and had a lot of fun.

I thank all the members of the wireless communication lab. In particular, I would like to thank Jia Tang, Jing Jiang, Lingjia Liu, Guoshen Yue, Wenyan He, Yang Yang, Janath Peiris, Hari Sankar, Salim El-Rouayheb and Nitin Nangare for numerous technical discussion.

I am indebted to my parents for their never-ending love and support. I dedicate this work to them.

TABLE OF CONTENTS

CHAPTER Page

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

FIGURE Page

FIGURE Page

FIGURE Page

CHAPTER I

INTRODUCTION

The first and second generation wireless communication systems focus on voice communication. In recent years, the quality and data rate provided by the wireless system are the most important requirements as data applications become more and more popular, e.g., video, multimedia. Multiple antennas are an important means to meet this challenge. It is well understood in systems with multiple transmit and receive antennas (MIMO systems), the capacity increases linearly with the number of antennas [1] [1].

The gain of multiple antennas is two-fold. First, through space diversity to combat wireless fading which is the key challenge to wireless communication. Essentially, each pair of transmit and receive antennas provides a signal path from transmitter to receiver. By sending signals that carry the correlate information through a number of different paths, multiple independent faded replicas of the data symbol can be obtained at the receiver end. By averaging these replicas, more reliable data reception can be obtained. The second factor is the multiplexing gain. Sending independent data streams through different signal paths by appropriately exploiting the so-called "degree of freedom," can achieve a much higher data rate of communication. There is a fundamental tradeoff between these two gains [2, 3].

Unlike conventional point-to-point communication, a wireless network, the overall throughput is interference limited. Multiple antennas provide the extra degrees of freedom to suppress the co-channel interference and effectively detect the intended sig-

---

The journal model is *IEEE Transactions on Automatic Control.*

[1] More accurately, it increases with the minimum of the transmit and receive antenna number.

nal and hence increase the spectrum efficiency of the system. Moreover, one evolution of multiple-antenna systems distributes the antennas in different mobile terminals to relax the physical size limitation of the communication device. Hence this creates the so-called "relay communication" where the relay nodes help the transmission from the source to destinations [4]. This distributed antenna relay communication poses some challenges for the transmission protocol design and fundamental limit analysis. Relay transmission has been adopted in some wideband wireless communication standards, e.g., IEEE 802.16j.

For a communication system, quality of service (QoS) is very important for most data communication applications. End-to-End distortion and transmission delay are two fundamental QoS metrics. Usually, the source is continuous amplitude and needs to be digitalized and transmitted over the wireless channel. Also, there should be a buffer to store the quantized bits before transmission. Therefore, for such a communication system with multiple antennas, analyzing the distortion and delay play an important role for future generation wireless system design.

This dissertation sets two goals in the framework of multiple-antenna communication systems. The first is physical layer signal processing: signal detection and interference suppression. The second is fundamental limit analysis for multiple-antenna systems. More specially, this dissertation studies the error exponent of the wideband relay channel and the distortion-delay tradeoff for an analogue source transmitted over the MIMO/SISO fading channel.

A. Dissertation Outline

The second chapter introduces the background of MIMO communication. The focus is the channel model, capacity results and signal processing in the transmitter and

receiver ends.

Chapter III introduces several low complexity sub-optimal MIMO detection schemes based on the List-BLAST algorithm which exhausts the constellation points in the first layer of a BLAST scheme to generate multiple candidate solutions from which the maximum likelihood solution is determined. The candidates can also be used as initial points for the space alternating generalized expectation-maximization (SAGE) algorithm to further improve performance. The proposed schemes can achieve close to optimal performance for both hard and soft output detection with lower complexity than that of sphere detection in our simulation settings.

Chapter IV studied decoding orthogonal space-time block codes without channel state information at the receiver. We used the inherent structure of these codes and adapted the sphere decoding method previously used under perfect channel knowledge to efficiently detect transmitted symbols. The resulting performance was within 1.5-dB of that of coherent detection and the proposed method had low average complexity in medium and high SNR regions.

In Chapter V, we presented algorithms to suppress the asynchronous co-channel interference (CCI) in MIMO OFDM systems; this is becoming the dominant limiting factor in the performance of emerging high-density WLANs. The key challenge is that the cyclic prefix of the interference signal does not line up with that of the intended signal due to asynchronous transmission in WLAN. Therefore, the orthogonality among the different tones of the interference signal is destroyed and conventional frequency domain minimum mean square error (MMSE) cancelation techniques that estimate the interference channel response for each tone cannot work effectively. To suppress the asynchronous interference, we designed an efficient estimator to measure the interference spatial covariance matrix using Cholesky decomposition and low-pass smoothing. Both an MMSE and a maximum a posteriori (MAP) receiver were de-

rived based on estimated interference statistics. Simulation results demonstrated the effectiveness of our solution.

Chapter VI investigated the error exponent of the wideband relay channel. By computing the random coding error exponent of three different relay strategies, i.e., amplify-and-forward (AF), decode-and-forward (DF) and block Markov code (BMC), we found that relayed transmission can enhance the wireless link reliability significantly in the wideband regime compared to direct transmission. We also studied optimal power allocation and relay placement by maximizing the reliability function. For DF and BMC relays, analytical and numerical results show that placing the relay node in the middle of source and destination provides the best link reliability. But for the AF relay scheme, the optimal relay placement depends on the path-loss exponent; for large path-loss exponents, half-way relay placement is also optimal.

Chapter VII examined the end-to-end distortion/delay tradeoff for an analogue source transmitted over a fading channel. The analogue source was quantized and stored in a buffer until it was transmitted. There are two extreme cases as far as buffer delay is concerned: no delay and infinite delay. We observed that there was a significant power gain by introducing a buffer delay. Our goal was to investigate the situation between these two extremes. Using the recently proposed effective capacity concept, we derived a closed-form formula for this tradeoff For the single-input single-output (SISO) case, an asymptotically tight upper bound for our distortion-delay curve was derived, which approached the infinite delay lower bound as $\mathcal{D}_\infty \exp(\frac{C}{\tau_n})$, with τ_n the normalized delay and C a constant. For the more general MIMO channel, we computed the distortion SNR exponent - the exponential decay rate of the expected distortion in the high SNR regime. Numerical results demonstrated that the introduction of a small amount of delay can save significant transmission power.

Finally, Chapter VIII concludes the dissertation and summarizes the new results

described in the dissertation.

B. A Note on Notation

Throughout this paper, normal letters indicate scalar quantities and boldface fonts denote matrices and vectors. For any matrix $\mathbf{M}$ we wrote its transpose as $\mathbf{M}^T$ and $\mathbf{M}^H$ as its conjugate transpose. x^* denotes the conjugate of x. $\mathbf{M}^{-1}$, $tr(\mathbf{M})$ and $\det(\mathbf{M})$ denote the inverse, trace and determinant of matrix $\mathbf{M}$, respectively; $\mathbf{I}$ denotes the identity matrix; $\mathbf{M}[i,j]$ denotes the $[i,j]^{th}$ entry of the matrix $\mathbf{M}$; x_i denotes the i^{th} element of the vector x; $\|\mathbf{M}\|_2$ and $\|\mathbf{M}\|_F$ denotes the L2 and Frobenius norm of the matrix and $\text{vec}(\mathbf{M})$ represents matrix vectorization by stack columns of $\mathbf{M}$. $\ln(\cdot)$ and $\log(\cdot)$ represents the natural and 2 based logarithm.

CHAPTER II

A BRIEF OVERVIEW OF MULTIPLE-ANTENNA SYSTEMS

This chapter first introduced the system model and the information limits of MIMO channels. Then briefly overviewed the transmitter techniques and receiver processing for MIMO communication. Finally, we introduced MIMO combined with orthogonal frequency-division multiplexing (OFDM) to provide high spectral efficiency for wideband wireless communication.

A. Channel and System Model

The channel model is first described. Fig.1 shows a communications link with M_t transmit antennas and M_r receive antennas. At each time instant, M_t signals, $[x_1, x_2, \cdots, x_{M_t}]$, satisfying an average power constraint, are transmitted using M_t antennas. Each of them reaches all M_r receive antennas. Mathematically, the channel model can be expressed as:

$$\mathbf{y} = \sqrt{\frac{\rho}{M_t}}\mathbf{H}\mathbf{x} + \mathbf{w} \ , \tag{2.1}$$

Where $\mathbf{H}$ is the channel matrix containing i.i.d. elements $h_{i,j} \sim \mathcal{CN}(0,1)$ (Rayleigh independent amplitude fading). $\mathbf{x}$ is the transmitted signal, the power of transmitted signal $\mathbf{x}$ is normalized so that is satisfies $\mathrm{tr}(\mathbb{E}[\mathbf{x}^H\mathbf{x}]) \leq M_t$. ρ denotes the signal-to-noise ratio (SNR), defined as the ratio of the average received signal energy per receiving antenna to the noise per-component variance. $\mathbf{w}$ is the complex additive Gaussian noise with i.i.d. entries $\mathcal{CN}(0,1)$. We defined $M_* = min(M_t, M_r)$ and $M^* = max(M_t, M_r)$.

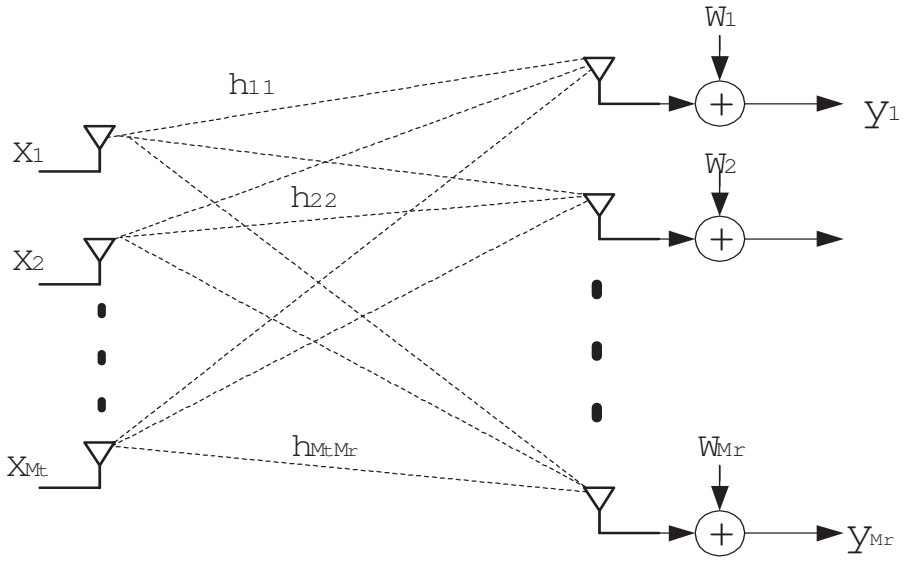

Fig. 1. Multiple-antennal channel

B. Information Limits of MIMO Channel

The shannon capacity is the maximum achievable rate of a channel with zero error probability. For infinite block length, the Shannon capacity is not defined since for any data rate there is a strict possibility that the the channel matrix $\mathbf{H}$ is unable to support it. Therefore, we can classify the capacities as ergodic capacity - averaged over long block length; and outage capacity - the maximum rate one can communicate with no more than a defined probability ϵ of channel outage. The mutual information of MIMO channel can be written as:

$$\mathbf{I}(\mathbf{x};\mathbf{y}|\mathbf{H}) = \log\det(\mathbf{I}_{M_r} + \mathbf{H}\mathbf{Q}\mathbf{H}^H) \ , \tag{2.2}$$

where $\mathbf{Q} = \mathbb{E}(\mathbf{x}\mathbf{x}^H)$ is the input signal covariance.

1. Ergodic Capacity

The ergodic capacity can be expressed as the average mutual information of all the channel realizations. The MIMO ergodic capacity depends on channel state information (CSI). In this section, we assumed the receiver has full CSI and we discussed the ergodic capacity with/without CSI at the transmitter. For an unknown channel at the receiver side (noncoherent) case, [5, 6, 7] has investigated the information limits.

a. CSI Perfectly Known Only at the Receiver

For this scenario, An independent (across transmit antennas) Gaussian input will achieve the capacity, i.e., $\mathbf{Q} = \frac{\rho}{M_t}\mathbf{I}_{M_r}$.

$$\begin{aligned}\mathbf{C} &= \mathbb{E}[\log\det(\mathbf{I}_{M_r} + \frac{\rho}{M_t}\mathbf{H}\mathbf{H}^H)] \\ &= \mathbb{E}[\log\det(\mathbf{I}_{M_t} + \frac{\rho}{M_t}\mathbf{H}^H\mathbf{H})] \ . \end{aligned} \tag{2.3}$$

A lower bound can be derived:

$$\mathbf{C} \geq M_* \log_2 \frac{\rho}{M_t} + \sum_{i=M^*-M_*+1}^{M^*} \mathbb{E}[\log_2 \chi^2_{2i}] \ , \tag{2.4}$$

where χ^2_{2i} is chi-square random variable with dimension of $2i$. Moreover, this lower bound is asymptotically tight at high SNR. We observe that this is equivalent to M_* parallelled sub-channels. In other words, MIMO has M_* degree of freedom to communication.

b. CSI Perfectly Known at Both the Transmitter and the Receiver

If the transmitter knows the channel realization, Let the SVD decomposition $\mathbf{H} = \mathbf{U}\mathbf{D}\mathbf{V}^H$, where $\mathbf{U}$ and $\mathbf{V}$ is the orthogonal matrix and $\mathbf{D}$ is a diagonal matrix. Since the $\mathbf{U}$ and $\mathbf{V}$ are known at both transmitter and receiver then we can precode (mul-

tiply) the transmit signal by $\mathbf{V}$ and post-filtering (multiply) the signal by $\mathbf{U}^H$ to transform the MIMO channel into M_* equivalent parallel channel. We denote these paralleled channel as eigen-channels. The channel capacity can be achieved by power water-filling. The water-filling gain is a power gain hence it is more significant at low SNR.

2. Outage Capacity

When the channel is slow fading the ergodicity does not hold, i.e., the codeword is no longer enough to average a large number of channel realizations. We can treat the mutual information as a random variable and assume the channel is quasi-static (remain constant for one block and change independently from block to block). The outage probability is given by

$$p_{out}(R) = \mathbb{P}(\mathbf{I}(\mathbf{x};\mathbf{y}) < R) \ . \tag{2.5}$$

Since the outage probability is a monotonically non-decreasing function of R. The outage capacity is defined as the the suprimum of the transmission rate that the outage probability is below some predefined value ϵ, and denoted as C_ϵ.

C. Transmission Techniques for MIMO Systems

In this section, we reviewed several classical transmission techniques for MIMO communication. This is by no means a complete list and which transmission technique to be used depends on the system settings and requirements .

1. Without CSI in the Transmitter

a. Spatial Multiplexing

When the transmitter does not know the channel, the independent Gaussian input will achieve the ergodic capacity. Transmitting independent data streams from the different transmit antennas spatial multiplex transmission can almost achieve the ergodic MIMO capacity although the receiver processing can be very complicated [8]. The spatial multiplexing transmission is sometimes called V-BLAST transmission. For the slow fading scenario, carefully designed signaling and coding (universal) are needed to achieve outage capacity. Design universal coding to achieve outage capacity is still an active research area. Moreover, for slow fading we need to consider two gains: space diversity and multiplex gain. The proposed space-time code is dedicated to utilize space diversity.

b. Space-time Coding

Tarokh et. al proposed the space-time code to provide transmitter diversity. The basic idea is to provide the transmitted signal redundancy and structure to protect the information from the fading detriment. The starting point was to minimize the pairwise error probability of two codewords. The conclusion was a design rule of rank criteria to maximize the diversity order and the determinant criteria to maximize the code gain. Based on these criteria, the manual designed space-time trellis codes were proposed to achieve $2-3$ dB from the outage capacity. However, the decoding of the space-time trellis code required a maximum-likelihood (ML) Viterbi decoding algorithm whose complexity increases exponentially as the trellis state number.

To reduce the decoding complexity, the orthogonal space-time block code (OS-TBC) is proposed, with a very simple - linear ML decoding scheme [9]. Due to the

orthogonality of the signals emanating from the different transmitter antennas, the inter-stream interference was avoided and the signals could easily be decoupled by a linear processing while still providing space diversity. The well-known Alamouti code [10] belongs to this family of codes with two transmit antennas. Normally, the linear OSTBC ML decoding requires channel knowledge at the receiver side. In Chapter IV, we have proposed a noncoherent decoder without CSI at both sides. Most OSTBC transmissions have some capacity loss and therefore suboptimal with regard to the information limit, except the 2×1 Alamouti code [11].

2. Full CSI at Both Transmitter and Receiver

When channel is known at both transmitter and receiver, the SVD can be used to decouple the MIMO channel into parallelled channels and water-filling can be used to achieve the capacity.

a. Transmitter Beamforming

When the water-filling gain is marginal compared with using only the coding and interleaving[1] and the transmitter antenna number is greater than the receiver antenna number then uniform power (equal power for each data stream) beamforming is always used. The basic idea is to transmit the information in the principal eigen-directions (the eigen-channels with the largest gain). The beamforming can be realized by pre-multiply the transmitted signal with the first several columns of $\mathbf{V}$ corresponding to the largest eigenvalues. When the transmitter antenna number is greater than the receiver antenna numbers (this is a typical setting for downlink transmission), the gain can be significant. Again, the beamforming gain is also a power gain.

[1]This is the case for most MIMO OFDM systems that use spatial-frequency interleaver and coding.

For time division duplex (TDD) systems, the CSI can be estimated at the transmitter by channel reciprocity. For frequency division duplex (FDD) systems, the channel is usually estimated at the receiver and beamforming matrix $\mathbf{V}$ has to fedback to the transmitter via a capacity limited link. The feedback overhead can be very large for some wideband MIMO-OFDM systems. Hence to efficiently transfer the CSI from the receiver to transmitter requires a carefully design.

D. Classical Detection Scheme

In this section, we review some classical signal processings at the receiver side. For OSTBC and beamforming transmission, usually an equivalent channel can be formalized, e. g., let the precode matrix of beamforming be $\mathbf{F}$, we can consider the matrix $\tilde{\mathbf{H}} = \mathbf{FH}$ as the equivalent channel. Hence, with regard to the receiver design, there is no difference from the spatial multiplex transmission by letting $\tilde{\mathbf{H}} = \mathbf{H}$. Therefore, without loss of generality, we can introduce MIMO signal detection techniques based on spatial multiplex transmission. We used the symbol-error rate for the uncoded system as the measure to compare the performance of different detectors.

1. Linear Detection

Assume $M_r \geq M_t$, the linear detector takes the received vector $\mathbf{y}$ and premultiplies it by a matrix $\mathbf{B}^H$. The resulting product $\tilde{\mathbf{x}}$ is passed to the minimum distance symbol by symbol decision. The matrix $\mathbf{B}$ can be optimized by different criteria. Two of the most popular schemes are zero-forcing (ZF), $\mathbf{B} = \sqrt{\frac{M_t}{\rho}}\mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}$; minimum mean square receiver (MMSE), $\mathbf{B} = (\mathbf{HH}^H + \frac{M_t}{\rho})^{-1}\sqrt{\frac{M_t}{\rho}}\mathbf{H} = \sqrt{\frac{M_t}{\rho}}\mathbf{H}(\mathbf{H}^H\mathbf{H} + \frac{M_t}{\rho})^{-1}$. The above equivalence of the two forms of the MMSE detector can be proved by the matrix inversion Lemma. The ZF detection chooses the pre-filter $\mathbf{B}$ to totally eliminate

inter-stream interference in $\tilde{\mathbf{x}}$, while the MMSE criteria chooses $\mathbf{B}$ to minimizing the variance of the error $\mathbf{e} = \mathbf{x} - \tilde{\mathbf{x}}$. When the SNR goes to large, the MMSE receiver will converges to the ZF receiver, as the interference dominates the noise in the high SNR.

The main disadvantage of linear detectors is poor performance in a symbol-error-rate (SER) sense. Since linear pre-filtering $\mathbf{B}$ makes the noise correlated, symbol-by-symbol detection, although very simple, is not optimal. From the more fundamental receiver diversity angle, ZF and MMSE have only diversity $M_r - M_t + 1$, compared to the diversity order of M_r for the ML detector. Hence, for a square matrix $\mathbf{H}$, the diversity order of such linear detectors is only 1. Intuitively, the ZF project the received signal in the null space of the $\mathbf{H}_i$ matrix, where $\mathbf{H}_i$ is the $\mathbf{H}$ matrix to remove the i^{th} column. The dimension of this null space is $M_r - M_t + 1$. Therefore, the resulted diversity order is $M_r - M_t + 1$ for the ZF detector. Since MMSE will converge to ZF in high SNR, MMSE has the same receiver diversity as ZF.

2. Decision Feedback Detection

Decision feedback detection, also called BLAST (nulling and canceling) is built on the linear detector by adding a feedback loop. Instead of making the decision concurrently for all of filtered output $\tilde{\mathbf{x}}$, decisions are made sequentially. One component a time. At the beginning, the detector first make decision on the first symbol of $\tilde{\mathbf{x}}$, denoted by $\hat{x}_1$, then the feedback loop is used to subtract the interference caused by x_1 from the remaining components of $\mathbf{B}^H\mathbf{r}$. Assuming that $\hat{x}_1$ is a correct decision, the process continues until all the components have been detected. Due to error propagation, the first symbol detection will dominate the vector error rate. Hence, usually the detected order is from the highest SNR information symbol to the lowest SNR to minimize error propagation. Since the performance is dominated by the first layer, the receiver

diversity order of this V-BLAST detection is the the same as the linear detection $(M_r - M_t + 1)$ even with the optimal detection order. In this dissertation, we have proposed a modified BLAST detection, called "Ordered List BLAST" to improve the performance of BLAST detection.

3. Maximum Likely Detection (Sphere Decoding)

When we assume all the vectors $\mathbf{x}$ are equally likely, the detector that maximizes the probability:

$$\tilde{\mathbf{x}} = \arg\max_{\Omega^{M_t}} f(\mathbf{r}|\mathbf{x}) \tag{2.6}$$

is optimal, where Ω^{M_t} denotes the set of constellation points in the complex M_t-dimenstion space. Assuming the noise is independent of $\mathbf{x}$ and i.i.d. Gaussian, the maximum likelihood (ML) detector is simplified to the minimum distance detector

$$\tilde{\mathbf{x}} = \arg\min_{\mathbf{x}\in\Omega^{M_t}} \|\mathbf{y} - \frac{\rho}{M_t}\mathbf{H}\mathbf{x}\|_2 \ , \tag{2.7}$$

where $\mathbf{H}$ is perfectly known. Since the transmitted signal is from the uncoded QAM or QPSK symbols. the optimization of (2.7) is an integer programming problem and NP-hard. Exhaustive search has exponential complexity and is practically impossible to implement. Sphere decoding (SD) can be used to reduce the complexity [12].

Let's Assume the QAM modulation has been used with the constellation size Q^2. We then transform the channel matrix as:

$$\mathbf{B} = \sqrt{\frac{12\rho}{M_t(Q^2-1)}} \begin{bmatrix} \mathrm{Re}\{\mathbf{H}\} & -\mathrm{Im}\{\mathbf{H}\} \\ \mathrm{Im}\{\mathbf{H}\} & \mathrm{Re}\{\mathbf{H}\} \end{bmatrix} \tag{2.8}$$

and let $\mathbf{y}_r = [\mathrm{Re}\{\mathbf{y}\} \quad \mathrm{Im}\{\mathbf{y}\}]^T$. Then the ML detector has following form

$$\tilde{\mathbf{x}} = \arg \min_{\mathbf{s} \in \mathbb{Z}_Q^{2M_t}} \|\mathbf{y}_r - \mathbf{B}\mathbf{s}\|_2 \ , \tag{2.9}$$

where $\mathbb{Z}_Q \triangleq \{0, 1, \cdots, Q-1\}$. We can consider $\mathbf{B}$ as a lattice generate matrix. Hence, the optimization problem is reduced to a closest lattice point search algorithm. Applying the QR decomposition to $\mathbf{B}$, we have

$$\mathbf{B} = \begin{bmatrix} \mathbf{Q} & \mathbf{Q}' \end{bmatrix} \begin{bmatrix} \mathbf{R} \\ 0 \end{bmatrix} \ , \tag{2.10}$$

where $\mathbf{R}$ is an $M_t \times M_t$ upper triangle matrix with positive diagonal elements, and $\mathbf{Q}$ (resp. $\mathbf{Q}'$) is an $M_r \times M_t$ (resp. $M_r \times (M_r - M_t)$) unitary matrix. Predefined a hypersphere $\mathcal{S}(\mathbf{y}_r, \sqrt{r_0})$ centered on the received signal , which is large enough to be included inside the optimal point with minimum Euclidean distance. Let the initial sphere radius be r_0. Therefore, the condition of the lattice points lies in the hypershpere, i.e., $\mathbf{B}\mathbf{x} \in \mathcal{S}(\mathbf{y}_r, \sqrt{r_0})$ can be written as

$$\begin{aligned} \|\mathbf{y}_r - \mathbf{B}\mathbf{x}\|^2 &\le r_0 \\ \left\| \begin{bmatrix} \mathbf{Q} & \mathbf{Q}' \end{bmatrix}^T \mathbf{y}_r - \begin{bmatrix} \mathbf{R} \\ 0 \end{bmatrix} \mathbf{x} \right\| &\le r_0 \\ \|\mathbf{Q}^T\mathbf{y}_r - \mathbf{R}\mathbf{x}\|^2 &\le r_0 - \|(\mathbf{Q}')^T\mathbf{y}_r\|^2 \\ \|\mathbf{y}' - \mathbf{R}\mathbf{x}\|^2 &\le r_0' \ , \end{aligned} \tag{2.11}$$

where $\mathbf{y}' \triangleq \mathbf{Q}^T\mathbf{y}_r$ and $r_0' \triangleq r_0 - \|(\mathbf{Q}')^T\mathbf{y}_r\|^2$. Due to the upper triangular form of $\mathbf{R}$, the last inequality implies series of conditions

$$\sum_{j=i}^{M_t} \|y_j' - \sum_{k=j}^{M_t} r_{j,k} x_k\|^2 \le r_0' \ , \quad i = 1, \cdots, M_t \ . \tag{2.12}$$

Therefore, given the last $M_t - i$ component values $\mathbf{x}_k^{M_t} \triangleq [x_{i+1}, \cdots, x_{M_t}]$, an upper and lower bound of x_i can be derived from the set of inequalities:

$$U_i(\mathbf{x}_{i+1}^{M_t}) = \left\lfloor \frac{1}{r_{i,i}} \left(y_i' - \sum_{j=i+1}^{M_t} r_{i,j} x_j - \sqrt{r_0' - \sum_{j=i+1}^{M_t} \left| y_j' - \sum_{k=j}^{M_t} r_{j,k} x_k \right|^2} \right) \right\rfloor$$

$$L_i(\mathbf{x}_{i+1}^{M_t}) = \left\lceil \frac{1}{r_{i,i}} \left(y_i' - \sum_{j=i+1}^{M_t} r_{i,j} x_j + \sqrt{r_0' - \sum_{j=i+1}^{M_t} \left| y_j' - \sum_{k=j}^{M_t} r_{j,k} x_k \right|^2} \right) \right\rceil \quad . \tag{2.13}$$

If

$$\sum_{j=i+1}^{M_t} \| y_j' - \sum_{k=j}^{M_t} r_{j,k} x_k \|^2 \geq r_0' \ , \tag{2.14}$$

then there is no value of x_i satisfying the inequality (2.12) and all the lattice points corresponding this choice of $\mathbf{x}_{i+1}^{M_t}$ do not belong to the sphere $\mathcal{S}(\mathbf{y}, s\sqrt{r_0})$, and can be pruned from the search space.

The search is started from the last layer and sequentially goes to the first layer as shown in Fig. 2. If the search engine arrives at the first layer, i.e., all the inequalities (2.12) are satisfied and a valid lattice point within the sphere has been computed. We then shrink the sphere radius to the distance of the found lattice point to the received signal. This radius update is very important to reduce the search complexity. The process is repeated until only the optimal point is left in the sphere. Essentially, sphere decoding is a depth first tree search process or branch-and-bound technique in the dynamic programming. It is a powerful method to solve the discrete ML optimal point search problem. Like the Viterbi ML search utilizes the finite state machine (FSM) trellis, the SD search reduces the complexity by the upper triangle structure of the lattice generating matrix.

The beauty of the SD algorithm lies in its approximated polynomial complexities for typical MIMO communication settings [13, 14]. In fact, for most case its expected

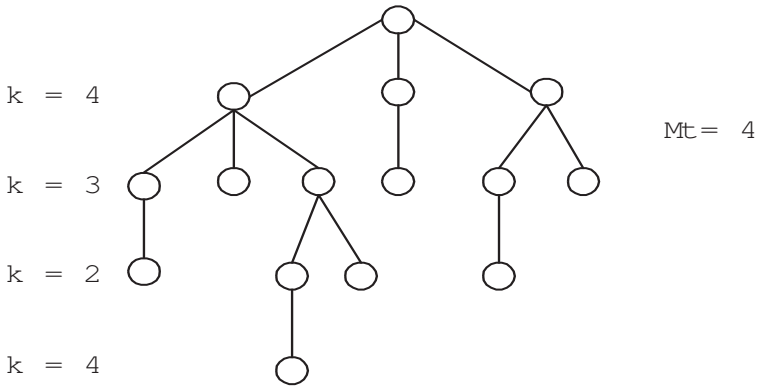

Fig. 2. Sample of the SD tree search in 4-dimensional hypersphere

search operation is around M_t^3. This low complexity is due to the fact that the received signal $\mathbf{y}$ is the transmitted lattice point perturbed by a Gaussian noise. Due to the the statistical distribution of channel matrix $\mathbf{H}$ and noise $\mathbf{W}$, SD is an efficient approximated polynomial complexity search algorithm for MIMO detection. For more general settings, [15] shows the expected search step can be expressed as $Q^{\gamma M_t}$, where $\gamma \in (0, 1]$ depending on the SNR value. Hence, strictly speaking, the algorithm is not polynomial. However, for large SNR, the factor $\gamma << 1$. This mean the complexity of SD is dominated by the polynomial term. As the SNR decreases, the γ increases. Therefore, SD has a lower complexity at high SNR than when operating at the low SNR.

4. Lattice Reduction Detection

Fig. 3 plot the decision boundary of different detectors for the 2×2 MIMO system. The red lines correspond to the two column vectors $[\mathbf{h}_1\ \mathbf{h}_2]$ of the channel matrix which are not orthogonal due to their random distribution. The red circle denotes the correct decision region. The larger the circle, the more noise power can be tol-

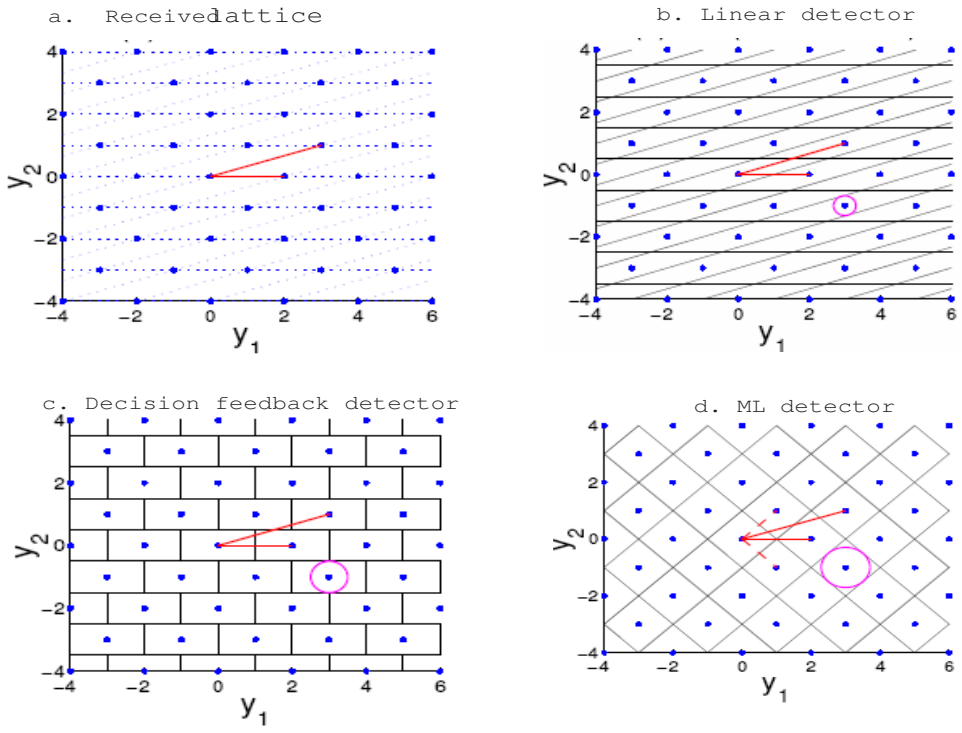

Fig. 3. A performance comparison of different detectors Yao'03

erated. Obviously, the ML has a large circle or better performance than the other detectors. To get the performance of the linear detector to achieve that of the ML detectors, one valid method is to make the columns of channel matrix H as orthogonal as possible. Hence some lattice reduction algorithms, e.g., Lenstra-Lenstra-Lovsz(LLL) algorithm, can be used to preprocess the channel matrix [16]. After the lattice reduction, the resulting matrix has quasi-orthogonal columns. Therefore, low complexity linear detectors can nearly achieve the ML performance.

E. Information Theory Aspect of MIMO Receiver

The previous comparison of MIMO detectors is based on the uncoded symbol error rate. The suboptimality of linear detection and decision feedback detection is due to the linear equalizer introduced correlation of noise across the antennas. Hence, uniform symbol-by-symbol quantization of equalizer output $\tilde{\mathbf{x}}$ is suboptimal and causes significant information loss. With respect to information theoretical comparison, the ML receiver is information lossless; hence, the sphere decoder algorithm can be easily extended to a "list sphere decodind" to generate the soft information. [8] Showing the list sphere decoding joint with turbo code can achieve near ergodic MIMO capacity though the complexity of such an ML receiver is very high.

The linear MMSE detector is seriously suboptimal in regard to the symbol error rate, however, the MMSE itself is information lossless. The output of MMSE equalizer is a sufficient statistic to detect x, i.e.,

$$I(\mathbf{x};\mathbf{y}) = I(\mathbf{x};\mathbf{W}_{MMSE}\mathbf{y}) \tag{2.15}$$

Therefore, according the chain rule of mutual information

$$\begin{aligned} I(\mathbf{x};\mathbf{y}) &= I(x_1, x_2, \cdots, x_{M_t};\mathbf{y}) \\ &= I(x_1;\mathbf{y}) + I(x_2;\mathbf{y}|x_1) + \cdots + I(x_{M_t};\mathbf{y}|x_1, \cdots, x_{M_t-1}) \ , \end{aligned} \tag{2.16}$$

The decision feedback MMSE receiver is optimal for achieving the capacity of the MIMO channel. This is a well-known fact proved by [3, 17]. Hence, if we use different capacity achieving codewords for each layer, the MMSE DFE is an optimal receiver to achieve MIMO ergodic capacity. This only applies to the MMSE receiver and is not suitable for other linear receivers, such as zero-forcing. As for the outage capacity, the capacity achieved scheme investigation will be more involved, since it is required to design the so-called universal code to be adapted with the different channel realizations. We have omitted the introduction here.

F. MIMO OFDM

The MIMO OFDM is a natural combination of two powerful techniques and has often been used in many high speed wireless communication standards, e.g., 3GPP LTE, 802.16 (WIMAX), 802.11 (WIFI). MIMO provides space diversity and multiplex gain. OFDM transforms the frequency selective wideband channel into parallelled flat fading channels while providing some frequency diversity. Hence, for each subcarrier, all the previously introduced techniques can be used accordingly. The subcarrier channels across the frequency are correlated. The correlation depends on the channel delay spread (frequency selectivity). Usually, a carefully designed interleaver is used across the subcarrier to utilize frequency and space diversity. The combination of MIMO and OFDM provides high speed reliable wireless data communication.

CHAPTER III

NEAR OPTIMAL LOW COMPLEXITY COHERENT MIMO DETECTOR

A. Introduction

The received signal at each receive antenna in a MIMO system is a superposition of transmitted signals from different transmit antennas. If the MIMO system has M_t transmit antennas and uses a constellation of size Q, maximum-likelihood (ML) detection, which searches through all the possible transmitted symbols, requires a complexity proportional to $O(Q^{M_t})$, which is hard to implement when Q and M_t are large. A number of suboptimal detectors were proposed to reduce complexity, such as BLAST detection [18, 19], zero-forcing (ZF) and MMSE detection. However, all these schemes perform fairly far from the ML detection scheme. Recently, the sphere detection algorithm which searches in the vicinity of the received signal vector for the optimum solution was proposed [12]. The average complexity of the proposed sphere detection algorithm in general is exponential in the problem dimension M_t [15], but could be dominated by polynomial terms of M_t, when M_t is small and the corresponding signal-to-noise ratio is chosen sufficiently large [15, 13, 14].

In this Chapter, we propose a suboptimal detection algorithm for MIMO systems based on searching a subset of all the possible transmitted symbols. The proposed algorithm, which we will refer to as the List-BLAST algorithm, is introduced in Section B with two improved versions. Section C derives the space alternating generalized expectation-maximization (SAGE) algorithm [20] for MIMO detection as a further enhancement to the List-BLAST. Section D compares the complexity of the proposed algorithms with that of the sphere detection. Section E discusses soft-output detection. Section F provides simulation results and Section G concludes. Most of the

work presented here is included in [21].

B. List-Blast Detection

Consider the discrete-time model of a MIMO frequency non-selective fading channel with M_t transmit antennas and M_r ($M_r \geq M_t$) receive antennas:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}, \tag{3.1}$$

where $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_{M_t}]$ is a $M_r \times M_t$ MIMO channel assumed to be perfectly known at the receiver side (We have incorporateed the SNR factor $\sqrt{\frac{\rho}{M_t}}$ into the channel matrix $\mathbf{H}$ without loss of generality). The maximum-likelihood (ML) detector then is:

$$\hat{\mathbf{x}}_{ml} = \arg \min_{\mathbf{x} \in \Omega^{M_t}} ||\mathbf{y} - \mathbf{H}\mathbf{x}||^2 \tag{3.2}$$

where Ω^{M_t} denotes the set of constellation points in the complex M_t-dimensional space. Since an exhaustive search for the ML solution over the whole set of Ω^{M_t} is too complex, we take a different approach by searching through only a subset of the candiates generated by manipulating the well-known BLAST detection scheme. We refer to this new approach and its various extensions as List-BLAST detection schemes in the sequel.

Let the QR decomposition of the channel matrix be $\mathbf{H} = \mathbf{Q}\mathbf{R}$, where $\mathbf{Q}$ is a unitary matrix and $\mathbf{R}$ is an upper triangular matrix. Letting $\mathbf{y}' = \mathbf{Q}^H\mathbf{y}$, the system in (3.1) can be expressed as $\mathbf{y}' = \mathbf{R}\mathbf{x} + \mathbf{w}'$, where $\mathbf{w}' = \mathbf{Q}^H\mathbf{w}$ has the same distribution as $\mathbf{w}$ since $\mathbf{Q}$ is unitary. In the triangularized model above, each row denotes a different transmission/detection layer with the k^{th} layer interfered only by layers with indices larger than k. In BLAST, one first detects $\hat{x}_{M_t}$; assuming $\hat{x}_{M_t}$ is correct, the interference of $r_{M_t-1,M_t}\hat{x}_{M_t}$ can be subtracted from layer $N_t - 1$ and $\hat{x}_{N_t-1}$

can be detected as in a scalar channel. Similarly, layers $N_t - 2, N_t - 3, \cdots, 1$ can be detected in order. In the proposed list-BLAST scheme, we perform an exhaustive search over all C discrete values $\hat{x}_{M_t}$ could take; for a given $\hat{x}_{M_t}$, we use the BLAST algorithm to detect the remaining elements of the vector $[\hat{x}_{M_t-1}, \hat{x}_{M_t-2}, \ldots, \hat{x}_1]$. This results in a list of C candidate points, each of which is a vector in the complex M_t dimensional space. Finally, we select the one which minimizes $||\mathbf{y}' - \mathbf{R}\hat{\mathbf{x}}||^2$ as the detected symbol vector. It can be easily shown that the list-BLAST algorithm for $M_t = 2$ is actually maximum-likelihood.

It is well known that the performance of BLAST detection can be improved by ordering the sequence of nulling and canceling. Each different order corresponds to a unique ranking of the columns of the channel matrix $\mathbf{H}$ in the above implementation using a QR decomposition. Thus, we can also extend the list-BLAST algorithm as follows.

- List-Ranked-BLAST: In this extension, the least reliable layer with the lowest signal-to-noise ratio is detected, or more accurately, listed first; the remaining layers are detected from the most reliable (with the highest SNR) to the least reliable. This is quite different from the optimal detection sequence in the traditional Ranked-BLAST detection, which is from the most reliable layer to the least reliable layer. The motivation is as follows: Since we do an exhaustive search over all transmitted symbols in the first layer, it is the most protected layer and therefore should be used against the lowest SNR. For each symbol in the list of all possible symbols in the first layer, the remaining layers are detected in the normal way, i.e. from the highest SNR to the lowest to minimize the chance of error propagation.

- List-Shifted-BLAST: We cyclicly shift (either right or left) the columns of $\mathbf{H}$ by

one, and apply the List-BLAST algorithm as described above to each shifted $\mathbf{H}$. If shifting is performed K times, where $1 \leq K \leq M_t$, we will have $C \times K$ candidates, from which the final detected signal vector will be selected based on Maximum Likelihood (ML) criterion. A larger value of K results in better performance, as will be shown in the simulation, but higher complexity. Therefore, the List-Shifted-BLAST algorithm provides a flexible trade-off between complexity and performance. Instead of cyclicly shifting the columns of $\mathbf{H}$, random permuting can also be used in a similar way.

C. SAGE-Aided List Blast Detection

Fessler and Hero [20] extended the classical EM algorithm [22, 23] to the SAGE algorithm, applied it to estimate superimposed signals in Gaussian noise and showed that SAGE converges more quickly than EM in this case. SAGE were applied to resolve interfering signals in CDMA multi-user detection [24] and channel estimation for multiple-antenna OFDM systems [25]. We will consider using SAGE algorithm to improve the List-Blast detection. In a detection problem where the parameter set is discrete, the convergence of the EM algorithm to even a local maximum has not been proven [26]. To improve the likelihood of converging to the true ML solution, we use the listed candidates as multiple initial points in SAGE to converge to another set of C points. We then compare these C points and select the one which minimizes $||\mathbf{y}' - \mathbf{R}\hat{\mathbf{x}}||^2$. We refer to this scheme as SAGE-aided List-BLAST detection in the sequel.

We include below a brief derivation for MIMO detection using SAGE. We choose the hidden data space $\mathbf{z}_i$ with respect to $\mathbf{x}_i$ for $i = 1, 2, \cdots, M_t$ alternately in each

iteration and associates all the noise variance with it. Therefore, we have

$$\mathbf{z}_i = \mathbf{h}_i x_i + \mathbf{w}, \;\; 1 \leq i \leq M_t, \tag{3.3}$$

$$\mathbf{y} = \mathbf{z}_i + \sum_{j=1, j\neq i}^{M_t} \mathbf{h}_j x_j. \tag{3.4}$$

Let $f(\mathbf{z}_i; \mathbf{x}_i)$ be the probability density function of $\mathbf{z}_i$ parameterized by $\mathbf{x}_i$ and $\hat{\mathbf{x}}^{(k)}$ the estimate of $\mathbf{x}$ at the k-th iteration. In the E-step we compute $U_i(\mathbf{x}, \hat{\mathbf{x}}^{(k)}) \triangleq \mathbf{E}\{\log f(\mathbf{z}_i; \mathbf{x}_i) | \mathbf{y}, \hat{\mathbf{x}}^{(k)}\}$ as expressed below, where the expectation operation $\mathbf{E}\{\cdot\}$ is with respect to the conditional distribution of $f(\mathbf{z}_i | \mathbf{y}, \hat{\mathbf{x}}^{(k)})$. We have:

$$U(\mathbf{x}, \hat{\mathbf{x}}^{(k)}) = c + x_i^* \mathbf{h}_i^H \bar{\mathbf{z}}_i + x_i \bar{\mathbf{z}}_i^H \mathbf{h}_i - |x_i|^2 \|\mathbf{h}_i\|^2, \tag{3.5}$$

where c is a constant not a function of $\mathbf{x}_i$; $\bar{\mathbf{z}}_i = E[\mathbf{z}_i | \mathbf{y}, \hat{\mathbf{x}}^{(k)}]$ is the conditional mean of $\mathbf{z}_i$ given $\mathbf{y}$ and $\hat{\mathbf{x}}^{(k)}$. Since $\mathbf{z}_i$ and $\mathbf{y}$ are jointly Gaussian, we have

$$\bar{\mathbf{z}}_i = \mathbf{h}_i \hat{x}_i^{(k)} + \left(\mathbf{y} - \sum_{j=1}^{M_t} \mathbf{h}_j \hat{x}_j^{(k)} \right). \tag{3.6}$$

Maximizing $U(\mathbf{x}, \hat{\mathbf{x}}^{(k)})$ with respect to $\mathbf{x}$ in the M-step, we have

$$\tilde{x}_i^{(k+1)} = \frac{\mathbf{h}_i^H \bar{\mathbf{z}}_i}{\|\mathbf{h}_i\|^2}, \;\; 1 \leq i \leq M_t. \tag{3.7}$$

To account for the fact that $\mathbf{x}$ is discrete and x_i $(1 \leq i \leq M_t)$ must be a constellation point, we invariably quantize $\hat{x}_i^{(k)}$ to its nearest constellation point in each iteration. Let $a_{i,j} \triangleq \mathbf{h}_i^H \mathbf{h}_j / \|\mathbf{h}_i\|^2$ and $b_i \triangleq \mathbf{h}_i^H \mathbf{y} / \|\mathbf{h}_i\|^2$. Substituting (3.6) in (3.7) and considering the quantization process we can summarize the SAGE iteration as follows:

- Initialize with some $\hat{x}_i^{(0)}$ for $1 \leq i \leq M_t$.
- At the $(k+1)^{th}$ iteration $(k = 0, 1, 2, ...)$:

For $i = 1 + [k \bmod M_t]$, compute

$$\hat{x}_i^{(k+1)} = Q\left(\hat{x}_i^{(k)} + \left[b_i - \sum_{j=1}^{M_t} a_{i,j}\hat{x}_j^{(k)}\right]\right), \tag{3.8}$$

for $1 \le j \le M_t$ and $j \neq i$, $\hat{x}_j^{(k+1)} = \hat{x}_j^{(k)}$.

D. Implementation and Complexity

For PSK modulation, sphere decoding can be implemented over the M_t dimensional complex space [8]. However, QAM modulation is usually handled by decoupling the real and the imaginary components; thus, the sphere detector need to search over a $2N_t$ dimensional real space. In contrast, the SAGE-aided List-BLAST schemes solve both QAM and PSK detection in the same fashion.

Assume a block fading channel. We need to consider the computational complexity for a whole block and that for each vector symbol in the block [27]. We denote the first kind of complexity as pre-detection complexity and the second kind as detection complexity. For List-BLAST detection, the pre-detection complexity requires $O(M_t^3)$ computations for the QR decomposition. If ordered BLAST is required, the asymptotic computational complexity is still of $O(M_t^3)$ by using some fast algorithm [28]. If SAGE is used, $\{a_{i,j}\}$ and $\{b_i\}$ can be pre-computed with complexity of $O(M_t^3)$. Similarly, sphere detection requires computing of both the QR decomposition and pseudo-inverse $\mathbf{H}$ with a complexity of $O(M_t^3)$ [8].

In the case of very slow fading, the channel remains constant during each transmission block which could be composed of hundreds of vector symbols, the pre-detection complexity can be very low per vector symbol and the detection complexity dominates. For computational overhead for each vector symbol detection, the List-BLAST and the List-Ranked-BLAST algorithms require the same computation

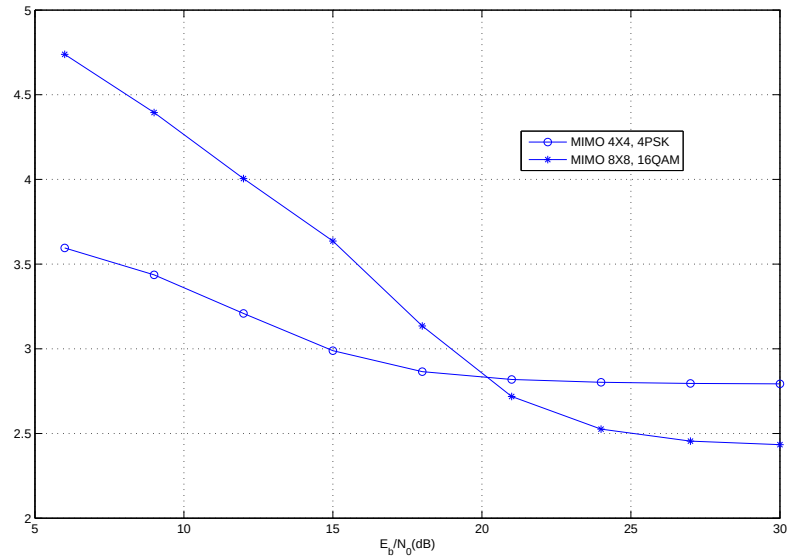

Fig. 4. Average detection complexity in number of FLOPS for sphere detectors. $m = 2N_t$

complexity of $O(CM_t^2)$. The List-Shifted-BLAST requires complexity of $O(KCM_t^2)$. Each run of SAGE starting from a single initial point requires a complexity of $O(M_t^2)$. Note that most of the time the SAGE algorithm converges in $1-3$ iterations; the number of iterations does not seem to be a function of M_t according to the observation in our simulations. Therefore, if we perform SAGE aided detection on top of List-BLAST or List-Ranked-BLAST, the complexity is still at $O(CM_t^2)$. If SAGE aided List-Shifted-BLAST is performed with $K = M_t$, the complexity is $O(CM_t^3)$. In contrast, the complexity of sphere detection is closely related to SNR and channel realizations. For practical value of M_t, at low SNR, it could require an average complexity of $O((2M_t)^4)$ - $O((2M_t)^{4.5})$ [29]. Some "bad" (with spread singular values of $\mathbf{H}$) channel realizations require more computation.

Besides asymptotic complexity measured with respect to M_t, we also compare the complexity in terms of average number of flops (floating point operations). The

Table I. Detection Complexity ($log_{2M_t}(\cdot)$) of List-BLAST Algorithms

Settings	List-BLAST	SAGE-aided List-BLAST	List-Shifted-BLAST	SAGE-aided List-Shifted-BLAST
4×4, 4PSk	2.6025	3.4754	3.2860	4.1449
4×4, 16QAM	3.2860	4.1449	3.9527	4.8115
8×8, 4PSk	2.4645	3.0916	3.2176	3.8421
8×8, 16QAM	2.9676	3.5921	3.7176	4.3421

average detection complexity of sphere decoding is shown in Fig. 4. The detection complexity of the list-BLAST algorithms is shown in Table I. The List-BLAST/List-Ranked-BLAST algorithm is the most efficient and has less complexity than that of sphere detection in low and medium SNR region. At high SNR, depending on the constellation size C and the number of shifts K ($K = M_t$ in Table I), SAGE-aided List-BLAST and List-Shifted-BLAST may have higher complexity than that of the sphere detection. The SAGE-aided List-Shifted-BLAST algorithm has relative higher complexity, therefore may not be efficient for hard detection. However, since it can list $2M_tC$ candidates without extra computations, it is highly efficient in soft-output detection as will be introduced next.

E. Soft-output Detection

The list-BLAST type algorithm provides us a natural way to decode and generate soft-information. We assume that the information bits have been encoded with a channel code, randomly interleaved, Gray-mapped to the constellation and then transmitted through M_t different antennas. Therefore, M_tM coded bits are transmitted per channel use, where $M = \log_2 C$.

At the receiver, MAP joint demodulation and detection can be used. The a priori L-value of the coded bits b_k, $k = 0, 1, \ldots, M_t M - 1$, is defined as $L_A(b_k) = \ln \frac{P[b_k=1]}{P[b_k=-1]}$. We can use the SAGE-aided List-Shifted-BLAST algorithm to generate a candidate signal set $\mathbb{L}$, which can be divided into two sets: $\mathbb{L}_{k,+1}$ if $b_k = 1$ and $\mathbb{L}_{k,-1}$ otherwise. Using the max-log approximation, the extrinsic L-value can be approximated as [**?**]

$$L_E(b_k|\mathbf{y}) \approx \max_{\mathbf{x}\in\mathbb{L}_{k,+1}} \left\{ -\frac{||\mathbf{y} - \mathbf{H}\mathbf{x}||^2}{N_0} + \frac{1}{2}\mathbf{b}_{[k]}^T \mathbf{L}_{A,[k]} \right\} - \max_{\mathbf{x}\in\mathbb{L}_{k,-1}} \left\{ -\frac{||\mathbf{y} - \mathbf{H}\mathbf{x}||^2}{N_0} + \frac{1}{2}\mathbf{b}_{[k]}^T \mathbf{L}_{A,[k]} \right\}, \tag{3.9}$$

where $\mathbf{b}_{[k]}$ denotes the sub-vector of $\mathbf{b}$ omitting its k^{th} element, and $\mathbf{L}_{A,[k]}$ is the vector of all L_A values, also omitting its k^{th} element. It is more desirable to include both the List-Shifted-BLAST solutions (the initial points in the SAGE algorithm) and the converged points after the SAGE iterations in $\mathbb{L}$ for two reasons. First, the List-Shifted-BLAST algorithm ensures that $\mathbb{L}_{k,+1}$ and $\mathbb{L}_{k,-1}$ will not be a null set due to the exhaustive listing of the constellation points for each transmit antenna. Second, the SAGE iteration will likely produce some candidates in the vicinity of the received vector [1]. These candidates are more reliable to be used in computing (3.9) using the max-log approximation. We note that the ML solution $\hat{\mathbf{x}}_{ml}$ may not necessarily be the candidate $\mathbf{x}$ which maximizes one of the two terms in the RHS of (3.9), which could be relatively far away from $\mathbf{y}$ due to the fact that turbo-coded systems usually operate at very low SNR. Therefore, if one uses sphere decoder to list the candidates as in [**?**], the search radius of the sphere decoder should be much larger than that in the case of high SNR, let alone one need to search back and forth in order to get multiple candidates. Therefore, in our simulations, we found that sphere detection requires significantly higher complexity than SAGE-aided List-Shifted-BLAST detection.

[1] $\mathbf{x}$ is in the vicinity of $\mathbf{y}$ in the sense that $||\mathbf{y} - \mathbf{H}\mathbf{x}|| < \delta$, where δ is small.

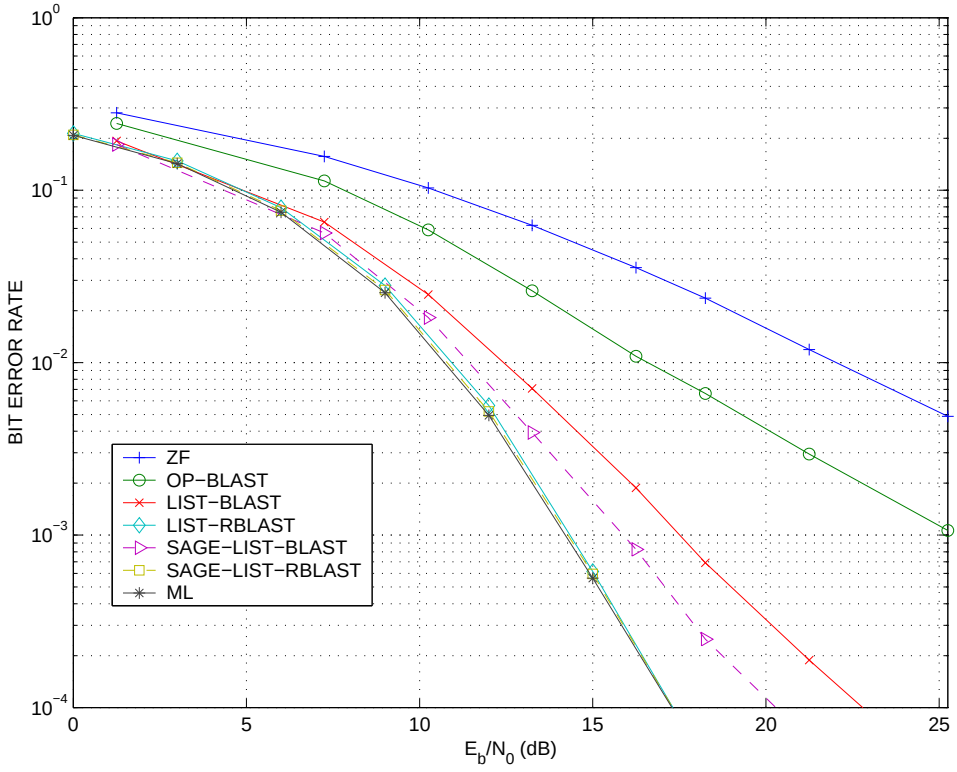

Fig. 5. Bit Error Rate of different detectors for a 4×4 MIMO system with uncoded 8-PSK modulation

F. Simulation Results

In the following simulations, we define E_b as the signal energy per transmitted information bit at the receiver. Thus, we have $\frac{E_b}{N_0} = \frac{E_s}{N_0} + 10\log_{10}\frac{M_r}{RN_tM}$, where M is the number of bits per transmitted symbol and R is the rate of the channel code. We assume an independently faded MIMO channel in each channel use. Note that for uncoded systems, the average bit error rate (BER) of the independently faded MIMO channel is the same as that of the block faded MIMO channel. We first consider an uncoded system, in which the channel code rate is $R = 1$.

Fig. 5 shows the BER of the ML detector implemented by sphere detection, the ZF detector, the zero-forcing BLAST detector with optimal detection order (the

layer with the highest SNR is detected first), the List-BLAST detector, the List-Ranked-BLAST detector, and the SAGE-aided detectors for a 4×4 MIMO system with uncoded 8PSK. Fig. 6 shows the BER of the different detectors for the same MIMO system with uncoded 16QAM. The ML detector achieves a spatial diversity order of four in this case, while the ZF detector achieves no spatial diversity. For both the QAM and the PSK modulation, the BLAST detector with optimal detection order, denoted as "OP-BLAST" in all the figures, achieves a spatial diversity order greater than one, but is outperformed by the list-BLAST detector denoted as "LIST-BLAST" and the List-Ranked-BLAST detector denoted as "LIST-RBLAST". Because of the exhaustive search in the first detection layer (the M_t^{th} layer) of the list-BLAST algorithm, the error probability is dominated by the $(M_t - 1)^{th}$ layer, which has spatial diversity of order two. We can achieve a further 2.5 dB gain by performing SAGE iterations as shown by the curve denoted as "SAGE-LIST-BLAST". In the List-Ranked-BLAST case, diversity order is further improved by ordering the nulling and cancelling as described in Section B. Actually, the List-Ranked-BLAST algorithm performs almost the same as the ML detector in this 4×4 MIMO case. Therefore, SAGE iterations cannot further improve its performance and are not required in this case. Since the List-Ranked-BLAST detection scheme has much lower complexity than that of the sphere detection, it is an excellent candidate for detection of 4×4 uncoded MIMO systems.

Fig. 7 shows the BER of the List-Ranked-BLAST scheme for an 8×8 MIMO system with uncoded 16-QAM modulation. In contrast to Fig. 6, the List-Ranked-BLAST detector is 2.5 dB worse at $BER = 10^{-6}$ than that of the ML detector. The SAGE-aided List-Ranked-BLAST detector can achieve an additional gain of 0.5 dB. Although not plotted, the performance of the List-Shifted-BLAST detector ($K = M_t$) in the 8×8 and the SAGE-aided List-Shifted-BLAST is almost the same as that of

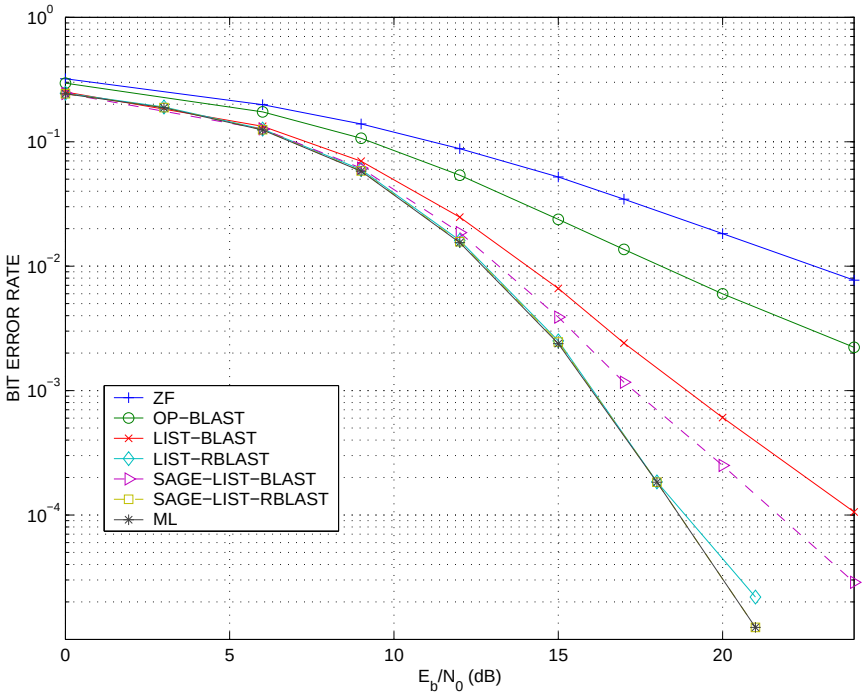

Fig. 6. Bit Error Rate of different detectors for a 4×4 MIMO system with uncoded 16-QAM modulation

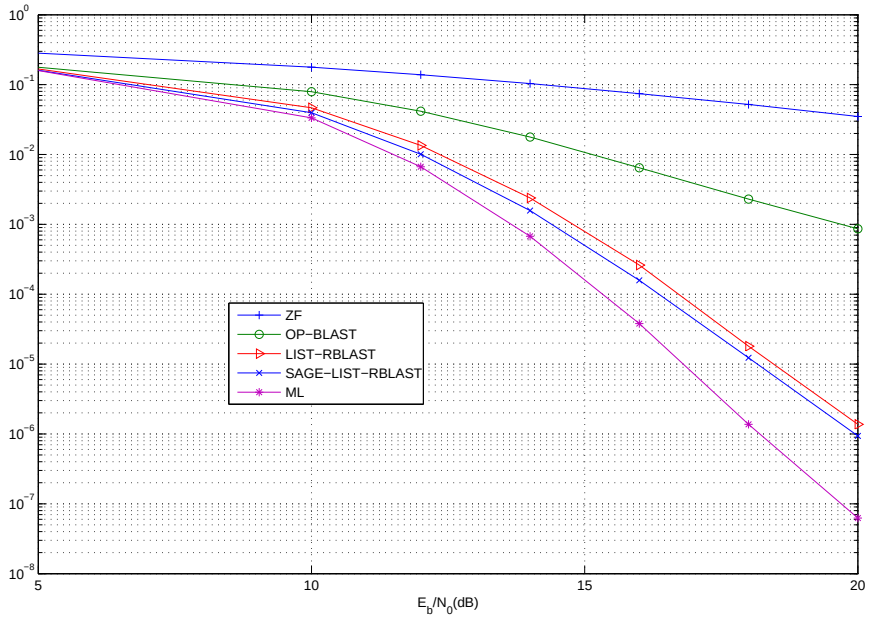

Fig. 7. Bit Error Rate of different detectors for a 8×8 MIMO system with uncoded 16-QAM modulation

the ML detector. However, they are more computationally complex than the List-Ranked-BLAST detector.

Finally, we compared the BER performance of the SAGE aided List-Shifted-BLAST algorithm and the sphere detection in a turbo-coded 4×4 MIMO system in Fig. 8. We choose the same system parameters as used in [8]. So, 16-QAM with Gray mapping and the rate $R = 1/2$ four state parallel turbo code with polynomial (7, 5) are used. The interleaver size of the turbo code is 9216 information bits and a random interleaver is used between the modulator and the turbo encoder. Both the initial points generated using the List-Shifted-BLAST algorithm ($K = 4$) and the converged points generated using the SAGE algorithm are included in the candidate set $\mathbb{L}$ whose size is 128, the same as used in the sphere detection. The curves associated with "4 iter" are generated using four Joint Demodulation and Detection (JDD) iterations. In

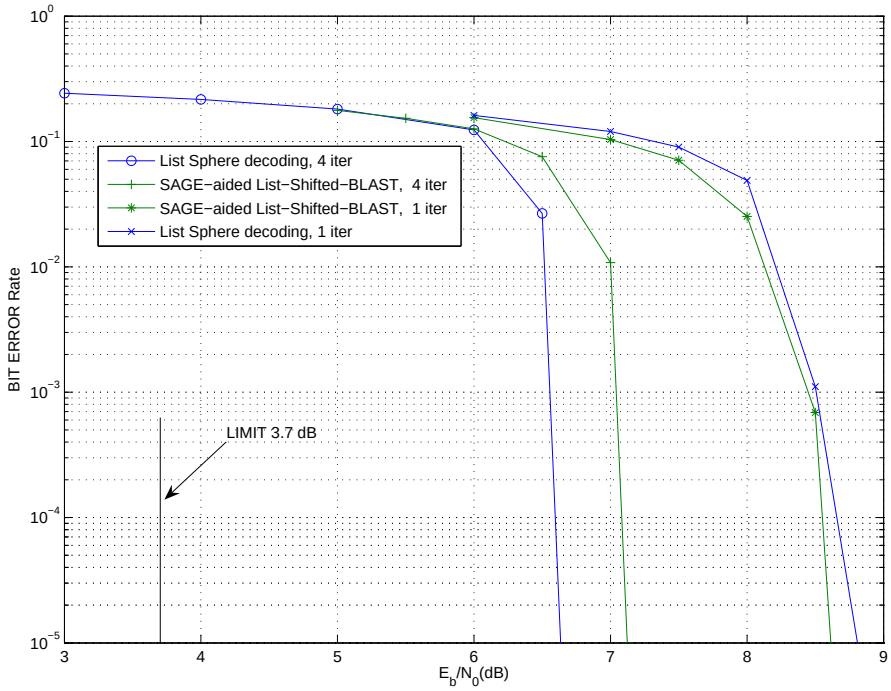

Fig. 8. Bit error rate of turbo-coded 4×4 MIMO systems with SAGE-aided List-Shifted-BLAST decoding and sphere decoding

each JDD iteration, the soft-output generated by the MIMO demodulator is passed to the turbo decoder, which has 8 iterations; the a posteriori probability of the coded bits after turbo-decoding is passed back to the demodulator for improved detection. The curves associated with "1 iter" are generated without feedback of decoder soft-output to the demodulator. The SNR at which the given rate is equal to the capacity of the MIMO systems is 3.7 (dB). The performance of the SAGE-aided List-Shifted-BLAST detection scheme without JDD is similar (surprisingly a little better) to that of the corresponding sphere detection. However, after four JDD iterations, the SAGE-aided List-Shifted-BLAST scheme is about 3 dB away from the capacity and 1 dB worse than that of the corresponding sphere detection scheme.

G. Conclusion

We introduce some low complexity sub-optimal MIMO detectors, which use the list-BLAST algorithm to generate multiple candidates from which a final one is selected based on ML principle. For hard-decision, we show that the List-Ranked-BLAST, which has a different detection order compared to the usual Ranked-BLAST, can achieve performance close to the ML detection with lower complexity at low and medium SNR than that of the sphere detection in our simulation settings. For soft-output detection, SAGE aided List-Shifted-BLAST can also achieve performance close to that of the sphere detection with much lower complexity.

CHAPTER IV

SPHERE DECODING OF ORTHOGONAL SPACE-TIME BLOCK CODES FOR NONCOHERENT CHANNELS

A. Introduction

Space-time codes are an efficient transmit diversity scheme that combats fading and achieves the high capacity promised by the multi-antenna systems. Orthogonal space-time block codes (OSTBC) [9] in particular provide a practical way to achieve the promise of spatial diversity at a reasonably low decoding complexity. Much of the previous work on decoding for space-time systems assumes knowledge of the channel fading coefficients at the receiver. There is some work for noncoherent multiple-antenna communication. For example, Hochwald and Marzetta [30] proposed unitary space-time modulation, but it suffers from exponential decoding complexity in rate and block length.

In this chapter, we propose a maximum-likelihood space-time decoding scheme that requires no channel state information at the receiver. By utilizing the structure imposed by orthogonal space-time block codes, we can convert this noncoherent detection problem into an integer quadratic programming problem and solve it efficiently by the sphere decoding algorithm. In the recent past, some attention has been paid to this problem; Ma et al., [31] use semi-definite relaxation (SDR) to solve this problem, but it is suboptimum and the complexity is of $O(N^{3.5})$, where N is the total number of symbols to be detected. Stoica and Ganesan [32] developed a blind cyclic detector to approximate the blind ML decoding problem by iterative channel estimation and symbol detection; there is some performance loss compared to our method and SDR. Uysal and Georghiades [33] implemented the ML detector with a

Viterbi algorithm. However, its complexity is exponential in the truncated memory length. The results was first included in [34].

The chapter is organized as follows. Section B introduces the signal transmission model. Section C details the ML detector for OSTBC without channel knowledge and Section D presents the sphere decoding solution for this problem. Complexity analysis and simulation results are provided in Section E and Section F concludes this paper.

B. Data Transmission Model

Consider space-time block coded transmission with M_t transmit antennas and M_r receive antennas. We assume the channel is block fading and frequency non-selective. Each space-time coded block consists of K channel uses and L space time blocks are concatenated as a long block for joint detection. The channel remains unchanged for $L \times K$ transmissions and then changes into another independent realization. The transmitted information symbols are denoted as $\{s_k^{(l)}\}_{k=1,l=1}^{K,L}$. K consecutive information symbols $\{s_1^{(l)}, \ldots, s_K^{(l)}\}$ are linearly mapped into one space-time code matrix $\mathbf{C}_{(l)}$. Mathematically, $\mathbf{C}_{(l)}$ can be written as:

$$\mathbf{C}_{(l)} = \sum_{k=1}^{K} \mathbf{X}_k s_k^{(l)} \in \mathbb{R}^{M_t \times K} \ , \tag{4.1}$$

where $\mathbf{X}_k \in \mathbb{R}^{M_t \times K}$ are fixed "elementary" code matrices; l represents the l^{th} space-time code block in the L concatenated data blocks. For OSTBC, the code matrices $\mathbf{X}_k$ satisfy the following condition:

$$\mathbf{X}_i \mathbf{X}_j^H = \begin{cases} \mathbf{I}, & i = j \\ -\mathbf{X}_j \mathbf{X}_i^H, & i \neq j \end{cases} \ , \tag{4.2}$$

with $M_t < K$, (see[**?**, 35]). For simplicity, we assume that $\mathbf{X}_k$ is real-valued. Using the properties (4.1), (4.2) of OSTBC, it can be easily shown that

$$\mathbf{C}_{(l)}\mathbf{C}_{(l)}^H = \mathbf{C}_{(l)}\mathbf{C}_{(l)}^T = (s_1^2 + s_2^2 + \ldots + s_K^2)\mathbf{I} \; . \tag{4.3}$$

If we assume the symbols are from a unitary constellation (constant energy), e.g., BPSK, then $\mathbf{C}_{(l)}\mathbf{C}_{(l)}^T = K\mathbf{I}$. Obviously, this product is independent of the transmitted symbols and can be omitted in the ML receiver design. Such orthogonal space-time structure satisfies Hochwald and Marzetta's unitary space-time modulation scheme [30]. As we will show later, the noncoherent ML OSTBC detector has polynomial decoding complexity.

For L continuous OSTBC transmissions, the received signal model is:

$$\mathbf{Y} = \sqrt{\frac{\rho}{M_t}}\mathbf{H}\mathbf{C} + \mathbf{W} \; , \tag{4.4}$$

$\mathbf{Y} \in \mathbb{C}^{M_r \times KL}$ is the received signal matrix for L space-time blocks, $\mathbf{Y} = [\mathbf{Y}_{(1)}, \ldots \mathbf{Y}_{(L)}]$; $\mathbf{Y}_{(l)}$ denotes the received signal for one OSTBC block transmission; $\mathbf{H} \in \mathbb{C}^{M_r \times M_t}$ is the complex channel matrix, whose (i, j) element $h_{i,j}$ is the complex fading gain from transmit antenna j to receive antenna i and is modeled as a circularly symmetric, complex, Gaussian random variable with zero mean and unit variance; the fading gains are assumed independent. $\mathbf{C} \in \mathbb{R}^{M_t \times KL}$ contains the L concatenated OSTBC transmitted matrices, $\mathbf{C} = [\mathbf{C}_{(1)}, \ldots \mathbf{C}_{(L)}]$, where $\mathbf{C}_{(l)}$ denotes the l^{th} OSTBC transmission with dimension $M_t \times K$; $\mathbf{W} \in \mathbb{C}^{M_r \times KL}$ represents complex, additive, circularly symmetric i.i.d. Gaussian noise with zero mean and unit variance. ρ is the average signal-to-noise ratio (SNR) at each receiving antenna.

C. The Maximum-Likelihood Noncoherent Detector

Given the transmitted matrix $\mathbf{C}$ which satisfies $\mathbf{C}^H\mathbf{C} = KLI$, the ML (incoherent) detector is is derived in [30] from the channel model and the statistics of random channel distribution:

$$\hat{\mathbf{C}} = \arg\max_{\mathbf{C}} P(\mathbf{Y} \mid \mathbf{C}) = \arg\max_{\mathbf{C}} \operatorname{tr}\{\mathbf{C}\mathbf{Y}^H\mathbf{Y}\mathbf{C}^H\} \ . \tag{4.5}$$

The Noncoherent can also derived as the generalized likelihood ratio testing (GLRT)

$$\hat{\mathbf{C}} = \arg\max_{\mathbf{C}} tr\left\{ -\left[\mathbf{Y} - \sqrt{\frac{\rho}{M_t}}\mathbf{C}\hat{\mathbf{H}}\right]^H \cdot \left[\mathbf{Y} - \sqrt{\frac{\rho}{M_t}}\mathbf{C}\hat{\mathbf{H}}\right]\right\} \ , \tag{4.6}$$

which use the coherent ML receiver with the unknown value of $\mathbf{H}$ replace by its ML estimate under the assumption the space-code matrix $\mathbf{C}$ is transmitted. Hence

$$\hat{\mathbf{H}} = \left(\frac{\rho}{M_t}\right)^{-1/2} (KL)^{-1}\mathbf{C}^H\mathbf{Y} \ . \tag{4.7}$$

The maximum-likelihood interpretation for the noncoherent receiver (4.5) assumes the channel matrix has independent elements that are distributed as $\mathcal{CN}(0,1)$, while the GLRT interpretation is less restrictive because it does not assume any thing about the statistics of the propagation matrix. Since the detector performance will depend on how good is the ML channel estimation $\hat{\mathbf{H}}$. It is well known that the unitary space-time signal constitute the optimal training signal. Specifically, if a known signal is transmitted to ML estimate the channel $\mathbf{H}$, the energy-constraint signal that minimizes the total error variance is a unitary space-time signal. Hence the fact of GLRT interpretation is further strengthened.

Similar to [31] and [32], we reformulate equation (4.5) into a quadratic form:

$$\begin{aligned}\hat{\mathbf{C}} &= \arg\max_{\mathbf{C}} \operatorname{tr}\{\mathbf{C}\mathbf{Y}^{\mathbf{H}}\mathbf{Y}\mathbf{C}^{\mathbf{H}}\} \\ &= \arg\max_{\mathbf{C}_{(l)}, l=1,\ldots,L} \|\sum_{l=1}^{L} \mathbf{C}_{(l)}\mathbf{Y}_{(l)}^{H}\|_F^2 \\ &= \arg\max_{\mathbf{C}_{(l)}, l=1,\ldots,L} \|\sum_{l=1}^{L} \operatorname{vec}\{\mathbf{C}_{(l)}\mathbf{Y}_{(l)}^{H}\}\|^2 \ . \end{aligned} \tag{4.8}$$

Writing $\mathbf{C}_{(l)}$ as the linear combination of transmitted symbols using (4.1), we obtain:

$$\operatorname{vec}\{\mathbf{C}_{(l)}\mathbf{Y}_{(l)}^{H}\} = \sum_{k=1}^{K} \operatorname{vec}\{\mathbf{X}_k\mathbf{Y}_{(l)}^{H}\}s_k^{(l)} = \mathbf{A}_{(l)}\mathbf{s}_{(l)} \ , \tag{4.9}$$

where $\mathbf{A}_{(l)} = [\operatorname{vec}(\mathbf{X}_1\mathbf{Y}_{(l)}^{H}), \ldots, \operatorname{vec}(\mathbf{X}_K\mathbf{Y}_{(l)}^{H})] \in \mathbb{C}^{M_tM_r \times K}$ and $\mathbf{s}_{(l)} = [s_1^{(l)}, \ldots, s_K^{(l)}]^T$ represents the information symbols of one OSTBC block transmission. With this result, we can further simplify the objective function, and rewrite the ML detector as:

$$\begin{aligned}\hat{\mathbf{s}} &= \arg\max_{\mathbf{s}_{(l)}, l=1,\ldots,L} \|\sum_{l=1}^{L} \mathbf{A}_{(l)}\mathbf{s}_{(l)}\|^2 \\ &= \arg\max_{\mathbf{s}} \mathbf{s}^H\mathbf{R}\mathbf{s} \end{aligned} \tag{4.10}$$

where vector $\mathbf{s} = [\mathbf{s}_{(1)}^T, \ldots, \mathbf{s}_{(L)}^T]^T$ is the transmitted data symbols over L transmission blocks, and $\mathbf{R}$ is a block positive definite matrix with (p, q) block given by

$$\mathbf{R}_{p,q} = \Re\{\mathbf{A}_{(p)}^H\mathbf{A}_{(q)}\} \tag{4.11}$$

So far we have formulated the ML detection problem for the noncoherent channel into an integer quadratic optimization problem (symbols $\mathbf{s}$ have discrete values). If there was no constraint on $\mathbf{s}$, a well-known solution is the eigenvector corresponding to the largest eigenvalue of $\mathbf{R}$. In our case, $\mathbf{s}$ belongs to a discrete signal constellation and the optimization problem involved is known to be NP hard. However, the optimization

problem can be solved efficiently using an integer programming algorithm, sphere decoding [12], presented briefly in the next section.

D. MIMO Sphere Decoding

We first transform the maximization problem into a minimization problem:

$$\begin{aligned}
\hat{\mathbf{s}} &= \max_{\mathbf{s}} \mathbf{s}^H \mathbf{R} \mathbf{s} \\
&= \min\{\xi \mathbf{I} - \mathbf{s}^H \mathbf{R} \mathbf{s}\} \\
&= \min_{\mathbf{s}} \mathbf{s}^H \underbrace{(\kappa \mathbf{I} - \mathbf{R})}_{\mathbf{W}} \mathbf{s} \qquad (4.12) \\
&\overset{(a)}{=} \min_{\mathbf{s}} \mathbf{s}^{\mathbf{H}} \mathbf{U}^{\mathbf{H}} \mathbf{U} \mathbf{s} = \min_{\mathbf{s}} \|\mathbf{U}\mathbf{s}\|^2 \ . \qquad (4.13)
\end{aligned}$$

κ is a constant that makes the matrix $\mathbf{W}$ strictly positive definite and $\xi = \mathbf{s}^H \kappa \mathbf{s}$; step (a) uses the Cholesky factorization of the positive definite matrix $\mathbf{W}$ and $\mathbf{U}$ is an upper triangular matrix. κ can take any value that is greater than the largest eigenvalue of matrix $\mathbf{R}$, e.g., $\mathrm{tr}(\mathbf{R})$ or $\lambda_{max}(\mathbf{R}) + \varepsilon$, where $\lambda_{max}(\mathbf{R})$ is the largest eigenvalue of $\mathbf{R}$ and ε is a positive value. We will optimize κ to reduce the decoding complexity in the next section. The optimization problem in (4.13) can be solved efficiently with Fincke and Pohst's lattice closest point search algorithm [36].

Brute-force searching of the whole signal space has an exponential complexity of 2^{QKL}, where Q is the signal constellation size. It's computationally intractable with large KL. Some suboptimal solutions have been proposed to reduce the search complexity at the cost of a considerable performance loss. Sphere decoding is an efficient optimal search algorithm with polynomial complexity. Instead of exhaustively searching the whole signal space, sphere decoding constraints the search in a predefined sphere. With the upper triangular structure of the lattice generator matrix $\mathbf{U}$,

upper and lower bounds for each dimension (symbol) can be recursively obtained. Updating the sphere radius with the norm of the latest valid lattice point (the latest output lattice point satisfying the upper and lower bounds for each dimension), the algorithm converges quickly to the optimal solution, especially for high SNRs.

Choose an $M = KL$ dimensional sphere with center at the origin and radius r, large enough to include at least one valid lattice point. $\mathbf{U}$ is the lattice generator matrix and $\mathbf{Us}$ is a lattice point corresponding to a valid transmitted signal $\mathbf{s}$; $\|\mathbf{Us}\|$ is the norm of the lattice point. We wish to find the minimum norm lattice point that lies within the sphere, i.e. the minimum norm lattice point satisfying

$$\|\mathbf{Us}\|^2 = \sum_{i=1}^{M} u_{ii}^2 \left[s_i + \sum_{j=i+1}^{M} \frac{u_{ij}}{u_{ii}} s_j \right]^2 < r^2 \ . \tag{4.14}$$

The sphere decoder establishes the bounds on $s_1, \ldots, s_M$, recursively from M to 1 [12]. For any lattice point within the sphere, the necessary condition for s_M is

$$\lceil -\frac{r}{u_{MM}} \rceil \leq s_M \leq \lfloor \frac{r}{u_{MM}} \rfloor \ , \tag{4.15}$$

where $\lceil \cdot \rceil$ and $\lfloor \cdot \rfloor$ are the standard ceiling and floor functions, respectively. Given $s_M, \ldots s_{i+1}$, we can establish an admissible value for s_i recursively using:

$$\begin{aligned} \lceil \frac{1}{u_{ii}} (- \sum_{j=i+1}^{M} u_{ij} s_j - \sqrt{r^2 - \sum_{j=i+1}^{M} |\sum_{l=j}^{M} u_{jl} s_l|^2}) \rceil \leq s_i \\ \leq \lfloor \frac{1}{u_{ii}} (- \sum_{j=i+1}^{M} u_{ij} s_j + \sqrt{r^2 - \sum_{j=i+1}^{M} |\sum_{l=j}^{M} u_{jl} s_l|^2}) \rfloor \ . \end{aligned} \tag{4.16}$$

The decoding algorithm is summarized as follows:

Sphere decoding algorithm (input: $(\mathbf{r}, \mathbf{U})$*, output:* $\hat{\mathbf{s}}$*)*

Step 1. Set $i := M, P_M := 0, \eta_M := 0, D = r^2$

Step 2. If $D < P_i$ go to Step4; otherwise

$$LB(s_i) := \max\{0, \lceil \frac{-\eta_i - \sqrt{D - P_i}}{u_{ii}} \rceil\} ,$$
$$UB(s_i) := \min\{Q - 1, \lfloor \frac{-\eta_i + \sqrt{D - P_i}}{u_{ii}} \rfloor\} . \tag{4.17}$$

set $s_i = LB(s_i) - 1$.

Step 3. $s_i := s_i + 1$ if $s_i \leq UB(s_i)$go to step 5, else go to step4.

Step 4. If $i = M$ terminate; else set $i := i + 1$ and go Step3.

Step 5. If $i > 1$, then let $\eta_{i-1} := \sum_{j=i}^{M} u_{i-1,j} s_j$, $P_{i-1} := P_i + |\eta_i + u_{ii} s_i|^2$, let $i := i - 1$ and go to Step2.

Step 6. Update the sphere radius as $D := P_1 + |\eta_1 + u_{11} s_1|^2$, update the solution as $\hat{\mathbf{s}} = \mathbf{s}$, and update all the upper boundaries

$$UB(s_l) = \min\{Q - 1, \lfloor \frac{-\eta_l + \sqrt{D - P_l}}{u_{ll}} \rfloor\}$$

for all $l = 1, \ldots, M$; go to Step3.

Sphere decoding has been shown to have polynomial complexity (approximately $O(N^3)$) [13, 14] for the problem at hand.

E. Performance

1. Complexity Analysis

In the implementation of the sphere decoder above, two parameters which will affect the complexity of the lattice search need to be optimized. One is the initial sphere

radius and the other is the parameter κ that transforms the maximization into a minimization problem.

Optimizing κ in (4.12) is a critical step in the efficient algorithm design. We have many choices, as long as κ is greater than the largest eigenvalue of $\mathbf{R}$, e.g., the trace of the matrix $\mathbf{R}$, which is easy to compute. However, if we use a large κ, the upper triangular matrix $\mathbf{U}$ in (4.13) is close to a diagonal matrix after the Cholesky factorization. Consequently, the sphere searching algorithm has to go into more levels to test the hypothesis whether the current signal value selection is out of the sphere constraint. This dramatically increases the search steps of the algorithm. The best choice of κ would be the smallest possible value, i.e. $\lambda_{max}(R) + \varepsilon$ where ε is a small positive value. Below, simulations of decoding complexity with different κ values are compared. Following the literature, we define the complexity exponent as:

$$e_c = \frac{\log C(M, r)}{\log M} \ , \tag{4.18}$$

where $C(M, r)$ is the number of elementary operations (additions, subtractions and multiplications) of the search algorithm and M and r are the dimension and radius of the search space, respectively. If e_c is a constant, the expected complexity is polynomial; otherwise, if e_c takes the form of $\frac{M}{\log M}$, the complexity is exponential. It will be seen below that for most reasonable SNR values the complexity exponent is less than 3 for the adapted sphere decoder. Hence for this problem the sphere decoding algorithm can be implemented efficiently.

2. Simulation Results

We adopt Ganesan's OSTBC scheme [35] which is full rate and full diversity with three transmit and four receive antennas. For simplicity, BPSK modulation is used.

The space-time component matrices are:

$$\mathbf{X_1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \quad \mathbf{X_2} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\mathbf{X_3} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ -1 & 0 & 0 & 0 \end{pmatrix} \quad \mathbf{X_4} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix} . \tag{4.19}$$

Here we let $L = 8$ (i.e., the channel remains constant for 8 coded blocks and then changes into an independent realization).

Fig. 9 gives the symbol error rate of different detection schemes without channel knowledge. The SNR denotes the signal-to-noise ratio per receiving antenna, which is defined as the ratio of the total transmit energy per channel use divided by the per-component noise variance. The results illustrate that sphere decoding performs similar to semi-definite relaxation (SDR) [31], outperforms all other schemes and is within 1.5 dB from the performance of a coherent detector. Compared with blind cyclic (BC) methods [32], sphere decoding has about a 2dB gain. As we will see below, sphere decoding has a lower complexity than SDR. In Fig. 1, we also included the pilot assisted technique, which is widely used in practice. The pilot assisted scheme achieve coherent demodulation by using one space-time block (4 symbols) as a pilot to estimate the channel. For fair comparison, we have considered the SNR penalty for the pilot assisted scheme. We see that sphere decoding method outperforms the pilot assist scheme as well. It is a well known result that for Rayleigh fading channel the DPSK performs 3 dB worse than the coherent BPSK detection and the diversity order is only one. However, from our noncoherent detector of OSTBC, there is only 1.5 dB loss and provide a diversity order the same as coherent detection.

To further demonstrate the performance of the proposed scheme, the symbol error rates v.s. the block length for different detectors are illustrated in Fig. 10. Intuitively, the longer the block, the better the symbol error rate. We observe that $L = 8$ is enough to achieve most of the gain from joint detection. Larger data blocks can not improve the performance, but increase complexity considerably.

We plot the complexity exponent for sphere decoding for different block lengths in Fig. 11. At low SNR, a longer block length has larger complexity exponent than a shorter one. However, at high SNR a shorter block length has slightly larger complexity. This behavior has to do with the quick convergence of the sphere decoding algorithm at high SNR pretty much irrespective of L. With SNR greater than 2 dB, Fig. 11 shows that the complexity exponent of sphere decoding is less than $O(N^{2.5})$, which is lower than SDR's $O(N^{3.5})$ [31].

In Fig. 12 we plot the complexity exponent corresponding to different κ values. The larger the value of κ, the higher the complexity exponent. Larger κ makes the matrix $\mathbf{U}$ more diagonally distributed. Hence, the search algorithm needs to go deeper in the tree to test the hypothesis whether the current assumed vector is within the sphere. If we assume κ equals $\mathrm{tr}(\mathbf{R})$ (in most channel realizations $\mathrm{tr}(\mathbf{R}) >> (\lambda_{max}(R)+200)$), the complexity exponent could be too large for large data blocks. Hence, the trace of the matrix is an unsuitable candidate for the maximization to minimization transform for this problem.

F. Conclusion

We investigated noncoherent decoding of OSTBC by adapting the sphere decoding algorithm previously applied to the coherent MIMO channel. The results show good performance in both symbol error rate and complexity, compared to other techniques.

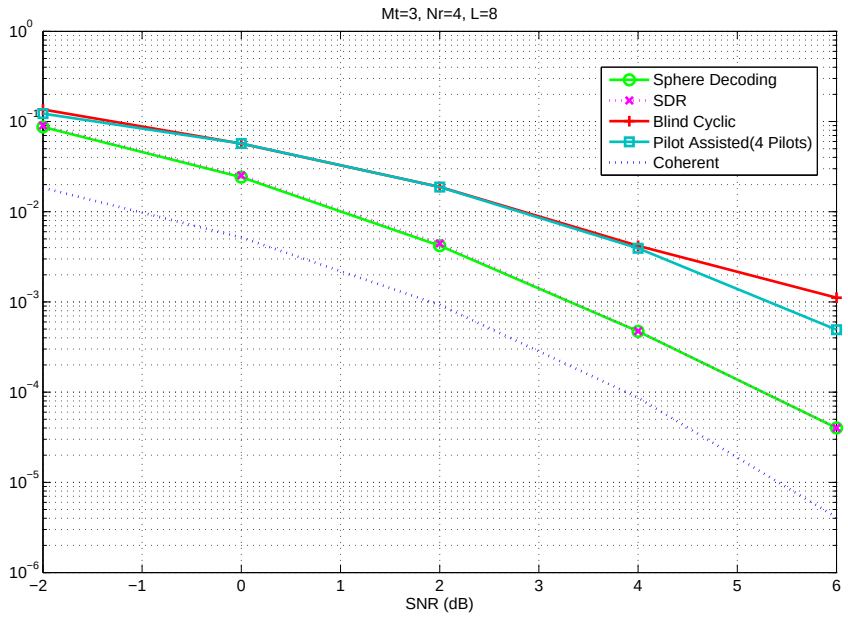

Fig. 9. Performance of the adapted sphere decoder vs other techniques for 3 transmit, 4 receive antennas and $L = 8$.

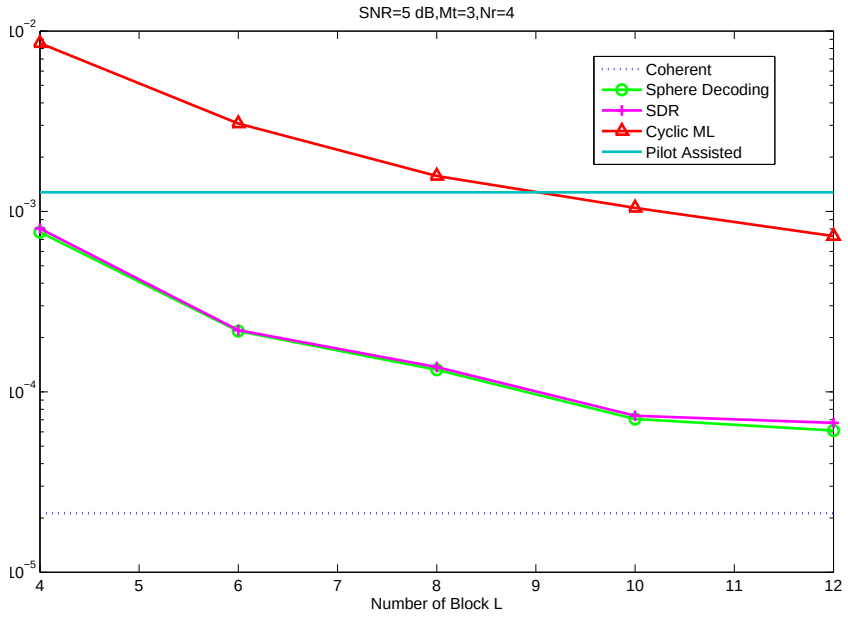

Fig. 10. Performance of different decoders as a function of block length: SNR=5, $M_t = 3$, $M_r = 4$

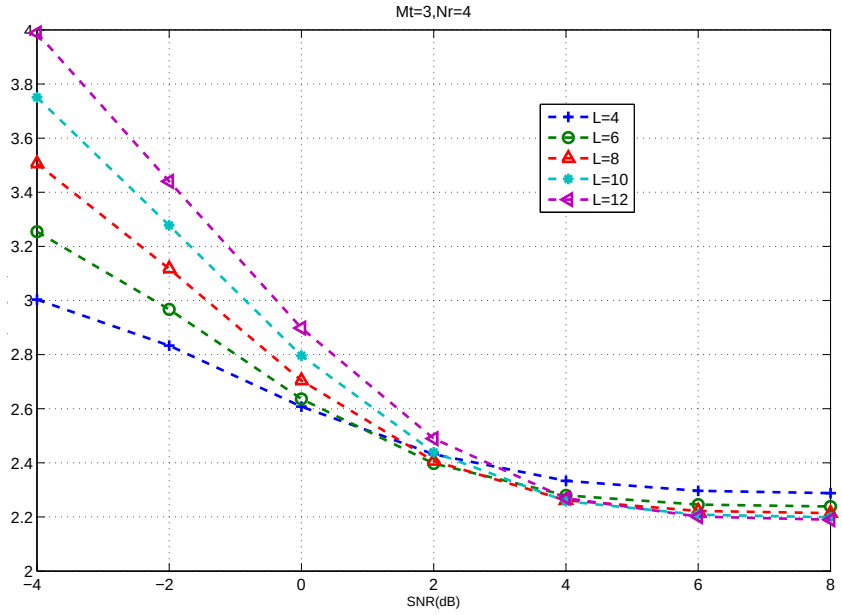

Fig. 11. The complexity exponent $\log_M$(average flops) v.s. SNR as a function of block length: $M_t = 3$, $M_r = 4$.

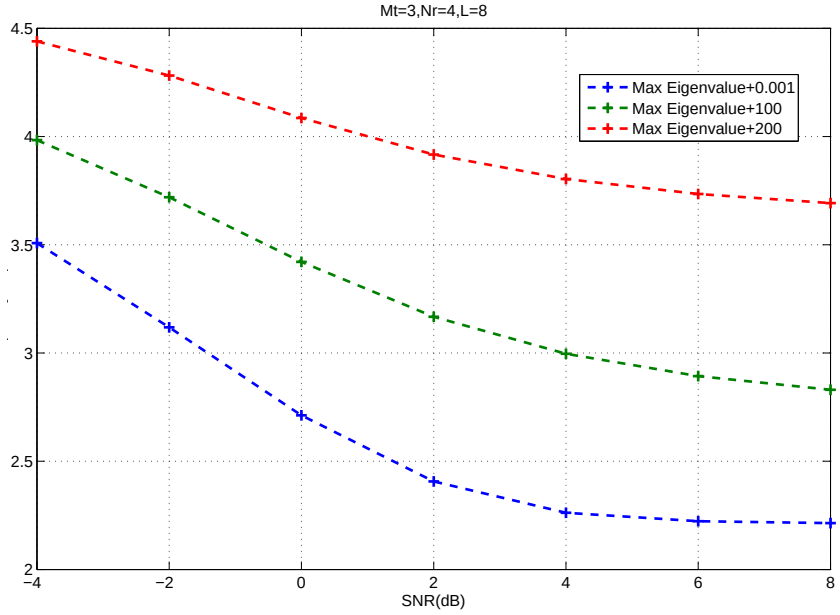

Fig. 12. The complexity exponent $\log_M$(average flops) v.s. SNR for various values of κ in (4.12): $M_t = 3$, $M_r = 4$, $L = 8$.

CHAPTER V

SIGNAL DETECTION WITH ASYNCHRONOUS CO-CHANNEL INTERFERENCE IN MIMO-OFDM SYSTEMS

A. Introduction

Increasingly, co-channel interferences (CCI) is becoming the dominant performance limiting factor in emerging high-density WLAN (HD-WLAN)[37]. The problem is exacerbated when more and more access points (AP)s are deployed in areas, such as office building, airport, university campus, etc., to provide network access for increasing number of mobile users. Only limited orthogonal channels (typically 3 or 8) are available. As a result, multiple cells that are operated on the same channel cannot be separated far enough and will interfere with each other if active at the same time, which is depicted in Fig 13. The next generation WLAN technology - 802.11n - combines orthogonal frequency division multiplex (OFDM) and multiple input multiple output (MIMO) techniques, providing good opportunities for achieving not only higher per-link throughput, but also better interference suppression capability.

Researchers have investigated the issue of CCI suppression extensively since Winters's seminal paper [38]. The use of multiple antennas brings extra degrees of freedom for CCI suppression. [39, 40] studies the throughput of interference-limited MIMO cellular system under different antenna configurations and transmission schemes. [41] investigated the MIMO capacity under interference with single-user detection. [42, 43] proposed a technique based on multiuser detection to cancel MIMO CCI for flat fading channels. Considering OFDM modulation and a time-varying channel, [44] designed an adaptive array processing scheme by using a MMSE diversity combiner. As pointed out in [45], the previous frequency domain approaches have difficulties in suppressing

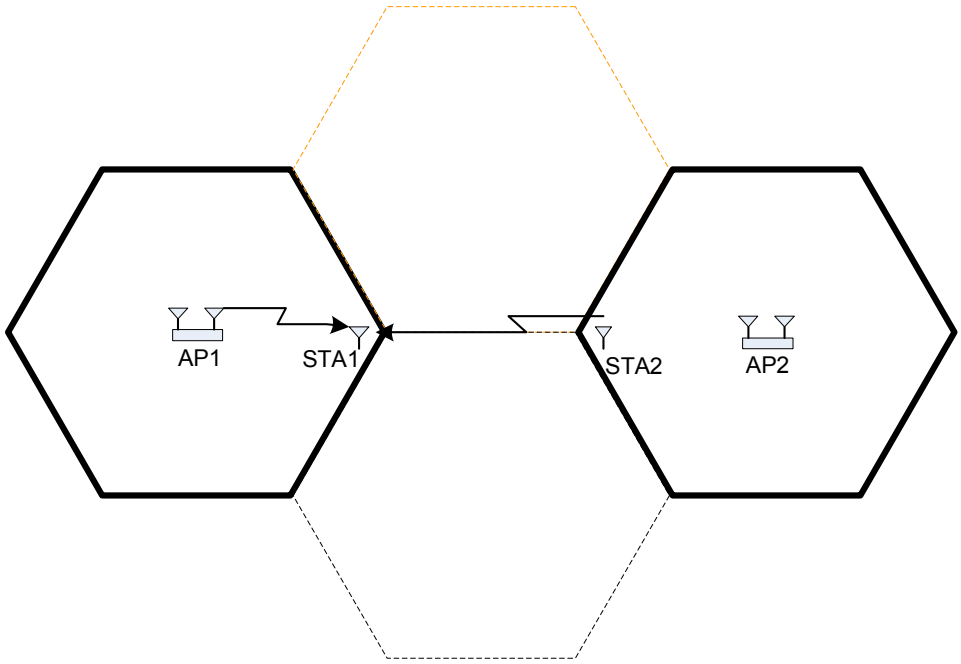

Fig. 13. Worst interference situation for 3 reusable frequency channels

asynchronous interference, so they proposed a space-time filter to suppress CCI by utilizing the OFDM cyclic-prefix structure. [46] is the most relevant to our work, which adopted a MMSE method and proposed to estimate the interference covariances for each subcarrier (or tone) by short training and utilizing the correlation among different tones. Besides the physical layer signal processing approaches, [37] designed a medium access control (MAC) based solution which adapts carrier sensing threshold to mitigate CCI from neighboring cells. It was shown by test-bed experiments that the proposed adaptive CSMA scheme can effectively address so-called "hidden and exposed terminal" problems and significantly improve network throughput.

Typically, CCI in a WLAN is asynchronous due to the use of a random access protocol, namely CSMA/CA (Carrier Sensing Medium Access/ Collision Avoidance). It was shown in [45] that the conventional frequency domain CCI cancelation by estimating both channels cannot work effectively because the cyclic padding OFDM modula-

tion structure to maintain inter subcarrier orthogonality has been destroyed. Hence, we adopted a statistical methodology – modeling the asynchronous (co-channel) interference as a zero-mean, time uncorrelated and spatially colored stationary Gaussian random process, and designed an efficient spatial covariance estimation algorithm by utilizing the OFDM symbol structure and matrix decomposition techniques. Simulation results show that our method can achieve packet error rate (PER) performance comparable to synchronized cancellation. The work in this chapter is given in a slightly different from in [47].

The rest of the chapter is organized as follows. Section B describes the system model, and introduces the effect of asynchronous interference. An efficient spatial covariance estimation method for MIMO OFDM signals is proposed in Section C. In Section D, the MMSE receiver enhanced with asynchronous CCI suppression capability is presented, as well as a modification for space-time coded systems is discussed. Then, the optimum MAP detector to minimize bit error probability is developed. Section E shows the performance of our algorithms by extensive simulations. Finally, Section F concludes.

B. System Model

Fig. 14 shows a MIMO OFDM system with M_t transmitters and M_r receivers [48]. The encoded packet is interleaved and partitioned into F blocks. Then the binary data blocks are mapped into $\{X(f,k)|1 \leq f \leq F, 1 \leq k \leq K\}$ using the selected modulation, where K is the number of subcarriers (tones). Assume the modulated symbols $X(f,k)$s have unit power. We denote $[X(f,1) \cdots X(f,K)]$ as one OFDM symbol. The OFDM symbols are space-time processed (through either space-time coded or spatial multiplexed), and then separated into M groups. Each group is transmitted

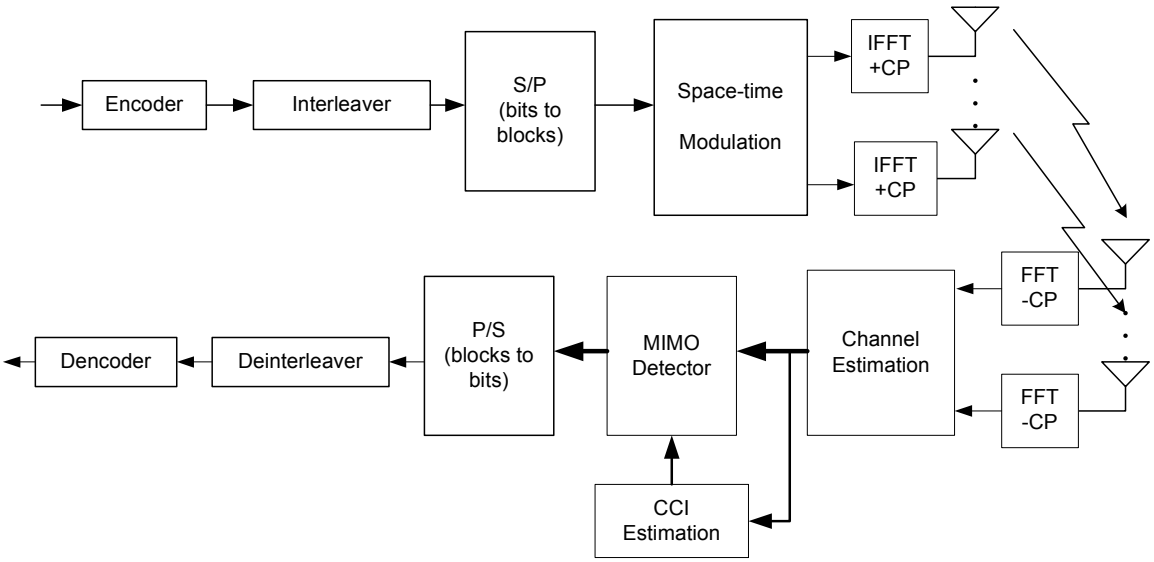

Fig. 14. MIMO OFDM system model

on one antenna. Before transmission, OFDM symbols are IFFT transformed into time domain and added cyclic prefix to minimize inter-symbol-interference (ISI) due to multi-path effect. The resulting transmission sequence is $\{x_i(n), i = 1 \cdots M\}$. We assume that the channel has L taps and remains unchanged within a packet. The same channel model is used for both intended and interference signals and a random delay τ is introduced to model the asynchrony of interference. Finally, we can express the received signal at the desired user's jth antenna as:

$$y_j(n) = \sum_{i=1}^{M_t} \sum_{l=0}^{L-1} h_{i,j,l} x_i(n-l) + \sum_{u=1}^{U} \sum_{i=1}^{M_t} \sum_{l=0}^{L-1} g^u_{i,j,l} z^u_i(n-l-\tau_u) + w_j(n) \ . \qquad (5.1)$$

where $h_{i,j,l}$ and $g^u_{i,j,l}$ define the the lth tap channel response for the desired transmitter and the uth interferer between the ith (transmit) antenna and the jth (receive) antenna, and $w_j(n)$ is the additive complex white Gaussian noise with zero mean and variance N_0. The second term in the above equation represents the co-channel interference.

Asynchrony destroys cyclic structure so that interference can no longer be modeled as the interferer's channel on the given subcarrier multiplied by the data symbol. All the taps of the time-domain channel response will contribute to the interference for each tone. Now, take one interferer's signal and denote it as $q(n)$. For simplicity, we assume one transmit antenna case, but extension to the MIMO channel is straightforward. We rewrite the interference as cyclic structure according to the timing of the desired signal [45].

$$q(n) = \sum_{l=0}^{L-1} g_l z([n-l]_K) + \sum_{l=0}^{L-1} g_l \{z(n-l) - z([n-l]_K)\} \mathbf{1}_{(n-l)<0} \ , \qquad (5.2)$$

where K is the FFT size of circular convolution; $[n]_K$ means n mod k, and the indicator function $\mathbf{1}_{(n-l)<0}$ is one if $(n-l) < 0$ and zero otherwise. Basically, the interference signal is described as a circular convolutional term plus a correction term. Taking FFT of (5.2) gives:

$$Q(k) = G(k)Z(k) + \sum_{n=0}^{K-1} \sum_{l=0}^{L-1} g_l \{z(n-l) - z([n-l]_K)\} \cdot \mathbf{1}_{(n-l)<0} \cdot e^{-j2\pi kn/K} \ , \quad (5.3)$$

with $G(k)$ and $Z(k)$ being the K-point FFT of g_l and $z(n)$ (for $0 \leq n \leq K-1$), respectively.

The second term of (5.3) implies we need L degrees of freedom to suppress the interference effectively by using the conventional MMSE receiver that estimates both the desired signal channel h_l and the interference channel g_l.

C. Spatial Covariance Estimation for Asynchronous Interference

Instead of estimating the interference channel response, we model it as a zero mean, spatially colored Gaussian stationary random process for each tone. Hence, the second moment - covariance completely characterizes the statistics of interference. We

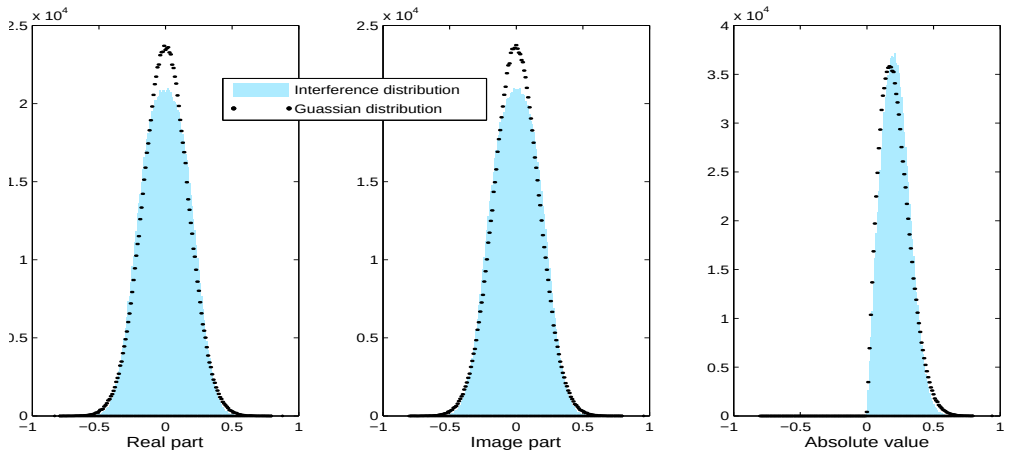

Fig. 15. Histogram of asynchronous interference (1 interferer)

proposed a statistical methodology to address the challenges of asynchronous CCI suppression. The Gaussian approximation is simple and easy for receiver design. Fig. 15 shows the accuracy of the Gaussian modeling of the asynchronous interference. We can observe that Gaussian distribution matches the asynchronous interference statistics well. The more the structure of the interference is exploited, the more effective the interference suppression algorithm is. We write the baseband received signal in the kth tone as follows:

$$\mathbf{Y}^k(n) = \mathbf{H}^k\mathbf{X}^k(n) + \mathbf{I}^k(n) \ , \tag{5.4}$$

where $\mathbf{I}(n) \in \mathbb{C}^{M_r \times 1}$ represents interference plus noise, i.e., we lumped (5.3) and additive Gaussian noise into $\mathbf{I}(n)$. The goal is to efficiently estimate the covariance of $\mathbf{I}(n)$ in each tone, and then design the Wiener filter or the optimum MAP detector to suppress interference.

The spatial covariance of $\mathbf{I}(n)$ in the kth tone can be expressed as:

$$\mathbf{R}_{II}^{k} = E\{\mathbf{I}^{k}(n)\mathbf{I}^{k}(n)^{H}\} = \frac{1}{P}\lim_{P\to\infty}\sum_{n=0}^{P-1}\{\mathbf{I}^{k}(n)\mathbf{I}^{k}(n)^{H}\} \ ,$$

where P is the training OFDM symbol number. However, it is not practical to measure the interference statistics over a long time, and therefore a parsimonious accurate spatial covariance estimator should be used. The most commonly used choice is the sample average, which is the maximum likelihood (ML) estimator and it is unbiased. However, the ML estimator is known to have a large tendency to spread the eigenvalues. This tendency is highly undesirable and often causes a substantial degradation in performance. Therefore, we proposed to utilize the correlation information of OFDM tones to refine the estimation. Such correlation is inherent in the OFDM modulation structure and the multipath characteristics of fading channels. Moreover, we use the Cholesky decomposition method to turn a constrained parameter estimation problem (positive definite matrix) into an unconstrained one.

1. Temporal Low-Pass Smoothing

Let $\tilde{\mathbf{R}}_{II}^{k} = \frac{1}{P}\sum_{n=1}^{P-1}\{\mathbf{I}^{k}(n)^{H}\mathbf{I}^{k}(n)\}$, where $\tilde{\mathbf{R}}_{II}^{k} \in \mathbb{C}^{M_r\times M_r}$. The matrix sequence $\{\tilde{\mathbf{R}}_{II}^{1}\cdots\tilde{\mathbf{R}}_{II}^{K}\}$ fully characterizes the statistics of the interference. The diagonal entries of the matrix sequence $\tilde{\mathbf{S}}_{nn} = \{\tilde{\mathbf{R}}_{II}^{1}[n,n]\cdots\tilde{\mathbf{R}}_{II}^{K}[n,n]\}$ are the estimated power spectral density (PSD) of signal from nth receive antenna. Similarly, the off-diagonal sequence $\tilde{\mathbf{S}}_{mn} = \{\tilde{\mathbf{R}}_{II}^{1}[m,n]\cdots\tilde{\mathbf{R}}_{II}^{K}[m,n]\}$ represents the estimate of mutual power spectral density between signals from the mth and nth antennas. We transform the auto/mutual PSD back to time domain with IFFT to get the cyclic auto/cross-correlation sequences.

$$\tilde{\mathbf{r}}_{mn} = \mathbf{F}^{-1}\tilde{\mathbf{S}}_{mn}, \qquad m,n = 1\cdots M_r \ , \tag{5.5}$$

where $\mathbf{F}$ is a $K \times K$ FFT matrix, $\tilde{\mathbf{r}}_{mn}$ denotes the correlation function.

We notice that the received signal is the sum of the OFDM signals which have propagated through the multipath channels with an additive Gaussian white noise. We assume the original signals sent by each transmit antenna are uncorrelated in the time domain. Let the maximum delay tap of the multipath channel be L. After the transmitted signal is convolved with the multi-tap channel response, two timing received samples will be correlated if separated by less than L, and uncorrelated otherwise,

$$\tilde{\mathbf{r}}_{mn} = \{\tilde{\mathbf{r}}_{mn}[0],\ \cdots, \tilde{\mathbf{r}}_{mn}[L-1]\ , 0,\quad \cdots,\ 0,\ \tilde{\mathbf{r}}_{mn}[K-L+2],\ \cdots, \tilde{\mathbf{r}}_{mn}[K]\} \quad (5.6)$$

Clearly, the correlation function has "low-pass" property, which will be exploited to smoothing the estimation. We null the terms for $L \leq k \leq K-L+1$ as shown in (5.7) before transforming $\tilde{\mathbf{r}}_{mn}$ back to frequency domain.

$$\begin{aligned} \hat{\mathbf{r}}_{mn} &= \mathbf{D}\tilde{\mathbf{r}}_{mn}, \qquad \mathbf{D} = \mathrm{diag}(d_k), \\ d_k &= [\underbrace{1, \cdots, 1}_{L}, 0, \cdots, 0, \underbrace{1, \cdots, 1}_{L-1}]\ . \end{aligned} \quad (5.7)$$

Hence, we get the smoothed spatial covariances estimations as

$$\begin{aligned} \hat{\mathbf{S}}_{mn} &= \mathbf{F}\hat{\mathbf{r}}_{mn} \\ &= \mathbf{F}\mathbf{D}\mathbf{F}^H\tilde{\mathbf{S}}_{mn} \\ &= \mathbf{P}\tilde{\mathbf{S}}_{mn}\ . \end{aligned} \quad (5.8)$$

For the case when only part of the tones are used to transmit data (for example, one 802.11 a/g/n OFDM symbol consists of 52 tones from 64 available), we try to estimated the $2L-1$ correlation parameters from the K_c data subcarriers. Let $\bar{\mathbf{r}}_{mn}$ be a $2L-1$ dimension vector, which was formed by removing zeros from $\tilde{\mathbf{r}}_{mn}$ in (5.6).

We have following least square estimation:

$$\hat{\bar{\mathbf{r}}}_{mn} = \min_{\bar{\mathbf{r}}_{mn}} \|\tilde{\mathbf{S}}_{mn} - \mathbf{F}_c \mathbf{B} \bar{\mathbf{r}}_{mn}\| \tag{5.9}$$

where F_c is the truncated FFT matrix with rows correspond to used data subcarriers, and

$$\mathbf{B} = \begin{pmatrix} \mathbf{I} & 0 & 0 \\ \underbrace{0}_{L} & \underbrace{0}_{K-2L+1} & \underbrace{\mathbf{I}}_{L-1} \end{pmatrix} . \tag{5.10}$$

Hence,

$$\hat{\mathbf{S}}_{mn} = \underbrace{\mathbf{F}_c \mathbf{B} \Big((\mathbf{F}_c \mathbf{B})^H (\mathbf{F}_c \mathbf{B}) \Big)^{-1} (\mathbf{F}_c \mathbf{B})^H}_{\mathbf{P}} \tilde{\mathbf{S}}_{mn} \tag{5.11}$$

Where the filter matrix $\mathbf{P}$ can be pre-computed and stored. Similar to the conventional low-pass filter in signal processing that can smooth temporal correlated signals, the above process can be regarded as a temporal low-pass filter to smooth spectral correlated signal. The concept is described originally in [46]. However, if operating the above low-pass filter on each entry vector independently, we will destroy the structure of the $M_r \times M_r$ matrix $\mathbf{R}_{II}^k$ that is Hermitian and positive definite (PD), and has $M_r(M_r + 1)/2$ parameters constrained. Next, we will address this issue by using Cholesky decomposition.

2. Cholesky Decomposition

In the area of multivariate statistics, it is a common approach to decompose the complicate covariance matrices into simpler components for further processing. There are three popular methods to use for matrix decomposition: variance-correlation decomposition, spectral decomposition (singular value decomposition (SVD)) and Cholesky

decomposition. While the entries of the correlation and orthogonal matrices in the variance-corrleation and spectral decompositions are still constrained, those in the lower triangle matrix of the Cholesky decomposition are always unconstrained. As a result, it becomes a unconstrained refinement if smoothing the Cholesky decomposition of spatial covariances across different tones instead of the covariance itself as in previous section, and the Hermitian and positive definite structure can be maintained. The low-triangle matrix of the Cholesky decomposition provides sufficient statistics for the covariance estimation, and can be written as:

$$\tilde{\mathbf{R}}_{II}^{k} = (\mathbf{U}^{k})^{H} \cdot \mathbf{U}^{k} \ , \tag{5.12}$$

where $\mathbf{U}^{k}$ is a upper triangle matrix, $\mathbf{U}^{k}$ is also called "square-root" of matrix $\tilde{\mathbf{R}}_{II}^{k}$;

Instead of filtering the entry vectors of $\tilde{\mathbf{R}}_{II}^{k}$, we now smooth that of upper triangle matrices $\mathbf{U}^{k}$. After the smoothing, we reconstruct the spatial covariance for each tone as $\hat{\mathbf{R}}_{II}^{k} = (\hat{\mathbf{U}}^{k})^{H} \cdot \hat{\mathbf{U}}^{k}$. Since the correlation among different tones still maintain for square-root matrix $\mathbf{U}$, we can use the filtering matrix $\mathbf{P}$ in equation (5.8). Other choice of smooth function might be possible, e.g., Kaiser-Bessel window. In our case, we observed that the matrix $\mathbf{P}$ provides good performance with wise choice of L. The algorithm is summarized in Table II.

In order to show the accuracy of different spatial covariance estimation approaches, we use three metrics to illustrate of the effect smoothing and Cholesky decomposition. Relative Frobenius norm, relative MSE of the eigenvalues and Stein's/entropy loss [49]. The relative accuracy with Frobenius norm is defined as:

$$\xi_{F}(\hat{\mathbf{R}}_{II}, \mathbf{R}_{II}) = \frac{\|\hat{\mathbf{R}}_{II} - \mathbf{R}_{II}\|_{F}}{\|\mathbf{R}_{II}\|_{F}} = \frac{\mathrm{trace}\{(\hat{\mathbf{R}}_{II} - \mathbf{R}_{II})(\hat{\mathbf{R}}_{II} - \mathbf{R}_{II})^{H}\}}{\mathrm{trace}\{\mathbf{R}_{II}\mathbf{R}_{II}^{H}\}} \tag{5.13}$$

where $\mathbf{R}_{II}$ is the theoretical spatial covariance, which is $\mathbf{H}\mathbf{H}^{H}$ for synchronous case.

Table II. Spatial Covariance Estimation Algorithm I

1. *Samples Average Estimation*: for $k = 1 \cdots K$, $\quad \tilde{\mathbf{R}}_{II}^{k} = \frac{1}{P} \sum_{n=0}^{P-1} \{\mathbf{I}^{k}(n)^{H} \mathbf{I}^{k}(n)\}$
2. *Cholesky Decomposition*: for $k = 1 \cdots K$, $\quad \tilde{\mathbf{R}}_{II}^{k} = (\tilde{\mathbf{U}}^{k})^{H} \cdot \tilde{\mathbf{U}}^{k}$
3. *Smoothing*: For each entry in $\tilde{\mathbf{U}}^{k}$, let $\tilde{\mathbf{u}} = \left[\tilde{\mathbf{U}}^{1}[m,n] \cdots \tilde{\mathbf{U}}^{K}[m,n]\right]^{T}$, $\quad \mathbf{v} = \mathbf{P} \cdot \tilde{\mathbf{u}}$, then $\mathbf{v} = \left[\hat{\mathbf{U}}^{1}[m,n] \cdots \hat{\mathbf{U}}^{K}[m,n]\right]^{T}$, $\quad$ construct $\hat{\mathbf{U}}^{k}$ from $\mathbf{v}$.
4. *Reconstructing*: Reconstruct the estimated covariance: $\hat{\mathbf{R}}_{II}^{k} = (\hat{\mathbf{U}}^{k})^{H} \cdot \hat{\mathbf{U}}^{k}$.

The relative MSE of eigenvalues is:

$$\xi_{E}(\hat{\mathbf{R}}_{II}, \mathbf{R}_{II}) = \frac{\sum_{i} (\lambda_{i,\hat{R}_{II}} - \lambda_{i,R_{II}})^{2}}{\sum_{i} (\lambda_{i,R_{II}})^{2}} \tag{5.14}$$

The Stein's/Entropy loss characterizes the relative distance of two Gaussian distributions.

$$\begin{aligned} L(\hat{\mathbf{R}}_{II}, \mathbf{R}_{II}) &= D\big(\mathcal{N}_{\mathbb{C}}(0, \mathbf{R}_{II}) \; | \mathcal{N}_{\mathbb{C}}(0, \hat{\mathbf{R}}_{II})\big) \\ &= \operatorname{trace}(\hat{\mathbf{R}}_{II} \mathbf{R}_{II}^{-1}) - \log\det(\hat{\mathbf{R}}_{II} \mathbf{R}_{II}^{-1}) - M_{r} \end{aligned} \tag{5.15}$$

We plot the numerical results for these three metrics in Fig. 16-18, which are the averaged values across different tones and channel realizations. From Fig. 16, we notice that for Frobenius norm, Cholesky decomposition based smoothing outperform both sample average and pure smoothing. In Fig. 17, we found the smoothing without Cholesky decomposition deviate the eigenvalue from the theoretical one with a large value and even worse than sample average. This demonstrates the pure covariance smoothing will destroy the spatial covariance matrices structure.

In Fig 18. we plot the percentage of reduction in average loss compared with the

sample mean covariance estimation:

$$PRAL = 100\frac{L_{sa}(\hat{\mathbf{R}}_{II}, \mathbf{R}_{II}) - L_{chol}(\hat{\mathbf{R}}_{II}, \mathbf{R}_{II})}{L_{sa}(\hat{\mathbf{R}}_{II}, \mathbf{R}_{II})} \tag{5.16}$$

It shows more than 50 percent reduction in the stein's loss for low interference v.s. noise ratio with the proposed spatial covariance estimation method. However, for less noisy case, the improvement decreases to around 20 percent. Therefore, our spatial covariance estimation algorithm provide substantial gain for noisy estimation environment.

The Cholesky decomposition method has been used in [50] for simultaneous estimation of several covariance matrix. It was also shown that the estimation of a covariance matrix is equivalent to estimating a squence of varying-coefficient and varying-order regression models with unconstrained coefficients.

D. Interference Aware Receiver Design

The enhanced parsimonious spatial covariance algorithm proposed the above allows for better estimation of the statistics of asynchronous co-channel interference. In this section, we will design an interference-aware receiver to suppress CCI by utilizing the estimated statistics, shown as Fig. 19. First, we use the classical Wiener filter, i.e., MMSE. The problem is invariant to the choice of OFDM tone k, without loss of generality, we will suppress the tone index - k.

1. MMSE Receiver for Co-channel Interference Mitigation

Denote the MMSE filter as $\mathbf{W}$,

$$\mathbf{W} = \mathbf{R}_{yy}^{-1}\mathbf{R}_{xy}^{H}\ , \tag{5.17}$$

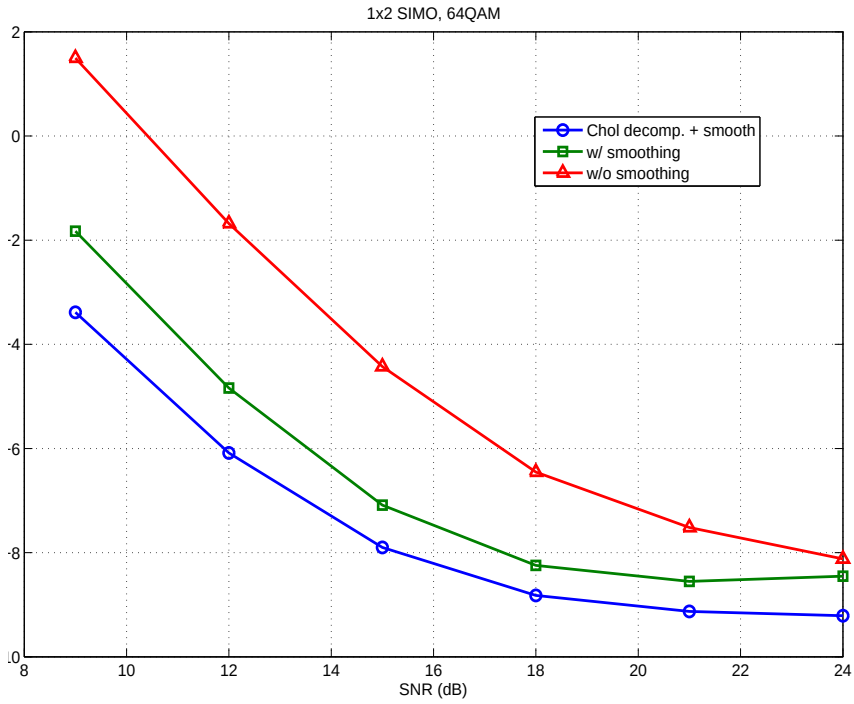

Fig. 16. Relative estimation accuracy for spatial covariance (F-norm)

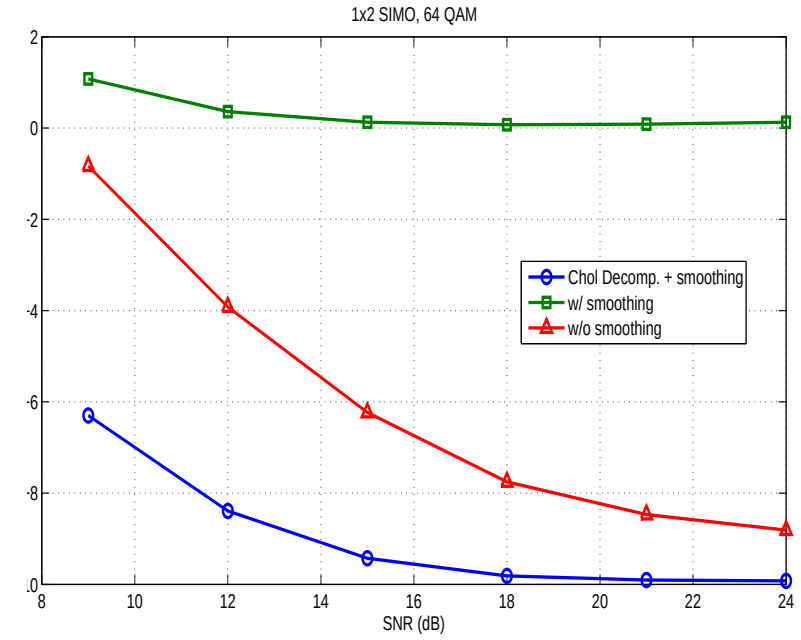

Fig. 17. Relative estimation accuracy for eigenvalue of spatial covariance

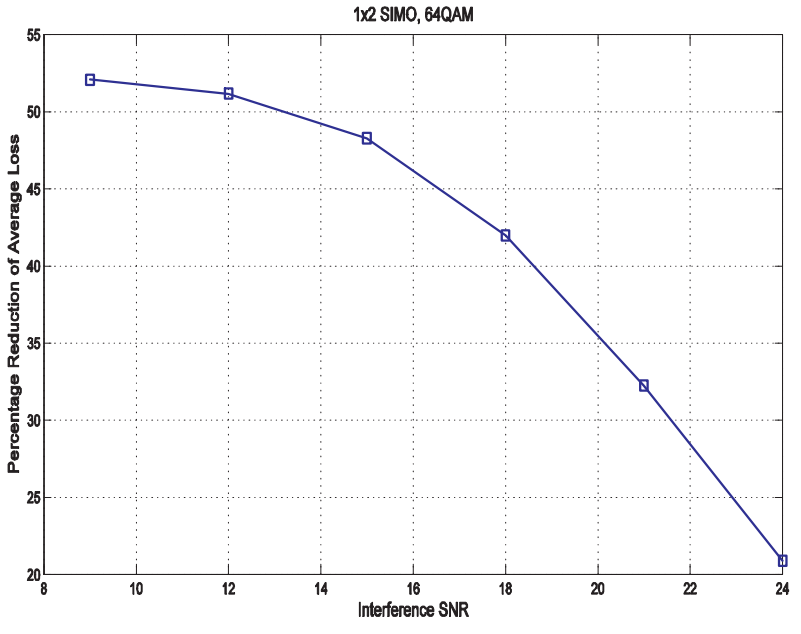

Fig. 18. Percentage reduction in the Stein's loss

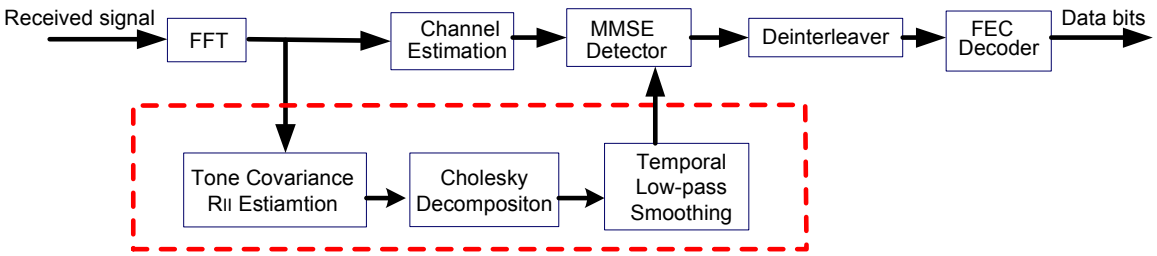

Fig. 19. Receiver structure

where $\mathbf{R}_{xy} = E\{\mathbf{x}(n)\mathbf{y}^H(n)\} = \mathbf{H}$, and $\mathbf{R}_{yy} = E\{\mathbf{y}(n)\mathbf{y}^H(n)\}$. In order to obtain the MMSE filter, we need to estimate the channel of desired signal $\hat{\mathbf{H}}$ and the covariance of received signal $\hat{\mathbf{R}}_{yy}$. The simplest way to estimate $\mathbf{R}_{yy}$ is to average the received signal vectors over a period of time. However, such method does not explore the structure of received signal efficiently and can not provide accurate covariance estimation even with long-time average, especially for high-order modulations. Our simulation results have verified this argument (not shown in this paper due to limited space) and show poor performance of 16QAM or above. Due to the independence among desired signal, interference and noise. We can rewrite the covariance of the received signal as:

$$\mathbf{R}_{yy} = \mathbf{H}\mathbf{H}^H + \mathbf{R}_{II} \ . \tag{5.18}$$

For the synchronous case, we have $\mathbf{R}_{II} = \mathbf{G}\mathbf{G}^H + N_0\mathbf{I}_{M_r}$, where $\mathbf{I}_{M_r}$ denotes the identity matrix and $\mathbf{G}$ indicates the channel response of interference. For the asynchronous case, we will use the algorithm proposed in the previous section to estimate $\mathbf{R}_{II}$. The receiver structure is shown in Fig. 14.

2. Enhancements for Space-time Block Coded (STBC) System

Space-time coding has recently emerged as a powerful approach to exploit the spatial diversity and combat fading in MIMO wireless communications systems. For sim-

plicity, we use the Alamouti code as an example, which has been adopted as one option by the next generation WLAN standard, e.g. 802.11n, and the analysis can be easily extended to other OSTBC. Modify the signal model in (5.4) to incorporate the space-time code. We can rewrite the received signal as:

$$\underbrace{\begin{pmatrix} y_1(n) \\ \vdots \\ y_{M_r}(n) \\ y_1^*(n+1) \\ \vdots \\ y_{M_r}^*(n+1) \end{pmatrix}}_{\mathbf{Y}} = \underbrace{\begin{pmatrix} h_{11} & h_{12} \\ \vdots & \vdots \\ h_{M_r1} & h_{M_r2} \\ h_{12}^* & -h_{11}^* \\ \vdots & \vdots \\ h_{M_r2}^* & -h_{M_r1}^* \end{pmatrix}}_{\tilde{\mathbf{H}}} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \underbrace{\begin{pmatrix} I_1 \\ \vdots \\ I_{M_r} \\ I_{M_r+1} \\ \vdots \\ I_{2M_r} \end{pmatrix}}_{\mathbf{I}} \tag{5.19}$$

Basically, we stacked the received signal vectors from time n and $n+1$ as one vector. $\mathbf{I}$ is the asynchronous co-channel interference, which is a space-time coded signal plus noise. If the intended and interference signals are synchronized (both for OFDM cyclic structure and orthogonal space-time modulation), we will have $2M_r - 2$ extra degree freedom. However, for random asynchronous interference, the term $\mathbf{I}$ is unstructured. Not only the degree of freedom is insufficient, but also we need double the dimention of the "spatial-temporal" covariance estimation. And even with the improved covariance estimation techniques in the previous section will not be able to provide good CCI suppression performance. Intuitively, it can be explained as the result of asynchrony making the space-time coded CCI acts as if we have a $2M_r \times 2M_r$ interference MIMO spatial multiplex transmission. Unfortunately, we only have $2M_r$ degree of freedom all together at the receiver, hence it is impossible to suppress the interference signal effectively.

Here, we propose a heuristic solution that is to block diagonalize the covariance matrix by zero-forcing the cross correlation information between two successive receive signal vectors from time n and $n+1$. The assumption is that these two signal vectors are separated far enough to be treated independently. More precisely, we write $\mathbf{R}_{II}$

can be write as:

$$\mathbf{R}_{II} = \begin{pmatrix} \mathbf{R}_{II}(n) & \mathbf{0} \\ \mathbf{0} & \mathbf{R}^*_{II}(n+1) \end{pmatrix} \tag{5.20}$$

Such process not only reduces the amount of estimation parameters by half, but also saves degree of freedom. The covariance matrix $\mathbf{R}_{II}(n)$ and $\mathbf{R}^*_{II}(n+1)$ can be estimated as previous by matrix decomposition and smoothing. Finally, for space-time codes system, we can use the MMSE receiver to suppress the interference by treat the STBC as an equivalent spatial multiplexing transmission with channel matrix as $\tilde{\mathbf{H}}$.

3. Bound of the Mean Square Error (MSE)

After the MMSE demodulator, the MSE is computed with the estimated spatial covariance,

$$\hat{MSE} = (1 + \mathbf{H}^H \hat{\mathbf{R}}_{II}^{-1} \mathbf{H})^{-1} \ . \tag{5.21}$$

And the post-equalizer SNR can be write as:

$$\hat{SNR}_m = \frac{1}{\hat{MSE}_{m,m}} - 1 \ . \tag{5.22}$$

The post-equalizer SNR for each data streams will be used to compute the soft informations of each bit, which is in turn used by the Viterbi decoder (convolutional code) or iterative message passing decoder (LDPC code) to decode the packet. Hence, the post-equalizer MSE will determine the receiver performance. In this section, we want to characterize the relationship of the estimated $\hat{MSE}$ and the true MSE. In order to demonstrate the impact of spatial covariance estimation on the receiver performance, we focus on the SIMO case, for which the intra-user interferences do not exist. Generalizing to MIMO case is straightforward.

Theorem 1. *The MSE of the MMSE receiver with the estimated spatial covariance is upper bounded by:*

$$\mathrm{MSE} \leq \hat{\mathrm{MSE}} + \left(\hat{\mathrm{MSE}} - \hat{\mathrm{MSE}}^2\right) \cdot \|\Delta \mathbf{R}_{II}\|_F \cdot \lambda_{max}(\hat{\mathbf{R}}_{II}^{-1}). \tag{5.23}$$

Proof. See Appendix A. □

The MSE upper bound has important implication. The smallest eigenvalue of $\hat{\mathbf{R}}_{II}$ has significant effect on MSE given the total estimation error $\|\Delta \mathbf{R}_{II}\|_F$. Sample mean ML estimator has the tendency to spreading the eigenvalues, therefore, it will increase the MSE of the equalizer.

4. Complexity Analysis

The complexity of MMSE receiver is well-known. At the beginning of each packet, we need a $M_r \times M_r$ matrix inversion, which is in the order of M_r^3 (can be reduce by some advance algorithms). In addition, the matrix $\mathbf{W}$ is the product of a square matrix of size $M_r \times M_r$ and a $M_r \times M_t$ matrix. The complexity of such a product is proportional to $M_t M_r^2$. For each received signal vector, the MMSE equalizer operation has a complexity of $M_t M_r$. The total complexity is multiply by K, since we have similar processing in each subcarrier. Next, we will focus on the complexity of the spatial covariance estimation. For the samples average operation, we have $PM_r(M_r+1)/2$ multiplications and $(P-1)M_r(M_r+1)/2$ additions given the P training received vectors are used. The Cholesky decomposition has a complexity of $M_r^3/3$. For the smoothing operation, we can eigher pre-computed smooth matrix and stored or use two FFT transform. For pre-computed and stored smoothing matrix, a K^2 matrix product operation is required. However, if we use the 2 FFT operations instead, the complexity is $2K \log K$ for all $M_r(M_r+1)/2$ entries. All together, the spatial

covariance estimation complexity at the beginning of decode the packet:

$$\frac{KPM_r(M_r+1)}{2}+\frac{KM_r^3}{3}+KM_r(M_r+1)\log K\ .$$

5. MAP Receiver for Co-channel Interference Suppression

Since we have model the interference as Gaussian random process with zero mean and covariance $\mathbf{R}_{II}$, we can derive the optimum MAP bit detector to minimize bit error probability. Let's assume a block of $M_t \log_2 C$ bits $\mathbf{b}$ have been transmitted per channel use for each tone, where C is the modulation constellation size. The *a psoteriori* log-likelihood ratio value (L-value) of bits $b_i, i = 0, \cdots, M_t \log_2 C - 1$, conditioned on the received vector $\mathbf{y}$ is

$$L_D(b_i|\mathbf{y}) = \ln \frac{P[b_i = +1|\mathbf{y}]}{P[b_i = -1|\mathbf{y}]}\ . \tag{5.24}$$

Assuming the $\{b_i\}$ are independent due to the random interleaver, (5.24) can be further expressed as:

$$L_D(b_i|\mathbf{y}) = L_A(b_i) + \ln \frac{\sum_{\mathbf{x}\in\mathbb{X}_{i,+1}} P[\mathbf{y}|\mathbf{x}] \cdot e^{\sum_{j\in\mathbb{J}_{i,x}} L_A(b_j)}}{\sum_{\mathbf{x}\in\mathbb{X}_{i,-1}} P[\mathbf{y}|\mathbf{x}] \cdot e^{\sum_{j\in\mathbb{J}_{i,x}} L_A(b_j)}}. \tag{5.25}$$

where $\mathbb{X}_{i,+1}$ is the set of $2^{M_t \log_2 C-1}$ bit vectors $\mathbf{x}$ having $b_i = +1$, $\mathbb{X}_{i,-1}$ is the set of $2^{M_t \log_2 C-1}$ bit vectors $\mathbf{x}$ having $b_i = -1$ and $L_A(b_j) = \ln \frac{P[b_j=1]}{P[b_j=-1]}$, is the *a priori* L-value. $\mathbb{J}_{k,x}$ is the set of indices j with

$$\mathbb{J}_{i,x} = \{j|j = 0, 1, 2, M_t \log_2 C - 1, j \neq i, b_j = 1\}. \tag{5.26}$$

The second term on the RHS of (5.25) is the extrinsic L-value, defined as $L_E(b_i|\mathbf{y})$ and used below. The sets $\mathbb{X}_{i,+1}$ and $\mathbb{X}_{i,-1}$ can be either generate by exhaustive listing for small antenna number and lower modulation order, or generate by the list sphere decoding for large antenna number and higher order modulation [8].

To compute the L-value (5.25) for the MAP detector, one essential step is to compute the likelihood function $P(\mathbf{y}|\mathbf{x})$, this can be found from the linear model of (5.4) and the estimated $\hat{\mathbf{R}}_{II}$

$$\begin{aligned} P[\mathbf{y}|\mathbf{x} = \text{map}(\mathbf{b})] &= \frac{1}{\pi^{M_r} \det(\mathbf{R}_{II})} \exp[(\mathbf{y} - \mathbf{H}\mathbf{x})^H \hat{\mathbf{R}}_{II}^{-1} (\mathbf{y} - \mathbf{H}\mathbf{x})] \\ &= \frac{1}{\pi^{M_r} \det(\mathbf{R}_{II})} \exp[||\hat{\mathbf{U}}^{-1}(\mathbf{y} - \mathbf{H}\mathbf{x})||^2] \ , \end{aligned} \tag{5.27}$$

where $\hat{\mathbf{U}}$ is the Cholesky decomposition of the $\hat{\mathbf{R}}_{II}$.

Using the max-log approximation, the extrinsic L-value can be approximated as

$$\begin{aligned} L_E(b_i|\mathbf{y}) \approx & \max_{\mathbf{x} \in \mathbb{L}_{i,+1}} \{ -\frac{||\hat{\mathbf{U}}^{-1}(\mathbf{y} - \mathbf{H}\mathbf{x})||^2}{N_0} + \frac{1}{2} \mathbf{b}_{[i]}^T \mathbf{L}_{A,[i]} \} \\ & - \max_{\mathbf{x} \in \mathbb{L}_{i,-1}} \{ -\frac{||\hat{\mathbf{U}}^{-1}(\mathbf{y} - \mathbf{H}\mathbf{x})||^2}{N_0} + \frac{1}{2} \mathbf{b}_{[i]}^T \mathbf{L}_{A,[i]} \}, \end{aligned} \tag{5.28}$$

where $\mathbf{b}_{[i]}$ denotes the sub-vector of $\mathbf{b}$ omitting its i^{th} element, and $\mathbf{L}_{A,[i]}$ is the vector of all L_A values, also omitting its i^{th} element. The MAP detector can output soft information and iterative (Turbo) exchange the extrinsic information with outer channel decoder to improve the performance. The complexity of the MAP detector is higher than the MMSE receiver. Basically, the square-root matrix $\mathbf{U}$ act as a pre-whitening filter to whiten the interference signal. Hence, an accurate spatial covariance estimation is also desirable for MAP detector.

E. Simulation Results

In this section, we provide simulation results to show the effectiveness of proposed spatial covariance estimation algorithms and the CCI suppression receiver. Simulation parameters are shown in Table III. Gray mapping is used for transmitted symbol modulation. Our performance metric is packet error rate (PER). A standard OFDM symbol level interleaver is used to combat frequency selectivity for the indoor multi-

Table III. Simulation Parameters

Number of subcarriers	64
Number of data subcarriers	48
Cyclic prefix length	16
FEC code	rate $\frac{1}{2}$ LPDC
Codeword length	1944
Channel model	TGn model D [51]
Packet size	972 bytes
Asynchrony	$1 - 80$ timing samples, uniform distribution
Training zero-padding	$4 - 6$ OFDM symbols
Interferer	1 interferer of NLOS

path scattering channel. For simplicity, we assume the desired signal and interference use the same transmission setup.

For the desired signal, we assume there is a line-of-sight (LOS) component in the first tap. But for interference signal, which usually located faraway (distance $>$ 10m 'breakpoint'), it only has NLOS path. The LOS path of the desired signal has an angle-of-arrival (AOA) of $\pi/4$. We assume one dominant co-channel interferer exists (For most situations, $1-2$ strong interferers are typical). We estimated the interference at the beginning of the packet decoding with zero-padding P OFDM symbols during transmission (i.e., the P OFDM signals consist of only interference and noise). Spatial covariances are parsimonious estimated by these P OFDM symbols.

We first assume 1×2 SIMO case. For which APs have legacy one transmitter antenna, but the mobile terminals have two receive antennas. Fig. 20 compares different receiver schemes. The SNR value is fixed for 20 dB. The desired signal's channel is estimated under the interference environment using one OFDM training symbols by the method in [52, 53]. Note that for SIR larger than 5 dB, channel estimation can achieve the required accuracy for most cases. 4 OFDM symbols zero-padding duration are used to estimate the spatial covariance. From the plot, we can observe 8 dB gain in SIR compared to MRC receiver for PER of 10^{-2}. The tone smoothing provides 1 dB gain and cholesky decomposition provides further 2 dB gain compared with the MMSE receiver without covariance estimation refinement. As a benchmark, we also show the results for the MMSE receiver for synchronized and clairvoyant interferer case (perfect know interfer's channel ($\mathbf{G}$)) - dot line in the figure. Interestingly, there is less than 2dB gap between our proposed method and the synchronized case.

In Fig. 21., high-order modulation 64QAM is used for different receiver schemes. 64QAM is the worst for estimating the spatial covariance. The proposed algorithm

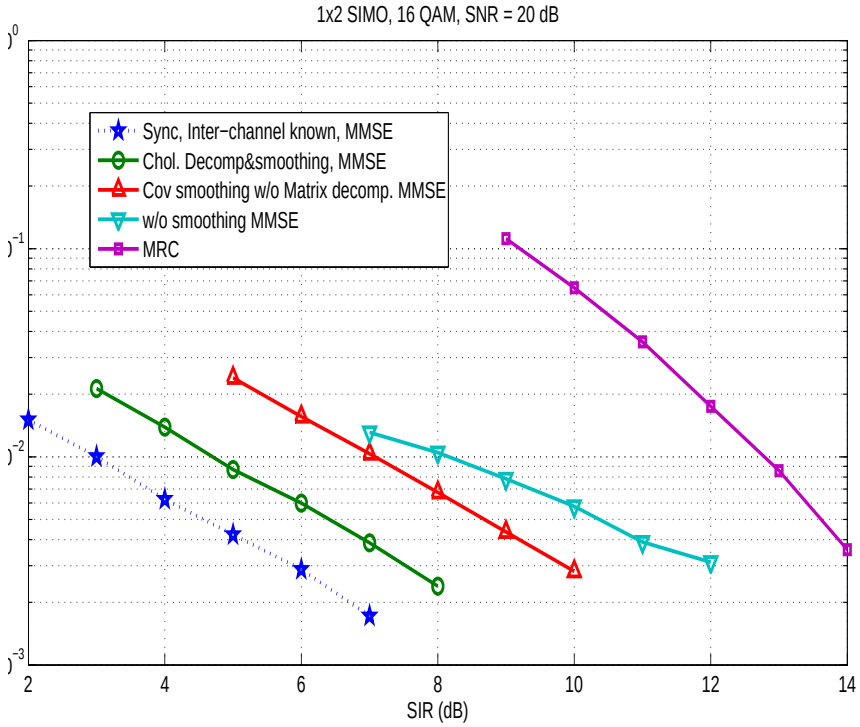

Fig. 20. Packet error rate of different receivers for 1x2 SIMO, 16 QAM, MMSE receiver

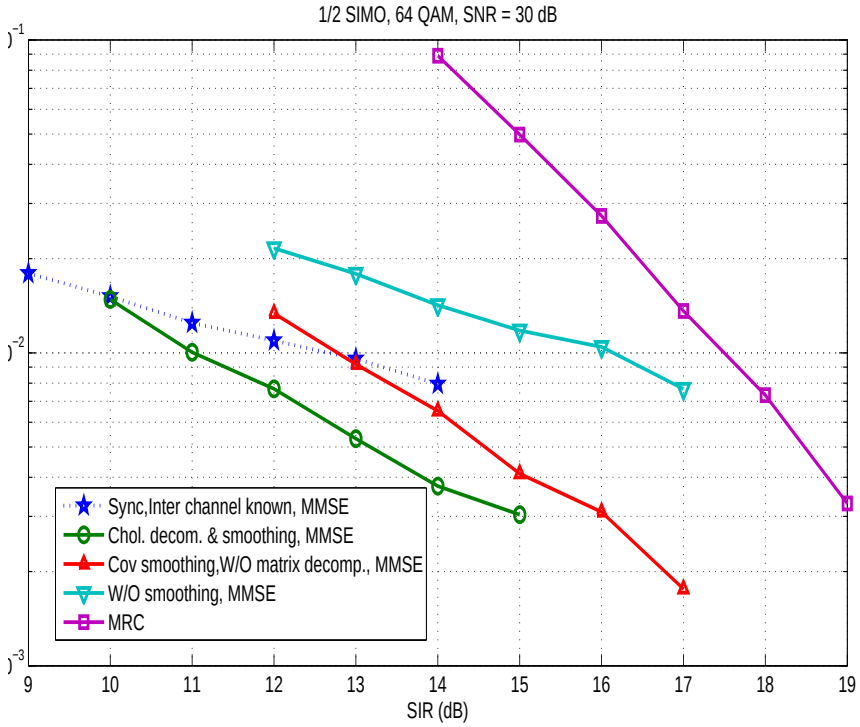

Fig. 21. Packet error rate of different receivers 1x2 SIMO 64QAM, MMSE receiver

still can provide a 6.5 dB gain compared with MRC receiver for PER of 10^{-2}. For 64QAM modulation, covariance smoothing can provided 3 dB gain compared with non-smoothing MMSE, and the Cholesky decomposition improves the PER curve by a further 2 dB gain. Again, the channel is estimated under the co-channel interference. Surprisingly, our proposed scheme perform even better than the synchronous clairvoyant curve - dot line.

Next, we consider the space-time coded transmission system. We use the Alamouti code with 2 transmitter antennas and 3 received antennas. We compared the PER of block diagonalized scheme and without the diagonalization. 6 OFDM symbols are used to zero pad for diagonalized case and 12 OFMD symbols for undiagonalized MMSE receiver in the covariance estimation. As we pointed out, MMSE without di-

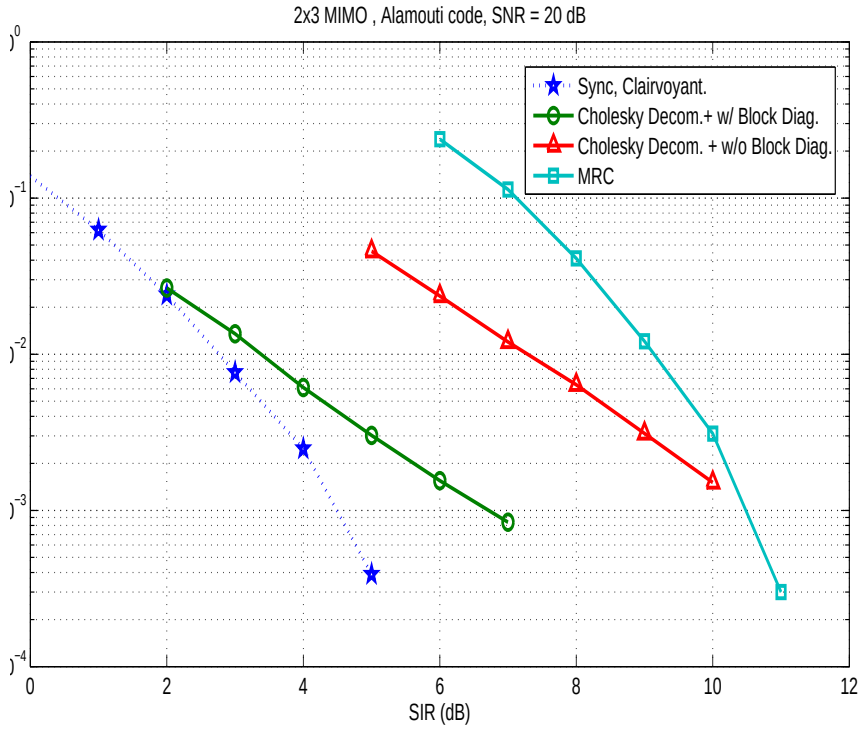

Fig. 22. Packet error rate for space-time coded system, 2x3 MIMO 16 QAM, MMSE receiver

agonization have freedom deficiency problem, which deteriorates the MMSE receiver performance. With diagonalization of $\mathbf{R}_{II}$. The proposed the algorithm approach the synchronized, interference channel perfect know curve (doted curve). Noticeably, MRC has better diversity gain, though the proposed the scheme provided 6 dB gain in SIR for PER of 10^{-2}. Therefore, the interference suppression will sacrifice the diversity gain of the space-time code. Also, from the Fig. 22., we notice that the diversity gain of synchronized interference case is better than asynchronous interference mitigation. We further demonstrate CCI suppression for space-time coded system in Fig. 23. with 64QAM modulation.

We plotted the MAP decoder with/without iteration and compared with the MMSE receiver in Fig. 24. In order to eliminate the effect of channel estimation

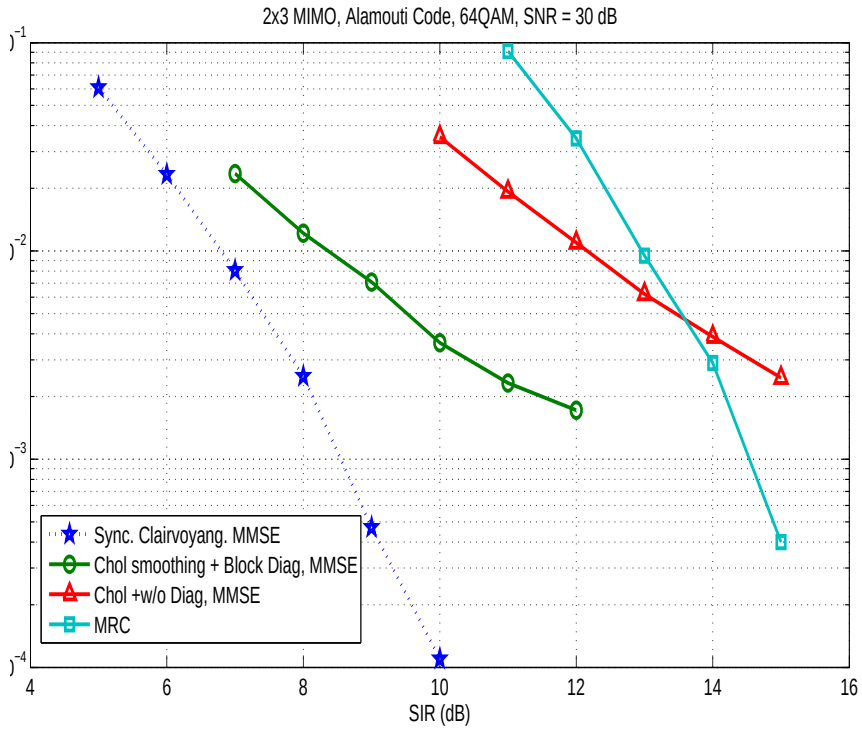

Fig. 23. Packet error rate for space-time coded system, 2x3 MIMO 64 QAM, MMSE receiver

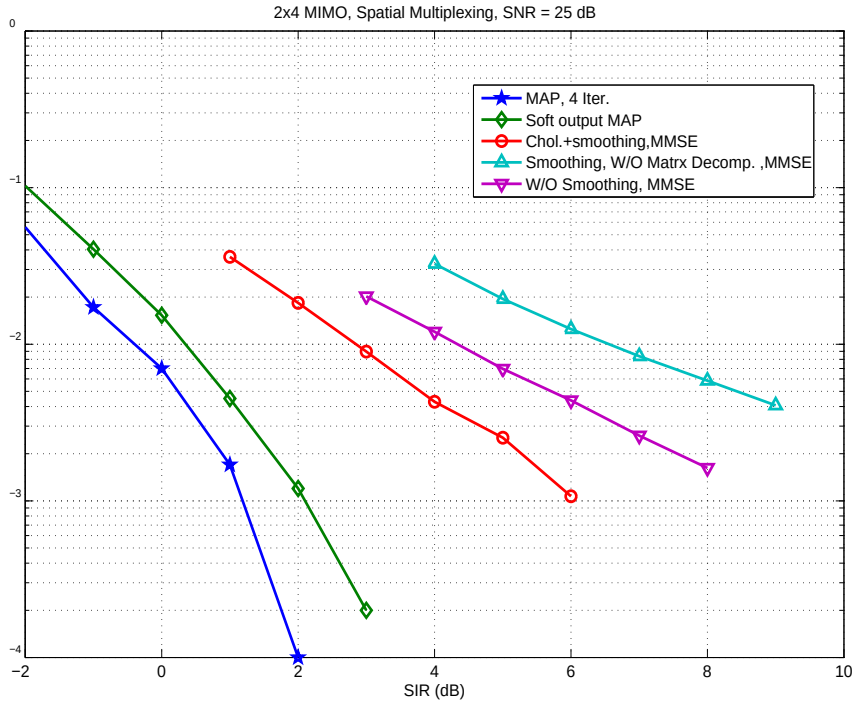

Fig. 24. Packet error rate for 2x4 MIMO system, 16QAM, MAP receiver

for low SIR. We assume the perfect channel knowledge of the desired signal. And 6 zero-padding OFDM symbols are used for interference statistics estimation. The MAP demodulator and LDPC decoder iterative exchange the extrinsic information. The iteration can provide a marginal gain of 1 dB compared with the MAP soft output demodulator. If the packet is successfully decoded, then the iterative process will be abort. By doing so, the receiver runs $1-2$ iterations for most cases. Apparently, the MAP demodulator provides more diversity gain in contrast to MMSE receiver. Surprisingly, without matrix decomposition, the smoothed MMSE perform even worse than the conventional MMSE. This can be explained as the result of the smoothing (low-pass filtering) destroys the covariance matrix structure, especially for larger matrix size (4 receiver antennas).

F. Conclusion

We have presented an efficient spatial covariance estimation method for MIMO OFDM system. The proposed method consists of a Cholesky decomposition step and a smoothing operation across OFDM tones. The algorithm can significantly improve the performance of a interference-aware receiver, demonstrated by the SIR gains for PER curves. We also designed the MMSE and MAP receiver based on the proposed interference statistic estimation method. The MAP receiver achieves better performance but at cost of higher complexity, compared with the MMSE one. The proposed algorithm has been tested in Intel's MIMO RF chain testbed and acquired the preliminary results.

In the future work, we will apply our schemes to high density WLAN, where we should consider partial interference and the "capture" effect. In other word, only part of the packet is interfered by CCI, hence there is a statistics "mismatch" problem. One potential solution to investigate is zero-padding in multiple positions and partitioning the packet into smaller blocks, which is a natural result if short length LDPC code is used.

CHAPTER VI

ERROR EXPONENT OF THE WIDEBAND RELAY CHANNEL

A. Introduction

Relayed transmission has received increasing attention as it can provide distributed space diversity to combat the fading impairment in the wireless network. The classical relay channel was introduced by van der Mulen [4], and then further explored by Cover and El Gammal [54]. Laneman et al., [55] analyzed the outage behavior and diversity order for several relay protocols. Their results characterized the diversity multiplexing trade-off at the high signal-to-noise ratio (SNR). Recently, Liang and Veeravalli [56] studied the optimal resource allocation problem for the Gaussian orthogonal relay channel. However, most previous work has primarily focused on narrow-band relay transmission, where the received SNR per degree of freedom is high. In this chapter, we study the performance of the relay channel in the wideband extreme, i.e., the available bandwidth is large and the resulting SNR per degree of freedom is low. Relevant examples are wireless ad-hoc, sensor networks and 802.16j. In [57], Verdu has investigated the spectral efficiency in wideband regime for general wireless channel.

We use Gallager's random code error exponent [58] (also known as the channel reliability function) as a tool to analyze different relay strategies. Error exponent provides a measure of how fast the decoding error probability decays exponentially as the code block length increases for rates below channel capacity. We show that, for orthogonal relaying, both AF and DF provide higher reliability than the direct transmission, and the DF scheme has better performance than AF for similar settings. If we relax the orthogonal constraint, i.e., the relay node can receive and transmit message at the same time (full duplex), block Markov coding scheme can be used to

boost the link reliability even more. The error exponent can serve as a performance measure to optimize the power allocation and relay node placement. We found that placing relay node in the middle between source and destination can provide the best link reliability for DF and BMC schemes. But for the AF scheme, the optimal position depends on the path-loss exponent of the physical wireless propagation model. Most results of this chapter are included in [59].

The remainder of this Chapter is organized as follows. Section B introduces the system model for the problem under consideration. Section C defines the the random coding error exponent. Section D and Section E give out the error exponent results for various relay strategies. Some numerical results are provided in Section F. Section G summarizes the main results of the paper.

B. System Model

In this work, communication occurs over a relay network, with one relay node and one source-destination pair, as is shown in Fig. 25. The source S broadcasts the message to both relay R and destination D. The relay processes the message and then sends it to the destination to assist the destination decoding the data. Based on the limitation of relay node, we focus on two kinds of relay: 1) orthogonal relay (half-duplex), i.e., transmitting and receiving in the different time or frequency sub-channels. The AF and DF schemes fall into this category. 2) Full duplex operation, including block Markov coding transmission. We model the wideband channel as a set of N parallel narrowband channels. We assume that the Doppler spread is negligible, which makes the narrowband channels have independently and identically distributed statistics (i.i.d.). Moreover, we assume that the coherent bandwidth is much larger than the bandwidth of the narrowband, such that each channel is flat faded. Using

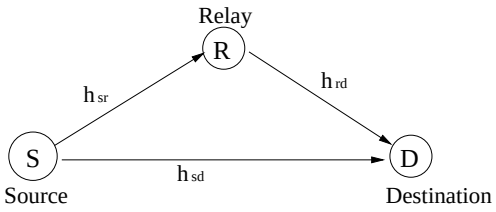

Fig. 25. Layout of relay network.

the sampling theorem, the received signal at the relay and the destination for n^{th} channel and k^{th} symbol time can be written, respectively, as

$$y_r[k,n] = \sqrt{\frac{P_s}{N}} h_{s,r}[k,n] x_s[k,n] + z_r[k,n] \tag{6.1}$$

$$y_d[k,n] = \sqrt{\frac{P_s}{N}} h_{s,d}[k,n] x_s[k,n] + \sqrt{\frac{P_r}{N}} h_{r,d}[k,n] x_r[k,n] + z_d[k,n] \ , \tag{6.2}$$

where $x_i[k,n]$ is the source/relay transmitted signal with $i \in \{s,r\}$. We assume $E[|x_i[k,n]|^2] = 1$, and let the transmit power at source and relay be P_s and P_r respectly. In $(1)-(2)$, $h_{i,j}[k,n]$ is the fading coefficient, where $i \in \{s,r\}$ and $j \in \{r,d\}$; $z_j[k,n]$ represents the additive white noise for $j \in \{r,d\}$. The pair (k,n) can be considered as index for time-frequency slot, or degree of freedom, to communicate. Statistically, we model the $h_{i,j}[k,n]$ as zero-mean, circularly-symmetric complex Gaussian random variables, which are independent across different narrowband channels and links. Additionally, we model $z_j[k,n]$ as zero-mean, independent, circularly-symmetric complex Gaussian random variables with variances N_0.

In this work, we simplify the model in Fig. 25. We assume that the distance between the source and destination is normalized to one, and the relay is located in a line between the source and destination. The parameter d represents the distance from source to relay, and $1-d$ is the distance from the relay to the destination. Using

physical path-loss propagation model for wireless communication [60], we assume $E[|h_{i,j}|^2] = \frac{1}{d_{i,j}^{\alpha}}$, where $d_{i,j}$ is the distance from transmitter i to receiver j, α is path-loss exponent.

Furthermore, we assume there is no decoding delay and coding is across the different narrowband channels. Our goal is to compute the error exponent of this wideband relay transmission, and to study the optimal power allocation and relay placement. Since we assume i.i.d. statistics across the narrowband channels, we can aim at one narrowband channel with source and relay power constraint $(\frac{P_s}{N}, \frac{P_r}{N})$. As N tends to ∞, the power allocated to each narrowband channel goes to 0. Equivalently, we can focus on analyzing a narrowband flat fading channel in the low SNR regime. For convenience, we omit the narrowband index n. With a little abuse of notation, let (P_s, P_r) represent the transmit power at source/relay for each narrowband channel, which can take a very small value.

C. The Random Coding Error Exponent

Gallager [**?**] established random coding techniques to upper-bound the achievable average error probability over a random code ensemble with maximum-likelihood decoding. Specifically, given a code $\mathbb{C}$ of length N over an alphabet χ with 2^{nR} codewords, we have

$$\bar{P}_e \leq \exp\left(-N(E_0(\rho, Q) - \rho R)\right) , \tag{6.3}$$

with $E_0(\rho, Q)$ defined as

$$E_0(\rho, Q) = -\ln\left(\int_{-\infty}^{\infty} \Big[\int_{-\infty}^{\infty} Q(\mathbf{X}) f(\mathbf{Y}/\mathbf{X})^{1/(1+\rho)} d\mathbf{X} \Big]^{1+\rho} d\mathbf{Y} \right) , \tag{6.4}$$

for $0 \leq \rho \leq 1$. $Q(\mathbf{X})$ is the code symbol (or input) distribution and $f(\mathbf{Y}/\mathbf{X})$ is the channel output distribution conditioned on the input. The random coding exponent

is defined to be the one that yields the tightest bound:

$$E_r(R) = \max_{\rho} \max_{Q} \{E_0(\rho, Q) - \rho R\} \ , \tag{6.5}$$

where the maximization is over Q and subject to the input power constraint. For linear Gaussian Channel model

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z} \ , \tag{6.6}$$

if we assume input symbol $\mathbf{x}$ has Gaussian distribution $\mathbf{x} \sim CN(0, \mathbf{P})$[1], and noise $\mathbf{z}$ has circular symmetric gaussian distribution $\mathbf{z} \sim CN(0, \mathbf{W})$. Substituting $Q(\mathbf{x})$ and $f(\mathbf{y}/\mathbf{x})$ into (6.5), we can get the following theorem.

Theorem 2. *(*Gaussian Error Exponent*) of (6.6):*

$$E_0(\rho, \mathbf{P}) = \rho \ln \mathbf{E}_H \left| \mathbf{I} + \frac{1}{1+\rho} \mathbf{W}^{-1} \mathbf{H} \mathbf{P} \mathbf{H}^{\dagger} \right| \ , \tag{6.7}$$

where $\mathbf{E}$ *denotes expectation, and* $|\cdot|$ *represents determinant of matrix.*

We omit the proof of this theorem because the result can be found in other literature [61]. If the channel model (6.6) reduces to the scalar one, Eq. (6.7) can be written as $E_0(\rho, P) = \rho \ln \mathbf{E}(1 + \frac{P|h|^2}{N_0(1+\rho)})$, which is the well known error exponent for the scalar fading channel.

D. Error Exponent for Orthogonal Relay Channel

For orthogonal relay operation, the relay node can not transmit and receive at the same time. We partition the transmission as two steps. First, source S broadcasts

[1]To choose $Q(x)$ as Gaussian is not optimal and a distribution concentrated on a "thin spherical shell" will give better results [**?**], nonetheless Gaussian error exponent is a convenient lower bound for the optimal error exponent.

message and relay R keeps silent, i.e.,let $x_r[k,n] = 0$ in Eq. $(1)-(2)$. In the next step, relay R transmits the processed message to destination and source S stops transmission. Mathematically, for the first step, the received signal of each equivalent narrowband channel can be written as

$$y_r = \sqrt{P_s} h_{s,r} x_s + z_r \tag{6.8}$$

$$y_d[1] = \sqrt{P_s} h_{s,d} x_s + z_d[1] \ , \tag{6.9}$$

For the next step, we obtain

$$y_d[2] = \sqrt{P_r} h_{r,d} x_r + z_d[2] \ . \tag{6.10}$$

1. Amplify-and-Forward (AF) Relay

Using the amplify-and-forward relay scheme, the relay node amplifies the message it received in the first phase and forwards it to the destination in the second phase, i.e.,

$$\sqrt{P_r} x_r = \beta y_r \ , \tag{6.11}$$

here we define the amplifier gain as $\beta = \sqrt{\frac{P_r}{P_s h_{sr}^2 + N_0}}$. Substitute (6.11) into (6.10) and write the received signal during the two phases in vector form

$$\underbrace{\begin{pmatrix} y_d[1] \\ y_d[2] \end{pmatrix}}_{\mathbf{y}} = \underbrace{\begin{pmatrix} h_{s,d} \\ h_{r,d}\beta h_{s,r} \end{pmatrix}}_{\mathbf{H}} \sqrt{P_s} x_s + \underbrace{\begin{pmatrix} 0 & 1 & 0 \\ h_{r,d}\beta & 0 & 1 \end{pmatrix} \begin{pmatrix} z_r \\ z_d[1] \\ z_d[2] \end{pmatrix}}_{\mathbf{z}}$$

Note that

$$E(\mathbf{z}\mathbf{z}^{\dagger}) = \begin{pmatrix} N_0 & 0 \\ 0 & |h_{r,d}\beta|^2 N_0 + N_0 \end{pmatrix} \ .$$

We observed that the AF relay is equivalent to a single-input-multiple-output (SIMO) channel. Using Theorem 2, we have following result.

Theorem 3. *(*Error Exponent of AF relay*) :*

$$E_r^{AF}(R) = \max_{0\le\rho\le 1}\left\{\frac{1}{2}\rho\ln\mathbf{E}\left(1+\frac{P_s}{(1+\rho)N_0}\left(h_{s,r}^2+h_{s,d}^2 - \frac{h_{s,r}^2}{|h_{r,d}\beta|^2+1}\right)\right)-\rho R\right\}. \tag{6.12}$$

For fair comparison with direct transmission, we have halved the degree of freedom and doubled the rate as $2R$ to account for the half-duplex transmission.

If we fix the total power budget as P, our goal is to optimize the power allocation (P_s, P_r) between source and relay transmission to maximize the error exponent of (6.12). Also, we try to find the optimal position in the line between source and destination to place the relay node. For wideband AF relay system, we assume the amplifier coefficient β takes the same value for all the parallel narrowband channel. Practically, it is a reasonable assumption and need not use passband filters for each narrow band channel.

Let us define $SNR = \frac{P^{dir}}{N_0 W}$, where P^{dir} is the direct transmission power in each channel use, and W is bandwidth of each narrowband channel. Then we have $P_s = 2\,SNR\gamma, P_r = 2\,SNR(1-\gamma)$, where $0 \le \gamma \le 1$, denotes the fraction of power allocated to the source transmission. Let $\mathbf{E}[|h_{i,j}|^2] = \lambda_{i,j} = \frac{1}{d_{i,j}^{\alpha}}$. Hence, we can express β as $\beta = \sqrt{\frac{2\,SNR(1-\gamma)}{2\,SNR\,\gamma\,\lambda_{sr}+1}}$.

Substituting all the terms into Eq. (6.12), and computing expectation value with respect to the channel gain, we have the following lemma.

Proposition 1. *For the AF relay channel, the error exponent is given by*

$$E_r^{AF}(R) = \max_{0\le\rho,\gamma,d\le 1} \left\{ \frac{\rho}{1+\rho}\gamma\, SNR\Big(1+\frac{1}{d^\alpha} + \frac{1}{d^\alpha} C\exp(C)E_i(-C)\Big) - \rho R \right\}, \tag{6.13}$$

where d is the distance between source and relay; $C = \frac{(2SNR\gamma+d^\alpha)(1-d)^\alpha}{2SNR(1-\gamma)d^\alpha}$; $E_i(\cdot)$ is the exponential integral function [62, pp. 925].

Remark 1. *The optimal values (d^*, γ^*) to maximize the error exponent (6.13) depend on the path-loss exponent α. For $\alpha \ge 4$, $d^* \approx 1/2$, hence placing the relay node in the middle point of source S and destination D is optimal for large path-loss exponent. The optimal value (d^*, γ^*) monotonically decreases from 1 to 0.5 as path-loss exponent α increases.*

Proof.

$$\begin{aligned} E_r^{AF}(R) &= \max_{0\le\rho\le 1} \left\{ \frac{1}{2}\rho \ln \mathbf{E}\bigg(1+\frac{P_s}{(1+\rho)N_0} \Big(h_{s,r}^2 + h_{s,d}^2 - \frac{h_{s,r}^2}{|h_{r,d}\beta|^2+1}\Big)\bigg) - \rho R \right\} \\ &\overset{(a)}{=} \max_{0\le\rho\le 1} \left\{ \frac{\rho}{1+\rho} SNR\,\gamma \bigg(\lambda_{sd} + \lambda_{sr}\mathbf{E}\Big(\frac{\beta^2 h_{rd}^2}{1+\beta^2 h_{rd}^2}\Big)\bigg) - \rho R \right\} \\ &= \max_{0\le\rho,\gamma,d\le 1} \left\{ \frac{\rho}{1+\rho}\gamma\, SNR\Big(1+\frac{1}{d^\alpha} + \frac{1}{d^\alpha} C\exp(C)E_i(-C)\Big) - \rho R \right\}, \end{aligned}$$

where (a) using the low SNR approximation $\ln(1+x) = x$. □

Maximizing the AF error exponent (6.13) over d and γ can be easily decoupled from maximization with respect to ρ. Hence, we can numerically search the two-dimensional space of d and γ. Although we were not able to show analytically

Table IV. Optimal Relay Position d^* and Power Allocation γ^*

α	2	3	4	5	6
d^*	0.99	0.79	0.52	0.50	0.50
γ^*	0.98	0.87	0.59	0.54	0.53

that $E_r^{AF}(R)$ is concave in (d, γ), our simulation results indicate it. Also, note that the relay placement and power allocation are independent of the SNR values. We summarize optimal value (d^*, γ^*) for typical α value in Table I.

2. Decode-and-Forward (DF) Relay

For DF relay, the relay node decodes the source message it received from the source as $\hat{\mathbf{x}}_s$ for N narrowband carriers, re-encodes the information and sends it to the destination in the second step. In this work, we assume the simple repetition-coded scheme. The relay retransmits the signal as

$$x_r[n] = \hat{x}_s[n] \ ,$$

where n is the narrowband channel index; $\hat{x}_s$ is the decoded data at the relay node that was sent from the source. The error probability of DF relay transmission is:

$$\begin{aligned} P_e^{DF} &= \exp(-NE_r^{SR}) + \left(1 - \exp(-NE_r^{SR})\right) \cdot \exp(-NE_r^{MAC}) \\ &\approx \exp(-NE_r^{SR}) + \exp(-NE_r^{MAC}) \ , \end{aligned} \tag{6.14}$$

where E_r^{SR} is the error exponent of source-relay transmission; E_r^{MAC} denotes the destination decoding error exponent given repeated transmission from source and relay in two steps. Here we have assumed the number of narrowband carriers N or code block length is large enough that the error probability of source-relay is very

small. According to Theorem 2, we have

$$E_r^{SR} = \max_{0\le\rho\le 1}\left\{\frac{1}{2}\rho\ln\left(1+\frac{P_s\lambda_{sr}}{N_0(1+\rho)}\right)-\rho R\right\} \tag{6.15}$$

$$E_r^{MAC} = \max_{0\le\rho\le 1}\left\{\frac{1}{2}\rho\ln\left(1+\frac{P_s\lambda_{sd}+P_r\lambda_{rd}}{N_0(1+\rho)}\right)-\rho R\right\}. \tag{6.16}$$

Again, we halved degree of freedom and doubled the rate to $2R$ for the half-duplex communication.

Proposition 2. (Error Exponent of fixed DF relay) *:*
$E_r^{DF} = \min\{E_r^{SR}, E_r^{MAC}\}$.

We want to maximize the error exponent by power allocation and relay placement. We are using physical path-loss model of wireless propagation, and let $P_s = 2SNR\gamma$, $P_r = 2SNR(1-\gamma)$. Mathematically, we have following the optimizing problem,

$$\max_{0\le d,\gamma\le 1}\min\left\{2SNR\gamma\frac{1}{d^\alpha},\ (2SNR\gamma+2SNR(1-\gamma)\frac{1}{(1-d)^\alpha})\right\}. \tag{6.17}$$

Since the first term monotonically increases as γ and d increase from 0 to 1, but the second term is a monotonically decreasing function of γ and d, the minimum in (6.17) can be achieved when the first term equals to the second one. Hence, we can reduce the problem to

$$\max_{0\le d,\gamma\le 1}\gamma\frac{1}{d^\alpha} \qquad \text{subject to} \quad \gamma\frac{1}{d^\alpha} = \gamma+(1-\gamma)\frac{1}{(1-d)^\alpha}. \tag{6.18}$$

It is a one dimensional maximization, we can readily get the solution. We summarize the above analysis of the optimal (d^*,γ^*) in the following remark.

Remark 2. *The optimal value to maximize DF error exponent is* $(d^*,\gamma^*) = (\frac{1}{2}, \frac{1}{2-2^{-\alpha}})$. *Hence placing the relay node in middle point of source-relay line is optimal to boost*

the link reliability, and the power allocation γ is close to one half as path-loss exponent α increase.

It is well known that adaptive type DF, i.e., switch back to direction transmission in the event of relay decoding error, can achieve full diversity in the high SNR regime. However, in our wideband relay case, adaptive DF amounts to choose the better error exponent between direct transmission and DF transmission. Since DF relay has much higher error exponent value than direct transmission, adaptive type DF can not improve the performance anymore in our case.

E. Error Exponent for Block Markov Coding (BMC)

In this section, we focus on the full-duplex relay operation, i.e., when relay node can receive and transmit at the same time. Block Markov Coding (BMC) was first proposed by Cover and El Gammal [54] to derive the lower bound for the relay channel capacity. For convenience, we briefly restate the BMC process in the wideband multi-carrier background. The information bearing bits stream (message) at the source is parsed into blocks, each with N symbols; each block of N symbols can be transmitted in N narrowband carrier for one channel use. Let $w_i \in [1, 2^{NR}]$ be the message sent by the source during i^{th} block. The set of message $\mathcal{W} = \{1, 2, \cdots, 2^{NR}\}$ is randomly partitioned into bins $\mathcal{S} = \{S_1, S_2, \cdots, S_{2^{NR_0}}\}$ with $R_0 < R$. A random codebook $\mathcal{X} = \{\mathbf{x}_1(w|s), \mathbf{x}_2(s)\}$ is generated based on the joint probability distribution $p(x_1, x_2)$, where $w \in [1, 2^{NR}]$ and $s \in [1, 2^{NR_0}]$. After the relay successfully decodes the message from the source during the $(i-1)^{st}$ block, it transmits a codeword $\mathbf{x}_2(s_i)$ in the i^{th} block to help destination decode the previously received message. For detailed description, please refer to [54]. If we assume the entries of codewords $\mathbf{x}_1(w|s)$ and $\mathbf{x}_2(s)$ are independent, identical Gaussian distribution with zero mean and unit variance. The

simultaneously transmitted signal vectors by source and relay in i^{th} block are given, respectively, by

$$\begin{aligned}\mathbf{x}_s &= \sqrt{P_1}\mathbf{x}_1(w_i|s_i) + \theta\sqrt{(1-\gamma_2)P_2}\mathbf{x}_2(s_i) \\ \mathbf{x}_r &= \sqrt{\gamma_2 P_2}\mathbf{x}_2(s_i)\ ,\end{aligned} \tag{6.19}$$

where P_1 and P_2 are transmitted power of $\mathbf{x}_1(w|s)$ and $\mathbf{x}_2(s)$; $\gamma_2 \in (0,1)$ denotes the fraction of power P_2 allocated to relay. θ is the phase tuning factor to assist source-relay combining, which satisfies $|\theta|^2 = 1$. The received vector at the relay and destination can be expressed, respectively, as

$$\begin{aligned}\mathbf{y}_r &= \mathbf{h}_{sr} \cdot \mathbf{x}_s + \mathbf{z}_r \\ \mathbf{y}_d &= \mathbf{h}_{sd} \cdot \mathbf{x}_s + \mathbf{h}_{rd} \cdot \mathbf{x}_r + \mathbf{z}_d,\end{aligned} \tag{6.20}$$

where $\mathbf{h}_{ij}$ represents channel coefficient vector for i.i.d. narrowband carriers, for $i \in \{s, r\}$ and $j \in \{r, d\}$; $(\cdot)$ denotes componentwise multiplication. For BMC relay strategies, there are two transmissions for each message, one for source-relay link; the other are the source and relay multiple-access to the destination. By the Theorem 1, we have the following result for BMC relay.

Proposition 3. (Error Exponent of BMC relay) *:* $E_r^{BMC} = \min\{E_r^{B-SR}, E_r^{B-MAC}\}$, *where*

$$E_r^{B-SR} = \max_{0\le\rho\le 1}\left\{\rho\ln\left(1+\frac{P_1\lambda_{sr}}{N_0(1+\rho)}\right) - \rho R\right\} \tag{6.21}$$

$$\begin{aligned}E_r^{B-MAC} = \max_{0\le\rho,\gamma_2\le 1,\gamma_2}\Bigg\{&\rho\ln\Big(1 \\ &+ \frac{P_1\lambda_{sd} + (1-\gamma_2)P_2\lambda_{sd} + \gamma_2 P_2\lambda_{rd}}{(1+\rho)N_0} - \rho R\Bigg\}\ .\end{aligned} \tag{6.22}$$

In our system model, $\lambda_{rd} \ge \lambda_{sd}$, so the E_r^{B-MAC} is maximized when $\gamma_2 = 1$.

Hence the transmitted signal in (6.19) reduces to

$$\mathbf{x}_s = \sqrt{P_1}\mathbf{x}_1(w_i|s_i), \quad \mathbf{x}_r = \sqrt{P_2}\mathbf{x}_2(s_i) \ . \tag{6.23}$$

Let us assume $\frac{P_1}{N_0} = \gamma_1 \,\mathrm{SNR}$ and $\frac{P_2}{N_0} = (1-\gamma_1)\,\mathrm{SNR}$. The error exponent of BMC relay degenerates to a form similar to DF relay, but with half rate and double degree of freedom. The difference here comes from the full-duplex relay, rather than the orthogonal operation in the DF scheme. The results of optimal power allocation and relay placement for DF relay can also be applied here directly.

F. Numerical Results

In this section, we present numerical results to illustrate the advantage of the relayed transmission. We assume the relay node is placed in the line connecting the source and destination; the distance between source-destination is normalized to one. For each link, we consider the physical path-loss channel model with $\alpha = 4$. We normalized the SNR value and degree of freedom for a fair comparison among direct transmission, orthogonal relay and BMC relay, i.e. for the orthogonal relay, the degree of freedom is halved, rate and SNR value for each message transmission is doubled.

Fig. 26 compares the error exponent of different transmission strategies with optimal power allocation and relay placement. The SNR value is -3 dB, which accounts for our wideband low SNR assumption. We observed that the BMC relay has the highest reliability because we allow the full-duplex operation. Note that the DF scheme has an advantage over AF schemes in the error exponent sense. This observation is in contrast to the existing results in the literature that both AF and adaptive DF achieve full diversity in the high SNR regime. All of the above relay transmissions provide a significant reliability gain over the direct transmission.

To further illustrate the advantage of using relay in the wideband wireless transmission, Fig. 27 plots the minimum number of narrowband carriers to achieve a prescribed decoding error probability. The SNR value is defined as total power per channel use divided by N_0. The rate represents the sum rate of all the narrowband carriers. It is required to solve for N in the following equation $P_e = \exp(-NE_r(\frac{SNR}{N}, \frac{R}{N}))$ Fig. 2 shows that the relay strategies require far fewer carriers to achieve the prescribed decoding error probability for the same SNR value and transmission rate, compared with the direct transmission. Hence, it requires less bandwidth or provides higher spectral efficiency.

G. Conclusion

Random coding error exponents provide more information than the capacity. For any rate below the capacity, they quantify (lower bound) the exponential decay rate of the maximum-likelihood decoding error probability averaged over randomly selected codes. In this paper, we derived the random error exponent of the relay channel wideband relay strategies, analytical and numerical results show that using relay can indeed improve the system reliability significantly for rate below the capacity, which can save power or reduce bandwidth required in the practical wireless system. Furthermore, using physical path-loss wireless propagation model, we investigated the optimal relay placement and power allocation to further boost the system reliability.

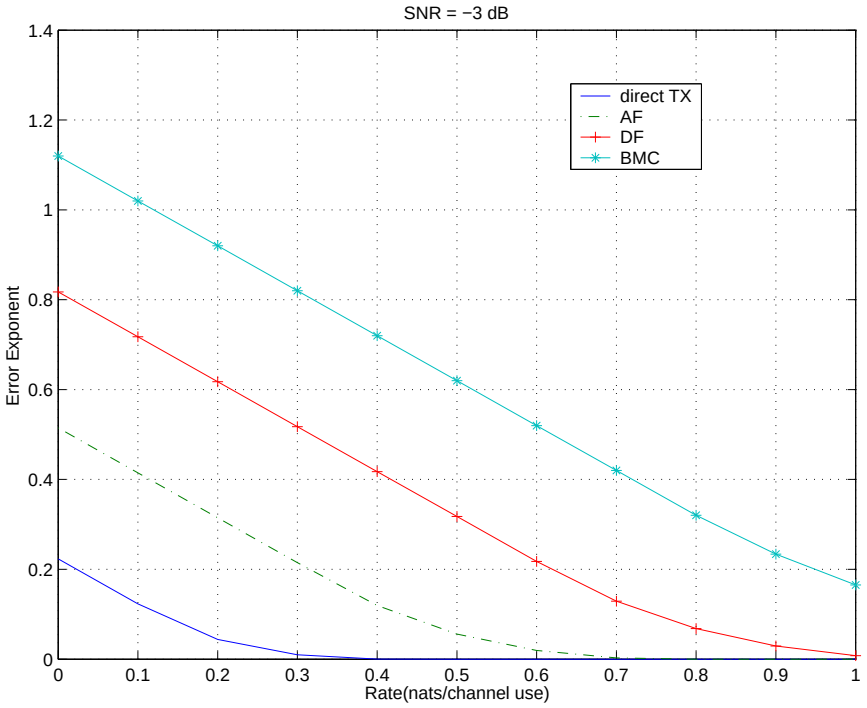

Fig. 26. Error exponent vs. rate with optimal power allocation and relay placement.

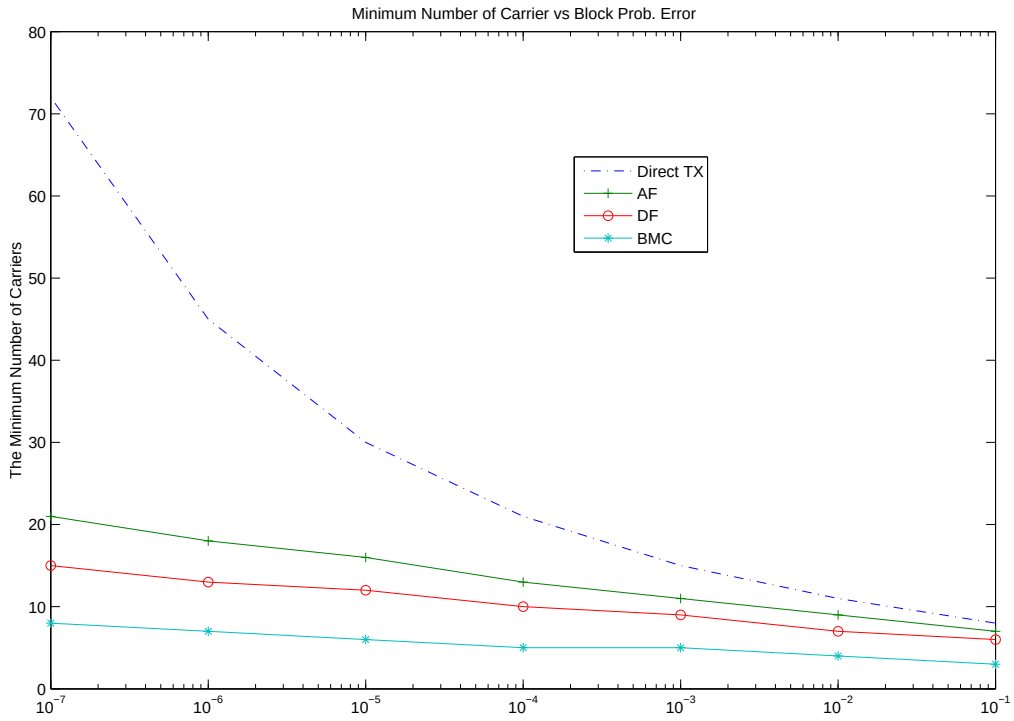

Fig. 27. Minimum no. of carriers needed to get prescribed error probability, SNR = 18 dB, R = 10 nats/channel use.

CHAPTER VII

QUALITY-OF-SERVICE FOR A BUFFERED TRANSMISSION OVER FADING CHANNEL

A. Introduction

Quality-of-Service (QoS) is a critical design objective for next-generation wireless communication systems. In general, data, voice and multimedia transmission over packet cellular networks, wireless LAN or sensor networks involves analogue observations transmitted to the end user over a wireless link. End-to-End distortion and transmission delay are two fundamental QoS metrics. Such QoS requirements pose a challenge for system design due to the unreliability and time varying nature of the wireless link.

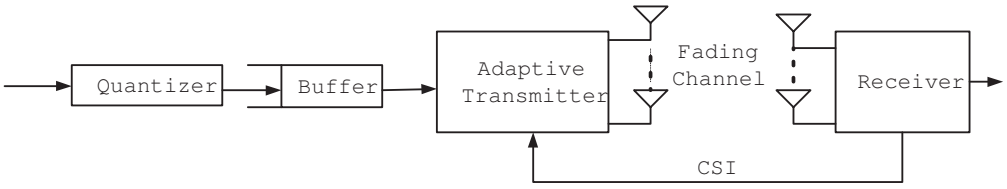

Fig. 28. System model

In this chapter, we consider transmission of an analogue source over a wireless time-varying fading channel as shown in Fig. 28. Our goal is to optimize the end-to-end distortion given a delay constraint. We first focus on the single antenna case (SISO) and derive the distortion and delay tradeoff for the wireless fading channel. We then extended our model to multiple input and multiple output (MIMO) block Rayleigh fading channel. We compute the SNR exponent [63] for the buffered transmission. To this end, we adopt a cross-layer approach shown in Figure 1. At this point, for simplicity we assume an independent and identically distributed (i.i.d.)

block fading channel model. Such a model is suitable for several practical communication scenarios, e.g., time hopping in TDMA, frequency hopping in FDMA and multicarrier systems. Extension to more practical time-correlated cases will be discussed later. Throughout this paper, we always assumed channel state information (CSI) was perfectly known at the receiver and that the transmitter only knew the instantaneous channel capacity via a feedback link (transmitter didn't need to know the exact channel realization).

We consider an i.i.d. complex memoryless Gaussian source $\sim \mathcal{CN}(0,1)$, which is quantized it and then fed into a buffer. Since the channel is time-varying, the transmitter adjust the transmission rate to the current channel status. The relevant performance criteria are the end-to-end quadratic distortion and the buffer delay. We aim to find the relationship between the distortion and delay for some average transmission power. The Gaussian source is a good approximation of more general source distribution in high resolution regime [64, 65]. We assume that each group of K source samples was transmitted over N channel uses on average. We define the corresponding *bandwidth ratio* as

$$\eta = \frac{N}{K}, \tag{7.1}$$

where K was large enough to consider the source as ergodic and N was large enough to design codes that could achieve instantaneous channel capacity. Our tools here were large deviation theory and information theory.

Recently, some researchers have considered such end-to-end quadratic distortion as their performance criteria. In [66], Holliday and Goldsmith rst investigated the end-to-end distortion for the MIMO block fading channel based on the source-channel separation theorem and Zheng & Tse's diversity-multiplexing trade-off. They also incorporated delay consideration into their model using the ARQ argument which

is different from our approach. In [64], Laneman et al., considered the problem of minimum average distortion transmission over parallelled channels. They introduced the distortion SNR exponent as a figure of merit for high SNR value and compared multiple description source coding diversity and channel coding diversity. Caire and Narayanan [63] pointed out that the the separation theorem does not hold for a delay constrained and the unknown channel at the transmitter end. They investigated the SNR exponent of the distortion function in high SNR regime. For this problem, an upper bound and lower-bound for the distortion SNR exponent were derived. [65] Gunduz and Erkip extended their results using a layered broadcast transmission scheme. For some bandwidth ratios, the optimum SNR exponent was achieved.

For the combination of queuing and information theories, in [67], Wu and Negi, first proposed the concept of effective capacity, an extension of Shannon's capacity, by incorporating the buffer delay into it. The effective capacity is the dual of Chang's effective bandwidth [68] in the network literature. Negi and Goel [69] united the effective capacity with the error exponent for more practical considerations. A QOS-aware rate and power control algorithm for wireless fading channel was proposed by Tang and Zhang [70].

For buffered transmission, Berry and Gallager investigated the power and delay tradeoff for communication over fading channel [71]. In [72], Tse analyzed the distortion for a xed line networks, but with an adaptive quantizer. Part of results in this chapter have been presented in [73].

The rest of this Chapter is organized as follows: in Section B, we have stated the problem and showed inserting a buffer can save significant power. We have introduced the system model and some preliminaries of the effective capacity in Section C. Section D develops our main results - distortion-delay function and an upper bound for SISO channel, some asymptotic analysis is also provided. In Section E, we have

extended the distortion analysis to the MIMO channel, and derived the SNR exponent for buffered transmission. Distortion-delay for a large antenna MIMO channel was also derived by utilizing the mutual information Gaussian approximation. Section F discusses some extension to a more general case. Finally, Section G concludes the paper.

B. Problem Statement

For buffered transmission over the fading channel, there are two extreme cases: 1) There is no buffer — no delay, 2) we have an infinite buffer size, i.e., we allow an infinite transmission delay. For the first case, we adaptively quantize the Gaussian source according to the CSI. Assuming perfect transmission, we can approximate the average achievable quadratic distortion by:

$$\mathcal{D}_0(\rho) = \mathrm{E}[\exp(-\eta \ln(1 + |h|^2 \frac{P}{N_0 W})] \ , \tag{7.2}$$

where P denotes the transmission power, W and N_0 resent the bandwidth and noise variance; h is the channel gain, a random variable with unit variance follow a certain statistical distribution. Here, we have used the information theoretical results: Gaussian distortion-rate function can be express as $\mathcal{D}(R_s) = \exp(-\eta R_c)$ and $C(\rho) = \log(1 + |h|^2\rho)$ is the instantaneous channel capacity-cost function [74]. For an infinite delay case, the average transmission rate can achieve the ergodic capacity of a fading channel and the quantizer can simply adopt a constant output rate. The average distortion is given by:

$$\mathcal{D}_\infty(\rho) = \exp(-\eta \, \mathrm{E}[\ln(1 + |h|^2 \frac{P}{N_0 W})]) \ . \tag{7.3}$$

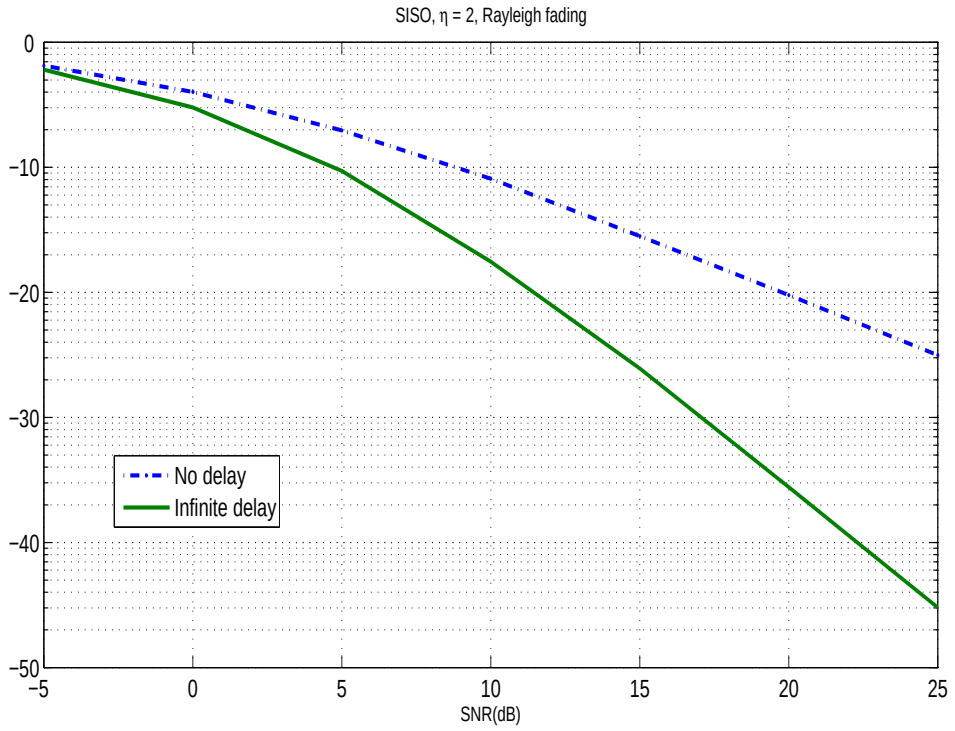

Fig. 29. Distortion of Gaussian source transmitted over i.i.d. Rayleigh fading channel.

The function $\exp(-(\cdot))$ is a covex function. Due to Jensen's inequality, the distortion $\mathcal{D}_0$ is low bounded by $\mathcal{D}_\infty$, i.e., $\mathcal{D}_0 \geq \mathcal{D}_\infty$. The two distortion functions are plotted in Fig. 29 for a Rayleigh fading channel. Notice that there is a gap between no-delay and infinite delay curves. We can call this transmission power gap "Jensen's gain." Note, we assume $\eta = 2$ and a complex Gaussian source, this is equivalent to a real source with bandwidth ratio of one. So introducing a buffer at the transmitter to match the source rate with the instantaneous quality of the channel can save a considerable amount of transmission power to meet some distortion requirement. Also, we have simplified the quantization step (constant rate). A natural question is therefore: if we allow only a finite delay or buffer, how much gain can we achieve? How fast does

the distortion curve converge to the infinite-delay lower bound as the delay increases? One of the the main result of this paper is to develop a clear characterization of the tradeoff between end-to-end quadratic distortion and delay to provide insight into the impact of the buffer delay on the achieved distortion function of the memoryless analogue source transmitted over a wireless fading channel.

To answer the question raised earlier, we combined ideas from the fields of queuing theory and communication/information theory to analyze the above problem. The tool we used is the concept of *effective capacity* [67], which is the dual of *effective bandwidth* in networking literature. The effective capacity synthesizes the channel statistics and QoS metric (delay and buffer overflow) into a single function using large deviation theory. It is a powerful and united approach for studying the statistical QoS performance of wireless transmission where the service process is time-varying. For i.i.d. SISO block fading channels we derived a closed-form expression for the distortion-delay curve which is hard to analyze due to some mathematically intractable special functions. We then gave out a tight upper bound for this distortion-delay function to theoretically and asymptotically analyze convergence behavior.

In Fig. 29, we find the power gain is marginal for low SNR. As the SNR value increases, the gain becomes significant because the $\exp(\cdot)$ and $\log(\cdot)$ functions are approximately linear in the low SNR regime. Hence, "Jensen's gain" is negligible at low SNR. We can view the slope of the distortionSNR curve as a similarity of the diversity order for the bit error rate in the wireless communication. Therefore, we will investigate the distortion SNR exponent for a buffered transmission. Introducing a buffer can provide some kind of time diversity. For the MIMO channel, besides time diversity, there is also space diversity. We will examine the interplay between these two diversities and the impact of buffer on the SNR exponent.

C. System Model

The system model is illustrated in Figure 1. There is an i.i.d. complex Gaussian source $\sim \mathcal{CN}(0,1)$ with total bandwidth B_w. We quantize the source samples using vector quantizer or trellis coded quantizer (TCQ). The quantization operated every K samples a time and fed into a buffer with size B bits. Let the K samples have time duration T_f, so each frame has $T_f \times B_w \times R_s = K \cdot R_s$ bits, where R_s bits is number of bits into which each Gaussian sample is quantized. K is large enough to ensure ergodic of the source.

We assume a MIMO i.i.d. block fading channel with M_t transmit and M_r receive antennas. The SISO, MISO and SIMO are special cases of this general model. The channel model can be expressed as:

$$\mathbf{y}_i = \sqrt{\frac{\rho}{M_t}}\mathbf{H}\mathbf{x}_i + \mathbf{w}_i, \qquad i = 1, \cdots, N \tag{7.4}$$

Where $\mathbf{H}$ is the channel matrix containing i.i.d. elements $h_{i,j} \sim \mathcal{CN}(0,1)$ (Rayleigh independent fading). $\mathbf{x}_i$ is the transmitted signal at time i, the codeword $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_N] \in \mathbb{C}^{M_t \times N}$ is normalized so that is satisfies $\mathrm{tr}(\mathbb{E}[\mathbf{X}^H\mathbf{X}]) \leq M_t N$. ρ denotes the signal-to-noise ratio (SNR), defined as the ratio of the average received signal energy per receiving antenn to the noise per-component variance. $\mathbf{Z} = [\mathbf{z}_1, \cdots, \mathbf{z}_N] \in \mathbb{C}^{M_r \times N}$ is the complex additive Gaussian noise with i.i.d. entries $\mathcal{CN}(0,1)$. We define $M_* = min(M_t, M_r)$ and $M^* = max(M_t, M_r)$.

1. Effective Capacity

The key idea of effective capacity is that, for a dynamic queuing system with stationary ergodic arrival and service process, the queue length $Q(t)$ converges in distribution to a random variable $Q(\infty)$. The probability of queue length exceeding a certain

threshold B decays exponentially as threshold B increases [67, 75]. Mathematically,

$$\lim_{B\to\infty} \frac{-1}{B} \ln Pr\{Q(\infty) > B\} = \theta \ , \tag{7.5}$$

where θ is the QoS parameter decided by the delay requirement of the queue system. A large value of θ leads to a stringent delay requirement, i.e., small delay. In particular, as θ goes to ∞, the system can not tolerate any delay. On the other end, when θ goes to 0, the system can tolerate an arbitrarily delay.

Let the sequence $\{R[i], i = 1, 2, \ldots\}$ denote the discrete-time instantaneous channel capacity, which is a stationary and ergodic stochastic process. Define

$$S[t] \triangleq \sum_{i=1}^{t} R[i] \tag{7.6}$$

as the accumulate service provided by the channel. Assume the Gärtner-Ellis limit of $S[t]$:

$$\Lambda_C(\theta) \triangleq \lim_{t\to\infty} \frac{1}{t} \ln \mathrm{E}\Big\{e^{\theta S[t]}\Big\}, \quad \forall\, \theta > 0 \tag{7.7}$$

exits and is a convex function differentiable for all real θ. Then, the effective capacity with delay constraint decided by θ is defined as

$$E_C(\theta) \triangleq -\frac{\Lambda_C(-\theta)}{\theta} = -\lim_{t\to\infty} \frac{1}{\theta t} \ln \mathrm{E}\Big\{e^{-\theta S[t]}\Big\} \ . \tag{7.8}$$

In particular, for i.i.d. cases, the effective capacity simply reduces to the ratio of log-moment generating function of the instantaneous channel capacity to the exponent θ

$$E_C(\theta) = -\frac{1}{\theta} \ln \mathrm{E}\Big\{e^{-\theta R[t]}\Big\} \ . \tag{7.9}$$

The effective capacity falls into the large deviation framework, which is asymptotically valid for a large queue size.

D. Distortion-Delay Function

We derived the closed-form expression for the end-to-end quadratic distortion given the delay constraint in this section. The starting point was vector quantization and delay bound violation probability using effective capacity. For a Gaussian source vector $\mathbf{u}$ with K samples that had support on $\mathbb{C}^K$, a KR_s-nats quantizer was applied to $\mathbf{u}$ via a mapping $\mathbf{u} \rightarrow \tilde{\mathbf{u}}$. The cardinality of discrete set $\tilde{\mathbf{u}}$ is e^{KR_s}. The average quadratic distortion was defined by

$$\mathcal{D}_Q(R_s) \triangleq \frac{1}{K}\mathrm{E}[|\mathbf{u} - \tilde{\mathbf{u}}|^2] \ , \tag{7.10}$$

where the expectation is with respect to $\mathbf{u}$. According to the distortion-rate theory, the distortion function $D_Q(R_s) = \exp(-R_s)$ is achievable for a complex Guassian source. When the quantized bits are transmitted over a statistical channel, let P_e denote the error probability of this channel. It has been shown in [76] that the achievable end-to-end distortion for such tandem scheme is upper bounded by

$$\mathcal{D}_{e-e}(R_s) \leq \mathcal{D}_Q(R_s) + O(1)P_e \ . \tag{7.11}$$

For our problem, if we assume using Gaussian code to achieve the instantaneous capacity, the delay bound violation (buffer overflow) probability will dominate the decoding error probability. From the effective capacity theory, we have the following approximation for P_e:

$$P_e \triangleq P_r\{Q(\infty) \geq B\} \approx \kappa e^{-\theta B} \ , \tag{7.12}$$

where θ is the QoS parameter, B is the buffer size; κ is a constant that denotes the probability that the buffer is non-empty. κ is large compared with P_e. Given the delay constraint at τ seconds, using Little's theorem, we have following result: $B = Rs \times B_w \times \tau$. B_w is the source bandwidth. Substitute (7.12), B and $D_Q(R_s)$

into (7.11), we may write the bound on the end-to-end distortion as

$$\mathcal{D}_{e-e}(R_s) \leq \exp(-R_s) + O(1)\kappa \exp(-\theta B_w R_s \tau) \ . \tag{7.13}$$

In order to get analytical results, we consider the asymptotically large delay and high SNR regime, i.e., small distortion. We can optimize the end-to-end distortion by choosing the two exponents equal to each other (exponential order tight) resulting in $\theta = \frac{1}{B_w \tau}$.

If we assume the transmitter doesn't know the channel realization, but does know the value of instantaneous capacity via the feedback link, the instantaneous capacity can be achieved by the Gaussian codebook. We have the following theorem.

Theorem 4. *Given a delay $\tau = \frac{1}{B_w \theta}$ and bandwidth raio η, the distortion upper bound function of the i.i.d MIMO block fading channel can be expressed as:*

$$\mathcal{D}(\theta) \leq \left[\mathbf{B}^{-1} \det[\mathbf{G}(\theta)]\right]^{\frac{1}{K\theta}} \ . \tag{7.14}$$

where $\mathbf{B} = \prod_{i=1}^{M_*} \Gamma(d+i)$, and $d = M^* - M_*$. And $\mathbf{G}$ is $M_* \times M_*$ Hankel matrix whose (i,j)th entry is defined to be

$$g_{i,j} = \int_0^\infty \left(1 + \frac{\rho}{M_t}\lambda\right)^{-\theta K \eta} \lambda^{i+j+d} e^{-\lambda} d\lambda, \quad i,j = 0, \cdots, M_* - 1 \ . \tag{7.15}$$

Γ is the complete Gamma function.

Proof. The Mutual information for the each MIMO block transmission can be expressed as:

$$R_s(\mathbf{H}) = K\eta \cdot \ln \det \left(\mathbf{I} + \frac{\rho}{M_t}\mathbf{H}\mathbf{H}^H\right) \tag{7.16}$$

plug into equation (7.9) and (7.13), we have

$$\mathcal{D}(\theta) \leq \left\{ \mathrm{E}\left[\det\left(\mathbf{I} + \frac{\rho}{M_t}\mathbf{H}\mathbf{H}^H \right) \right]^{-\theta K \eta} \right\}^{\frac{1}{\theta K}}$$
$$= \left\{ \int_0^\infty \prod \left(1 + \frac{\rho}{M_t}\lambda_i \right)^{-\theta K \eta} f(\boldsymbol{\lambda}) d\boldsymbol{\lambda} \right\}^{\frac{1}{\theta K}} . \tag{7.17}$$

Where $0 \leq \lambda_1 \leq \cdots \leq \lambda_{M_*}$ denote the ordered eigenvalues of $\mathbf{H}\mathbf{H}^H$. The joint distribution of the λ_i's follows the Wishart pdf given by

$$f(\boldsymbol{\lambda}) = K_{M_t,M_r}^{-1} \prod_{i=1}^{M_*} \lambda_i^{M^* - M_*} \prod_{i<j} (\lambda_i - \lambda_j)^2 \exp\left(- \sum_i \lambda_i \right) , \tag{7.18}$$

where K_{M_t,M_r} is a normalization constant. Follow the results of [77], we can get the distortion function as (7.14). □

Remarks

- If we assume the quantization process is independent of the channel status, we can show the the constant quantization rate is the optimum one. First, for a buffered system with independent arrival and departure processes, the constant arrival process is optimal with respect to the buffer overflow probability for all the arrival processes that have the same average rate [68]. Second, given a buffer overflow probability, the constant rate quantization will minimize the distortion according to Jensen's inequality. Therefore, the constant rate quantization is optimal if the quantization process is independent of channel mutual information. Another advantage of constant rate quantization is that it reduces the quantizer design complexity.

- When the quantizer rate selection is according to the buffer state status, we cannot prove that the constant rate quantization is optimal. Hence, the distortion of(7.14) is an upper bound. One extreme case is that the quantizer is

chosen to make sure no buffer overflows, i. e. , the quantization rate selection is to match the channel mutual information profile. This scheme will degenerate to no buffer (delay) case. Therefore, it is seriously suboptimal. The optimal quantizer rate should balance the "Jensen's gain" and the reduced distortion by decreasing the buffer overflow probability via the quantization rate matching the buffer status.

The introduced buffer delay in (7.17) at first shrinks the integrand near to 1 as $\theta \to 0$, and then restores it after taking the expectation. From Fig. 30, we can observe that after the contraction function of $(\cdot))^{\theta}$ as θ goes to zero, the integrand function becomes more linear. This observation can explain why we have a large gain after introducing a buffer delay mathematically, and provides some intuition of the distortiondelay function. Moreover, Fig. 30 shows that the larger bandwidth ratio η is, the more effective the shrink operation (larger gain). Therefore, introducing a buffer delay provides a larger gain for the high bandwidth ratio scenario, or high resolution quantization. We theoretically confirmed the result later by deriving the SNR exponent.

The result of Theorem 1. is very complicated; very little insight can be gained from the expression itself. In the ensuing part of this paper, we first investigate the distortion-delay of SISO, MISO / SIMO case, in which a simpler form can be arrived. Then, for a more general MIMO channel, we considered the high SNR regime and computed the distortion SNR exponent. Guassian approximation of MIMO mutual information was also used to derive an approximation for large antenna system.

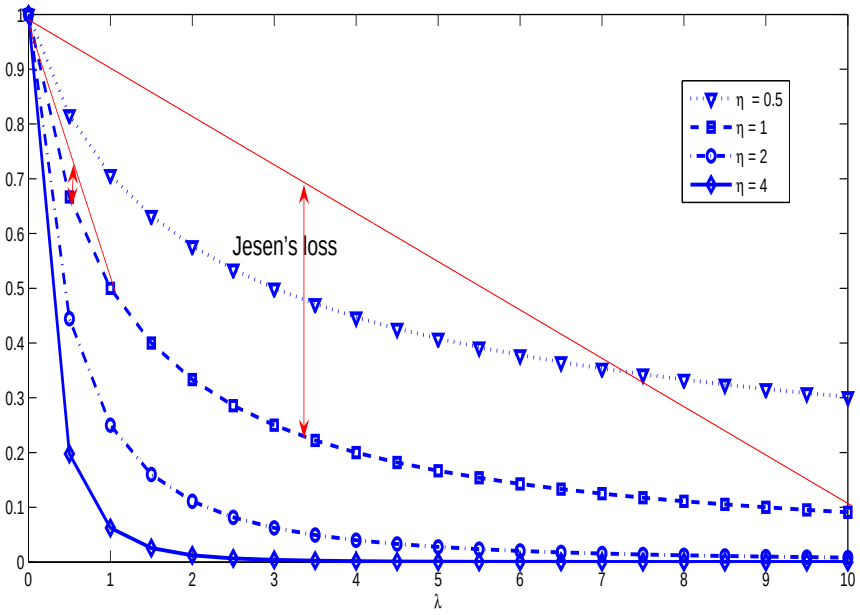

Fig. 30. Illustration of buffer delay effect on the distortion

1. Single Antenna System (SISO)

For simplicity, we introduced the normalized delay as $\tau_n = \tau/T_f = \frac{1}{\theta B_w T_f} = \frac{1}{K\theta}$. For the SISO Rayleigh fading channel, the channel matrix degenerated to a scalar channel. We have following Corollary.

Corollary 1. *For SISO system, the distortion-delay upper bound is*

$$\mathcal{D}(\lambda\eta) \leq \left[\rho^{-\lambda\eta}\exp\left(\frac{1}{\rho}\right)\gamma\Big(1-\lambda\eta,\frac{1}{\rho}\Big)\right]^{\frac{1}{\lambda}}, \tag{7.19}$$

where $\lambda = \frac{1}{\tau_n}$ *and* $\gamma(\cdot,\cdot)$ *is the incomplete Gamma function.*

Proof. For SISO channel, the (7.14) is reduced to the scaler case,

$$\mathcal{D}(\lambda) \leq \left[\int_0^\infty \Big(1+\rho x\Big)^{-\lambda\eta} e^{-x} dx\right]^{\frac{l}{\lambda}}, \tag{7.20}$$

by the formula of [62], we can complete the proof. □

The closed-form expressions of (7.19) is very difficult to analyze due to the special functions. In order to analyze distortion as the delay constraint increases, it is desirable to reduce the function into some simple form that is easy to handle. This objective motivates us to derive an asymptotically tight upper bound for the distortion-delay function in next section.

a. Asymptotic Analysis

We started by characterizing the behavior of the tail of the distortion-delay curve $\mathcal{D}(\tau_n)$, hence we are interested in the asymptotically large delay regime. We only considered the Rayleigh fading SISO case. In this part, we assumed $\eta = 1$ for simplicity, generalizing to other bandwidth ratio is straightforward. We tried to show that $\mathcal{D}(\tau_n) \to \mathcal{D}(\infty)$ as $\tau_n \to \infty$. In addition, we proveed that the limit is approached

as $e^{\frac{C}{\tau_n}}$ by finding the upper bound on the distortion-delay function and then show the bound is asymptotically tight. The ergodic capacity of m_{th}-order diversity Raleigh fading channel with a constant transmission power can be expressed as [78]:

$$C_{erg} = \frac{\gamma(m, -m/\rho)}{\Gamma(m)} E_1(m/\rho) + \sum_{k=1}^{m-1} \frac{1}{k} \frac{\gamma(k, m/\rho)\gamma(m-k, -m/\rho)}{\Gamma(k)\Gamma(m-k)} \ , \tag{7.21}$$

where $\gamma(\cdot,\cdot)$ and $\Gamma(\cdot)$ denote incomplete and complete Gamma functions; $E_1(\cdot)$ presents the exponential integration function. Hence for $m = 1$, the lower bound of distortion/delay function can be written as:

$$\mathcal{D}(\infty) = \exp\Big(- e^{\frac{1}{\rho}} E_1(1/\rho) \Big) \ . \tag{7.22}$$

Next, We tried to derive the asymptotic upper bound on $\mathcal{D}(\tau_n)$ of (7.19) to achieve the limit $\mathcal{D}(\infty)$. We mean asymptotically in the sense of $\tau_n \to \infty$ or $\lambda \to 0$.

Theorem 5. *An asymptotic upper bound for $\mathcal{D}(D_n)$ can be expressed as:*

$$\mathcal{D}_{upper}(\lambda) = \left[\frac{1}{\lambda - 1}\big(e^{\frac{1}{\rho}} - 1\big) + \frac{1}{1 - \xi\lambda + \phi\lambda^2} \rho^{-\lambda} e^{\frac{1}{\rho}} \right]^{\frac{1}{\lambda}} \ , \tag{7.23}$$

where $\xi = 0.577215$ and $\phi = \frac{1}{12}(6\xi^2 - \pi^2)$. As $\lambda \to 0$ this upper bound is asymptotically tight and approaches $\mathcal{D}(\infty)$ as $\mathcal{D}(\infty) \cdot e^{C\lambda}$, where C is some constant.

Proof. See Appendix B. □

b. Example

We present some numerical results to verify our findings. Suppose we have a real Gaussian source $N(0,1)$ with bandwidth $100kHz$, bandwidth ratio $\eta = 1$[1]. We assume an i.i.d. block Rayleigh fading channel model. Let the duration of each time

[1]A real Gaussian source is equivalent to a complex one with doubled bandwidth ratio

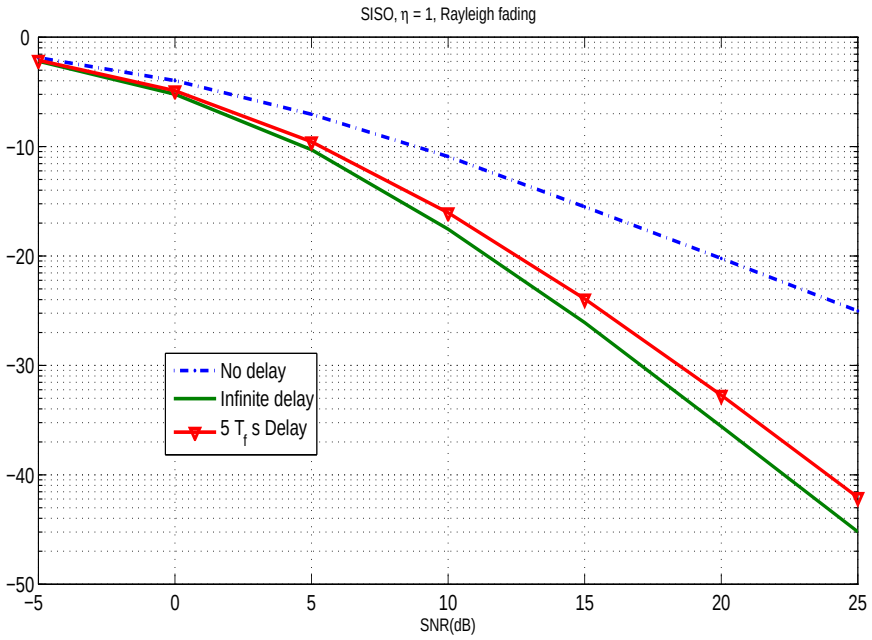

Fig. 31. Distortion of real Gaussian source transmitted over i.i.d.Rayleigh fading channel.

frame be $2ms$ such that each data frame consists of 200 source samples. Fig. 31 shows that a normalized delay of 5Tf can achieve most of the gains, especially for high transmission power. The gap between this curve and the infinite delay case is less than 1dB for a typical SNR value. In Fig. 32, we plotted the end-to-end quadratic distortion vs. SNR and delay. It clearly characterized the distortion and delay tradeoff for the Gaussian source transmitted over the wireless fading channel. Note that the higher the SNR value, the faster the distortion converges to the infinite delay lower bound. For an SNR value of 25 dB, less than 2Tf delay can achieve most of the Jensen's gain.

Fig. 33 shows the upper bound for the distortion/delay $\mathcal{D}(D_n)$ curve at SNR = 15dB. The ergodic Shannon capacity in this case is 3.0015 nats/symbol and the

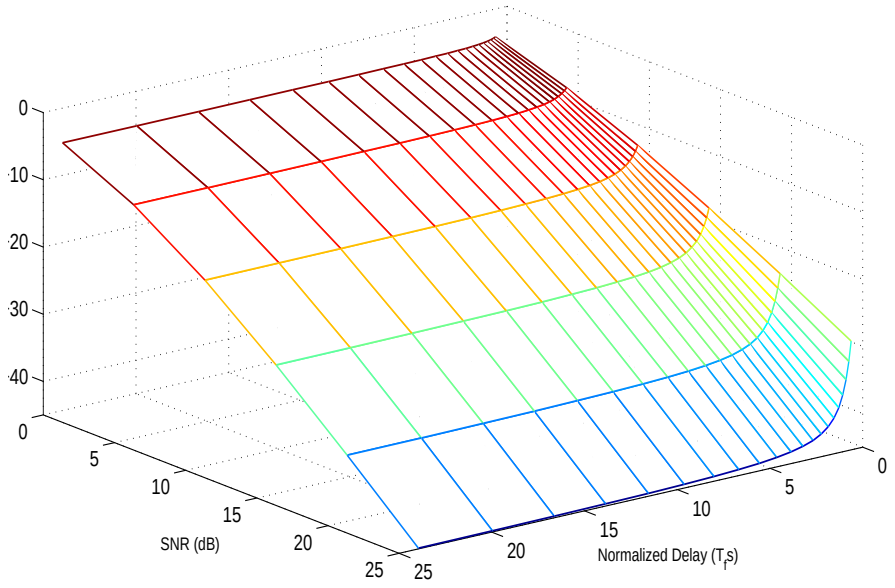

Fig. 32. Distortion vs. delay and SNR

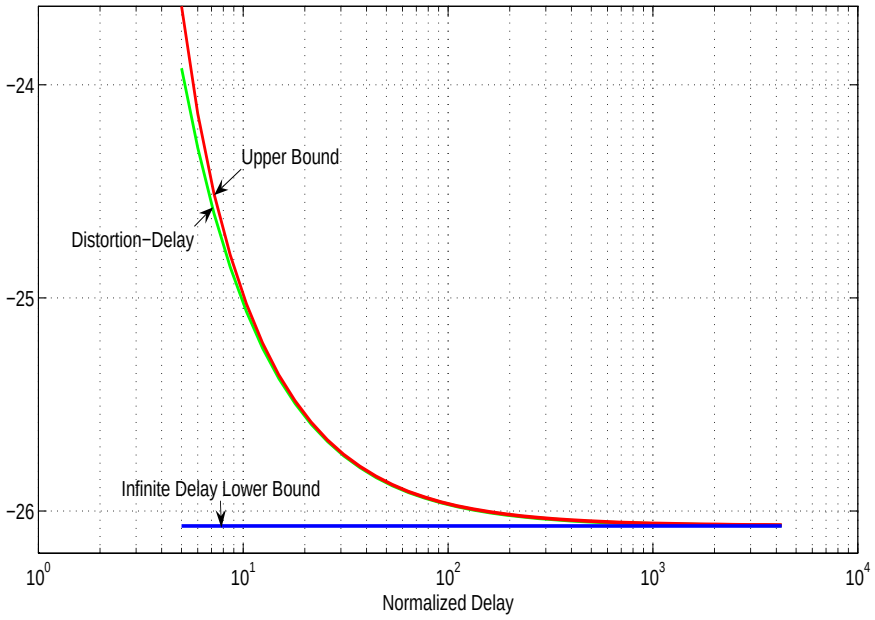

Fig. 33. Upper bound of distortion/delay function (SNR=15dB)

distortion $\mathcal{D}(\infty)$ is 0.0025. The rate of distortion/delay curve and the upper bound converge to the infinite delay lower bound is clearly illustrated in Figure 5. It shows the upper bound is asymptotically tight and converges. From this upper bound and the distortion/delay function, we observed that introducing some finite delay can help achieve the $\mathcal{D}(\infty)$ lower bound very quickly. In some practical applications, e.g., video transmission over a wireless fading channel, which can tolerate a certain amount of delay, our results suggested that inserting a buffer between quantizer and transmitter will significantly enhance the image quality. Intuitively, a transmission delay can be thought of as some delay diversity corresponding to space diversity in the MIMO channel. Hence there is also some diversity-rate tradeoff for our problem which can lead to results similar to those in [63].

2. SIMO/MISO Antennas System

For a SIMO channel of m receiver antenna. We can consider such channel as a m_{th}-order combining diversity Rayleigh fading channel. Again, for simplicity, we here assume $\eta = 1$. The channel gain after combining is Chi-square distributed with $2m$ degrees of freedom, and the probability density function (pdf) is given by:

$$f(x) = \frac{1}{(m-1)!} x^{m-1} e^{-x}, \quad x > 0 \ . \tag{7.24}$$

Corollary 2. *For the SIMO Rayleigh fading channel with m receive antennas. The distortion-delay upper bound has a closed-form expression:*

$$\mathcal{D}_m(\tau_n) \leq \left[\frac{\Gamma(\lambda - m)}{\Gamma(\lambda)} \rho^{-m} {}_1F_1\Big(m; m-\lambda+1; \frac{1}{\rho}\Big) + \frac{\Gamma(m-\lambda)}{\Gamma(m)} \rho^{-\lambda} {}_1F_1\Big(\lambda; \lambda-m+1; \frac{1}{\rho}\Big) \right]^{\tau_n} , \tag{7.25}$$

where $\lambda = 1/\tau_n$.

Proof. We start from Eqn. (7.14), with SIMO case

$$\begin{aligned} \mathcal{D}(\theta) &= \Big(\int_0^\infty (1+\rho x)^{-\lambda} f(x) d\rho \Big)^{\tau_n} \\ &= \left(\frac{1}{(m-1)!} \cdot \int_0^\infty (1+\rho x)^{-\lambda} x^{m-1} e^{-x} dx \right)^{\tau_n} , \end{aligned} \tag{7.26}$$

where we have used the expression of $f(x)$ in (7.24). We know that [62, Ch. 3.383.5]:

$$\int_0^\infty e^{-px} x^{q-1} (1+ax)^{-v} dx = a^{-q} \Gamma(q) \Psi\big(q, q+1-v; \frac{p}{a}\big) \ , \tag{7.27}$$

where $\Psi(\cdot,\cdot;\cdot)$ denotes the degenerate Hypergeometric function. Reducing to the more commonly used confluent hypergeometric function, we have following relation:

$$\Psi(x,y;z) = \frac{\Gamma(1-y)}{\Gamma(x-y+z)} {}_1F_1(x;y;z) + \frac{\Gamma(y-1)}{\Gamma(x)} z_1^{1-y} F_1(x-y+1; 2-y; z) \ . \tag{7.28}$$

Let $p = 1$, $q = m$, $v = \lambda$ and $a = \rho$. Plugging (7.28) into (7.27), we can prove Lemma

1. □

For MISO case[2], it is similar to the SIMO case but the power is divided by m. Even for the SIMO/MISO case, the distortion-delay upper bound function is very complicated. We can get only some numerical results. Therefore, for a more general MIMO channel, we resorted to the SNR exponent in the high SNR regime to demonstrate the buffer gain.

E. Distortion Exponent of MIMO Block Fading Channel

For MIMO block fading channel with a buffered transmission, Eqn. (7.14) is very hard to analyze and provides less insight. We can only use the numerical method to compute the function since "Jensen's gai" is negligible in the low SNR regime and becomes significant at the high SNR. Therefore, we are more interested in the high SNR behavior of the expected distortion. We defined the figure of merit of *distortion exponent* [63] with bandwidth ratio η:

$$\alpha(\eta) = -\lim_{\rho \to \inf} \frac{\log \mathcal{D}(\rho, \eta)}{\log \rho} \ . \tag{7.29}$$

A distortion exponent of α means that the expected distortion decays as $\rho^{-\alpha}$ with increasing SNR value ρ when the SNR is high. We want to characterize the buffer delay and bandwidth ratio's impact on the SNR exponent.

Theorem 6. *[63] (No Buffer) For transmission of memoryless, complex Gaussian source over a MIMO block fading channel, the distortion exponent with perfect known channel is given by*

$$\alpha(\eta) = \sum_{i=1}^{M_*} \min\Big(\eta, 2i - 1 + |M_t - M_r|\Big). \tag{7.30}$$

[2]We assume transmitter has CSI for MISO case for beamforming transmission

The proof of Theorem 6, using the technique of [2]. Intuitively, when the bandwidth ratio is low, the distortion is limited by the η and the degree of freedom of the MIMO channel – the total degree of freedom utilized to transmit the information. On the other hand, when the bandwidth ratio is high, we needed more diversity to provide transmission reliability. Hence, for the high bandwidth ratio, the system is diversity limited and the SNR exponent is determined by the second term.

Theorem 7. *(with buffer delay) For transmission of memoryless, complex Gaussian source over a MIMO block fading channel, If the quantized bits are stored in a buffer before transmitting over the fading channel. Assume the transmitter know exactly the instantaneous channel capacity, the distortion SNR exponent is given by*

$$\alpha(\eta) = \tau_n \min\left\{\frac{\eta}{\tau_n}, 2i - 1 + |M_r - M_t|\right\} . \tag{7.31}$$

Proof. Proof can be found in Appendix II. □

Remarks

- We found the SNR exponent of Theorem 4 is similar as the one of joint encoding and decoding of L MIMO fading blocks. However, the joint encoding increased transmitter and receiver complexity. Introducing a simple buffer delay can produce the same SNR exponent by utilizing the time diversity.

- For the SIMO/MIMO case, the SNR exponent reduces to $\min\{\eta, \tau_n M\}$, where M is the receiver / transmitter antenna number. We can consider $\eta = \tau_n M$ as a corner point. Below this point, the system's degree of freedom is limited, hence introducing more antenna will not improve the SNR exponent. Beyond this point, the system is diversity limited. By increasing the antenna number to provide more combining branches to increase diversity, the SNR exponent is also increased.

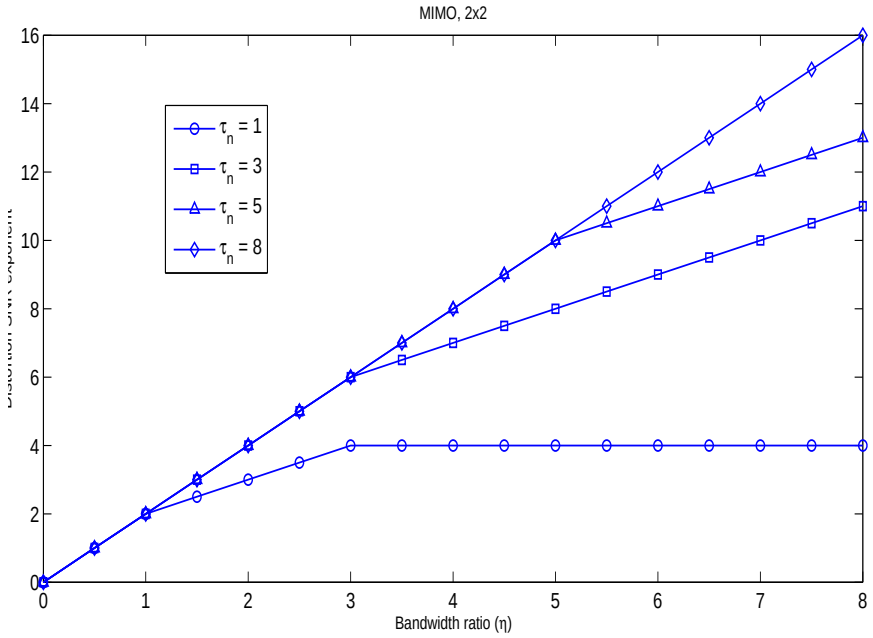

Fig. 34. Distortion exponent v.s. bandwidth ratio for block fading 2x2 MIMO channel.

In Fig. 34, we fixed the MIMO channel as 2×2, and plotted the SNR exponent v.s. the bandwidth ratio curves for different delays. As the delay increases, we have more time diversity to combat fading, hence the corner point of the exponent-bandwidth ratio curve also increases. For $\tau_n = 1$, the maximum SNR exponent can be achieved for $\eta = 3$. It is useless to increase channel bandwidth ratio beyond 3 in the high SNR. In Fig. 35, We fixed the normalized delay as $\tau_n = 5$ and showed different SNR exponent-bandwidth ratio curves for different antenna settings. For the SISO channel, the SNR exponent will not increase any more as the bandwidth ratio increase beyond 5.

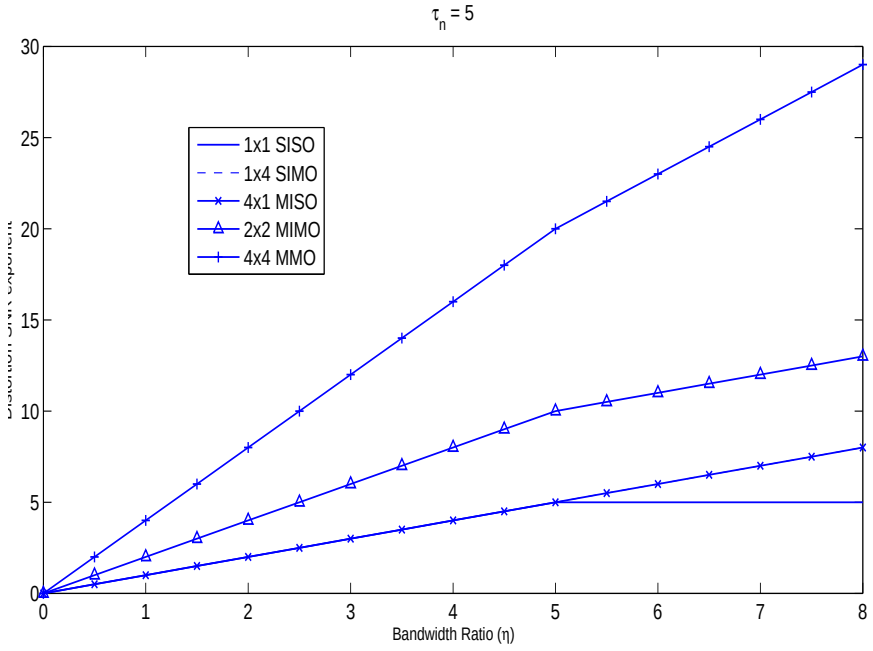

Fig. 35. Distortion exponent v.s. bandwidth ratio for normalized delay = 5.

1. MIMO Mutual Information Gaussian Approximation

Due to the difficulty in handling of Eqn. (7.14), we used some approximations for the MIMO mutual information. The mathematical operation of $\log\det(cdot)$ involved an extensive amount of average. Therefore the Lyapunov's central limit theorem can be applied. The mutual information can be approximatd as a Gaussian distribution for large antenna systems. In [79], the mean and variance of different antenna settings was derived. We used the results of [79] to derive the distortion-delay approximations for different antenna settings.

a. Large M_r, fixed M_t

For this case the mutual information obeys

$$\mathcal{I} \sim \mathcal{N}\left(M_t \ln\left(1+\frac{M_r\rho}{M_t}\right), \frac{M_t}{M_r}\right) . \tag{7.32}$$

The well-known moment generate function of the Gaussian distribution is $\mathrm{E}(e^{sx}) = \exp(sm_x + \frac{1}{2}s^2\sigma_x^2)$, where m_x and σ_x^2 is the mean and variance of the Gaussian variable x. By plugging (7.32) into (7.9) and after some straightforward math manipulations, we can get the effective capacity and distortion delay function as

$$E_c(\theta) = M_t\eta \ln\left(1+\frac{M_r}{M_t}\rho\right) - \frac{1}{2}\theta K \frac{M_t}{M_r}\eta^2 \tag{7.33}$$

$$\mathcal{D}(\tau_n) \leq \left[1+\frac{M_r\rho}{M_t} - \exp\left(\frac{M_t}{2M_r}(\frac{\eta^2}{\tau_n})\right)\right]^{-M_t\eta} \tag{7.34}$$

From Eqn. (7.33, 7.34), the effective capacity approaches to the ergodic capacity as $\theta \to 0$ or $M_r \to \infty$ (channel hardening). The SNR exponent is $M_t\eta$, which is the same as Theorem 4, as M_t fixed and M_r goes to infinity. Hence, the SNR exponent is determined by the first term in Eqn. (7.31). We found that the Guassian

approximation did reveal the distortion-delay tradeoff asymptotically.

b. Large M_t, fixed M_r

the mutual information obeys

$$\mathcal{I} \sim \mathcal{N}\bigg(M_r \ln\Big(1+\rho\Big), \frac{M_r\rho^2}{M_t(1+\rho)^2}\bigg) \ . \tag{7.35}$$

The effective capacity and distortion delay curve is

$$E_c(\theta) = M_r\eta\ln(1+\rho) - \frac{1}{2}\theta K\eta^2\frac{M_t}{M_r}\frac{\rho^2}{1+\rho^2} \tag{7.36}$$

$$\mathcal{D}(\tau_n) \leq \left[1+\rho-\exp\Big(\frac{M_r}{2M_t}(\frac{\eta^2}{\tau_n})\frac{\rho^2}{1+\rho^2}\Big)\right]^{-M_r\eta} \tag{7.37}$$

Again, the effective capacity approaches to the ergodic capacity as $\theta \to 0$ or $M_t \to \infty$ The SNR exponent is $M_r\eta$, which confirmed the results of Theorem 4.

c. Large M_t and M_r, Fixed $\beta = M_r/M_t$, High SNR

The mutual information obeys

$$\mathcal{I} \sim \mathcal{N}\bigg(M_t\mu(\beta,\rho), \sigma^2(\beta)\bigg) \ , \quad \beta \geq 1 \tag{7.38}$$

$$\sim \mathcal{N}\bigg(M_r\mu\Big(\frac{1}{\beta},\beta\rho\Big), \sigma^2\Big(\frac{1}{\beta}\Big)\bigg) \ , \quad \beta \leq \ 1 \ . \tag{7.39}$$

Where $\mu(\beta,\rho) = \ln\rho + F(\beta)$, $F(\beta), \sigma^2(\beta)$ are functions only depends on β. The effective capacity capacity and distortion-delay function is:

$$E_c(\theta) = M_r\eta\ln(\rho) - \theta C_1 \tag{7.40}$$

$$\mathcal{D}(\tau_n) \leq \left[\rho - C_2\right]^{-M_r\eta}, \quad \beta \geq 1 \tag{7.41}$$

$$E_c(\theta) = M_t\eta\ln(\rho) - \theta C_3 \tag{7.42}$$

$$\mathcal{D}(\tau_n) \leq \left[\rho - C_4\right]^{-M_t\eta}, \quad \beta \geq 1\ , \tag{7.43}$$

Where C_1, C_2, C_3, C_4 are some constants. As both M_r, M_t grow larger with fixed β, hence the $|M_t - M_r|$ also goes large, the SNR exponent is still $M_*\eta$.

F. Discussion and Remarks

In previous sections, we clearly characterized the distortion/delay curve. However, we depended on some ideal assumptions, e.g., that the instantaneous channel capacity is achievable and the CSI is perfectly known at the transmitter.

Remark 3. *(*Decoding Error Probability*) In previous discussions we have assumed using the the Gaussian code to achieve instantaneous capacity. In reality, we had to take the decoding error probability into account for short codewords. [69] has integrated the physical layer decoding error into the effective capacity function through a random coding error exponent. This showed that a joint queuing/coding exponent exists. Such an exponent can fit well into our distortion and delay analytical frame work.*

Remark 4. *(*Power Control*) Since we have perfect CSI at the transmitter, given an average transmission power budget, we can control the transmission power to maximize the effective capacity or minimize the end-to-end distortion for some delay constraint. In other words, the transmission power is not necessarily constant. Recent*

work [70] shows that the optimum power adaptation policy is related to the delay constraint. As the delay goes to infinity, the power control policy approaches a water-filing solution. On the contrary, for stringent delay constraints, the optimum power control policy becomes more like a "truncated channel inversion." Our future work will investigate how optimum power control affects the distortion/delay curves. Optimal power control policy for SISO/MIMO block fading has been investigated in [78, 80].

Remark 5. *(*Channel Correlation*) Although an i.i.d. block fading channel is easy to analyze and has several practical applications, this model is not always valid. It is more general and practical to consider channel correlation. We can use Jake's model to characterize the correlated channel fading process. The autocorrelation of channel gain $R(\tau)$ can be expressed as*

$$R(\tau) = J_0(2\pi f_d \tau) \ , \tag{7.44}$$

where $J_0(\cdot)$ denotes the zero-th order Bessel function of first kind and f_d represents the maximum Doppler frequency. Channel correlation will reduce the effective capacity[67]. Intuitively, correlation may cause the fading channel to stay in the bad status for a longer time compared with i.i.d. block fading. [70] shows that given a correlated fading channel with the same marginal statistics as i.i.d. case, the effective capacity of such a correlated channel is a linear shift in delay axis in logarithmic scale, the shift value is proportional to the Doppler frequency f_d. Hence the i.i.d. block fading distortion/delay tradeoff can be easily extended to the correlated case.

G. Conclusion

We investigate the fundamental problem of distortion/delay tradeoff for the analogue source transmitted over wireless fading channels. We derive a close-form analytical

formula to characterize this relationship using recently proposed effective capacity. Based on this closed-form expression, we derived an upper bound that is asymptotically tight to study the convergence behavior of the distortion/delay function for an SISO channel. We also characterized the SNR exponent of a MIMO block fading channel in the high SNR regime. Simulation results show that a small delay can result in a significant transmission power saving. The framework of this paper is applicable to a broad class application, e.g., video transmission.

CHAPTER VIII

CONCLUSION

This dissertation has studied the various topics of signal detection, interference suppression, error exponent and QoS under the framework of the multiple-antennas system. The main contributions of this dissertation are summarized as follows:

- Some near optimal low complexity coherent MIMO detectors are proposed. The detectors are suitable for both coded/uncoded MIMO communication.

- We proposed a noncoherent ML detector for OSTBC transmission. The detector efficiently utilizes the sphere search to reduce computation complexity. The performance of this receiver is within 1.5 dB of the channel perfectly known coherent receiver, yet maintains the full diversity order of the space-time code.

- We designed an efficient spacial covariance estimator to suppress the asynchronous co-channel interference in the MIMO-OFDM settings. The designed MMSE interference suppression receiver performed within $1-2$ dB of the synchronous and the channel perfectly known receiver. Moreover, the optimal MAP receiver that was incorporated with the covariance estimator is designed to show superior performance. With the designed open-loop interference suppression capability, the overall throughput of the high-density multicell wireless network can be increased significantly.

- We have derived the error exponent of a wideband relay channel with different relay schemes. Based on the computed error exponent, we optimized the power allocation and relay node placement. The performances of different relay schemes are compared based on the error exponent which provides a different

angle to denote how fast the error decay with the subcarrier channel number, which is a good complementary of the fundamental capacity analysis.

- We also analyzed the QoS – delay and distortion of a buffered transmission under the framework of the MIMO channel. From the cross-layer approach, the effective capacity concept has been used to investigate the distortion-delay tradeoff To characterize the asymptotical behavior, a tight upper bound was derived for special SISO case. For a more general MIMO channel, the SNR exponent was computed for different buffer delay values which represented the decay rate of the distortion with the increased high SNR value. The result provided guidelines for the system design that must satisfy some QoS requirements.

Multiple-antenna communication have attracted considerable research interest in the past few years. Moving from SISO communication to the MIMO channel is a revolutionary step in wireless communication. This dissertation is our effort in that direction.

REFERENCES

[1] I. E. Telatar, "Capacity of multi-antenna gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, Nov/Dec 1999.

[2] L. Zheng and D. Tse, "Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels," *IEEE Transactions on Information Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[3] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge: Cambridge University Press, U.K., 2005.

[4] E. C. Van Der Meulen, "Three-terminal communication channels," *Adv. Appl. Prob.*, vol. 3, pp. 120–154, 1971.

[5] T. J. Marzetta and B. M. Hochwald, "Capacity of mobile multiple-antenna communication link in a rayleigh flat-fading environment," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 139–157, March 1999.

[6] L. Zheng and D. N. Tse, "Communication on the grassmann manifold: a geometric approach to the noncoherent multiple-antenna channel," *IEEE Transactions on Information Theory*, vol. 48, no. 2, pp. 359–383, February 2002.

[7] B. Hassabi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless link," *IEEE Transactions on Information Theory*, vol. 49, no. 4, pp. 951–963, April 2003.

[8] B. M. Hochwald and S. T. Brink, "Achieving near-capacity on a multiple-antena channel," *IEEE Transactions on Communications*, vol. 51, pp. 389–399, March 2003.

[9] V. Tarokh, N. Seshadri and A. R. Calderbank, "Space-time block code from othogonal designs," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1456–1467, July 1999.

[10] S. Alamouti, "A simple transmitter diversity scheme for wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1451–1458, October 1998.

[11] B. Hassabi and B. M. Hochwald, "High-rate codes that are linear in space and time," *IEEE Transactions on Information Theory*, vol. 48, no. 7, pp. 1804–1824, July 2002.

[12] M. O. Damen, H. E. Gamal and G. Caire, "On maximum-likelihood detection and search for the closest lattice point," *IEEE Transactions on Information Theory*, vol. 49, pp. 2389–2402, October 2003.

[13] B. Hassabi and H. Vikalo, "On the sphere-decoding algorithm I. expected complexity," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 2806–2818, August 2005.

[14] B. Hassabi and H. Vikalo, "On the sphere-decoding algorithm II. generalization, second-order statistics, and applications to communicatinos," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 2819–2834, August 2005.

[15] J. Jalden and B. Ottersten, "On the complexity of sphere decoding in digital communication," *IEEE Transactions on Signal Processing*, vol. 53, no. 4, pp. 1474–1484, April 2005.

[16] H. Yao, "Efficien signal, code and receiver design for mimo communication systems," 2003, Ph.D. Dissertation, Massachusetts Institute of Technology, Cam-

bridge.

[17] A. Matache and R. D. Wesel, "Universal trellis codes for diagonally layered space-time systems," *IEEE Transactions on Signal Processing*, vol. 51, pp. 2773–2783, November 2003.

[18] G. J. Foschini, "Layered space-time architecture for wireless communications in a fading environment when using multielement antennas," *Bell Labs Technical Journal*, vol. 1, no. 6, pp. 41–59, Autumn 1996.

[19] G. D. Golden, G. J. Foschini, R. A. Valenzuela and P. W. Wolniansky, "Detection algorithm and initial laboratory results using the v-blast space-time communication architecture," *IEEE Electronics Letters*, vol. 35, no. 1, pp. 14–15, January 1999.

[20] J. A. Fessler and A. O. Hero, "Space-alternating generalized expectation-maximization algorithm," *IEEE Transactions on Signal Processing*, vol. 42, no. 10, pp. 2664–2677, October 1994.

[21] Y. Z. Xie, Q. Li and C. N. Georghiades, "On some near optimal low complexity detectors for mimo fading channels," vol. 4, no. 4, pp. 1182–1186, April 2007.

[22] A. P. Dempster, N. M. Laird and D. B. Rubin, "Maximum-likelihood from incomplete data via the em algorithm," J. Roy. Statist. Soc., vol. 39, 1977.

[23] M. Feder and E. Weinstein, "Parameter estimation of superimposed signals using the em algorithm," *IEEE Transactions on Signal Processing*, vol. 36, no. 4, pp. 477–489, April 1998.

[24] H. Dai and H.V.Poor, "Iterative space-time processing for multiuser detection in multipath cdma channels," *IEEE Transactions on Signal Processing*, vol. 50,

no. 9, pp. 2116–2127, September 2002.

[25] Y. Xie and C. N. Georghiadesr, "Two em-based channel estimation algorithms for ofdm with transmitter diversity," *IEEE Transactions on Communications*, vol. 51, no. 1, pp. 106–115, January 2003.

[26] P. Spasojevic, "Sequence and channel estimation for channels with memory," 1999, Ph.D. Dissertation, Texas A&M University, College Station.

[27] H. Artes, D. Seethaler and F. Hlawatsch, "Efficient detection algorithms for MIMO channels: A geometrical approach to approximate ML detection," *IEEE Transactions on Signal Processing*, vol. 51, no. 11, pp. 2808–2820, November 2003.

[28] J. Benesty, Y. Huang and J. Chen, "A fast recursive algorithm for optimal sequential signal detection in a blast system," *IEEE Transactions on Signal Processing*, vol. 51, no. 7, pp. 1722–1730, July 2003.

[29] M. O. Damen, K. Abed-Meraim and M. S. Lemdani, "Further results on the sphere decoder," in *IEEE ISIT*, Washington D. C., June 2001, p. 333.

[30] B. M. Hochwald and T. L. Marzetta, "Unitary space-time modulation for multiple-antenna communication in rayleigh flat fading," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 543–564, March 2000.

[31] Wing-Kin Ma, P. C. Ching, T. N. Davidson and X.-G. Xia, "Blind maximum-likelihood decoding for orthogonal space-time block codes: A semidefinite relaxation approach," in *IEEE Globecom*, San Francisco, CA., December 2003, pp. 2094 – 2098.

[32] P. Stoica and G. Ganesan, "Space-time block code: Trained, blind and semi-blind detection," in *IEEE ICASSP*, Orlando, FL., May 2002, pp. 1609 – 1612.

[33] M. Uysal and C. N. Georghiades, "An efficient implementation of a maximum-likelihood detection for space-time block coded system," *IEEE Transactions on Communications*, vol. 51, pp. 521–524, April 2003.

[34] Qiang Li and C. N.Georghiades, "Sphere Decoding of Orthogonal Space-time Block Code for Incoherent Channels," in *IEEE International Symposium on Intelligent Control*, Cyprus, June 2005, pp. 702 – 706.

[35] G. Ganesan; P. Stoica, "Space-time diversity using orthogonal and amicable orthogonal design," in *IEEE ICASSP*, Istanbul, June 2000, pp. 2561 – 2564.

[36] U. Fincke and M. Pohst, "Improved methods for calculating vectors of short length in a lattice, including a complexity analysis," Mathematics of Computation, vol. 44, pp. 463–471, April 1985.

[37] J. Zhu, B. Metzler, X. Guo and Y. Liu, "Adaptive CSMA for scalable network capacity in high-Density WLAN: A hardware prototyping approach," in *IEEE INFOCOM*, Barcelona, Spain, April 2006, pp. 1–10.

[38] J. Winters, "Optimum combining in digital mobile radio with co-channel interference," *IEEE Journal on Selected Areas in Communications*, vol. 52, pp. 528–538, July 1984.

[39] S. Catreux, P. F. Diessen and L. J. Greenstein, "Attainable throughtput of an interference-limited multiple-input multiple-output (mimo) cellular system," *IEEE Transactions on Communications*, vol. 49, pp. 1307–1311, August 2001.

[40] S. Catreux, P. F. Diessen and L. J. Greenstein, "Simulation results for an interference-limited multiple-input mulitple-output cellular system," *IEEE Communication Letters*, vol. 4, pp. 334–336, August 2001.

[41] R. S. Blum, "Mimo capacity with interference," *IEEE Journal on Selected Areas in Communications*, vol. 21, pp. 793–801, June 2003.

[42] H. Dai, A. F. Molisch and H. V. Poor, "Downlink capcity of interference-limited mimo systems with joint detection," vol. 3, pp. 442–452, June 2003.

[43] Göran Klang, "On interference rejection in wireless multichannel systems," 2003, Ph.D. Dissertation, Kungliga Tekniska Hogskolan (Royal Institute of Technology), Stocholm Sweden.

[44] Y. Li and N. R. Sollenberger, "Adaptive antenna arrays for ofdm systems with cochannel interference," *IEEE Transactions on Communications*, vol. 47, pp. 217–229, Febuary 1999.

[45] T. A. Thomas and F. W. Vook, "Asynchronous interference suppression in broadband cyclic-prefix communication," in *IEEE WCNC*, New Orleans, LO, March 2003, pp. 568–572.

[46] A. Maltsev, R. Maslennikov and A. Khoryaev, "Comparative analysis of spatial covariance matrix estimation methods in ofdm communication systems," in *IEEE Symposium on Signal Processing and Information Technology (ISSPIT)*, Vancouver, BC, August 2006, pp. 551–555.

[47] Qiang Li, Jing Zhu, Xingang Guo and C. N. Georghiades, "Asynchronous Co-Channel Interference Suppression in MIMO-OFDM Systems," in *IEEE ICC*, Glasgow, Scotland, June 2007, pp. 5744–5750.

[48] B. Bangerter, E. Jacobsen, M. Ho, A. Stephens, A. Maltsev, A. Rubtsov and A. Sadri, "High-throughput wireless lan air interface," Intel Technology Journal, vol. 7, pp. 47– 57, August 2003.

[49] Stein, "Some Problems in Multivariate Analysis, Part I," 1956, Technical Report, Dept. of Statistics, Stanford University.

[50] M. Pourahmadi, M. J. Daniels and T. Park, "Simultaneous modelling of the cholesky decomposition of several covariance matrices," Journal of Multivariate Analysis, 2006, To Appear.

[51] "*TGn channel models* ieee 802.11-03/940r4," May 2004, [Online] Availiable: http://www.802wirelessworld.com.

[52] N. Seshadri Y. Li and S. Ariyavisitakul, "Channel estimation for ofdm systems with transmitter diveristy in mobile wireless channels," *IEEE Transactions on Communications*, vol. 17, pp. 461–471, March 1999.

[53] Y. Li, "Simplified channel estimation for ofdm systems with multiple transmit antennas," *IEEE Transactions on Wireless Communications*, vol. 1, pp. 67–75, January 2002.

[54] T. M. Cover and A. A. El Gammal, "Capacity theorem for the relay channel," *IEEE Transactions on Wireless Communications*, vol. I-25, pp. 572–584, September 1979.

[55] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Transactions on Information Theory*, vol. 50, pp. 3062–3080, December 2004.

[56] Y. Liang and V. V. Veeravalli, "Gaussian orthogonal relay channel: Optimal resource allocation and capacity," *IEEE Transactions on Information Theory*, vol. 51, pp. 3284–3289, September 2005.

[57] S. Verdu, "Spectral efficiency in the wideband regime," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1319–1343, June 2002.

[58] R. G. Gallager, *Information Theory and Reliable Communication*, New York: Wiley, 1968.

[59] Qiang Li and C. N. Georghiades, "The error exponent of wideband relay channel," in *EUSIPCO*, Florence, Italy, September 2006, pp. 500–505.

[60] T. S. Rappaport, *Wireless Communications: Principles and Practice, 2nd ed.*, Upper Saddle River, NJ: Prentice Hall PTR, 2001.

[61] T. Guess and M. K. Varanasi, "Error exponent for maximum-likelihood and successive decoders for the gaussian cdma channel," *IEEE Transactions on Information Theory*, vol. 46, pp. 1683–1691, July 2000.

[62] I.S. Gradshteyn and I.M. Ryzhik, *Table of Integral, Series, and Products*, New York: Academic Press, 1992.

[63] G. Caire and K. Narayanan, "On the distortion snr exponent of hybrid digitalanalog spacetime coding," *IEEE Transactions on Information Theory*, vol. 53, pp. 2867–2878, August 2007.

[64] J. N. Laneman, E. Marinian, G. W. Wornell and J. G. Apostolopoulos, "Source-channel diveristy for parallel channels," *IEEE Transactions on Information Theory*, vol. 51, pp. 3518–3559, October 2005.

[65] D. Gunuz and E. Erkip, "Joint source-channel codes for mimo block fading channels," *IEEE Transactions on Information Theory*, Submitted.

[66] T. Holliday and A. Goldsmith, "Optimizing end-to-end distortion in mimo systems," in *IEEE ISIT*, Adelaide, SA, September 2005, pp. 1671–1675.

[67] D. Wu and R. Negi, "Effective capacity: A wireless link model for support of quality of service," *IEEE Transactions on Wireless Communications*, vol. 2, no. 4, pp. 630–643, July 2003.

[68] C.-S. Chang and J. A. Thomas, "Effective bandwidth in high-speed digital networks," *IEEE Journal on Selected Areas in Communications*, vol. 13, pp. 1091–1100, Aug. 1995.

[69] R. Negi and S. Goel, "An information-theoretic approach to queue in wireless channels with large delay bounds," in *IEEE Globecom*, Dallas, TX, November 2004, pp. 116–122.

[70] J. Tang and X. Zhang, "Quality-of-service driven power and rate adaptation over wireless links," *IEEE Transactions on Wireless Communications*, vol. 6, pp. 3058 – 3068, August 2007.

[71] R. Berry and R. G. Gallager, "Communication over fading channels with delay constraints," *IEEE Transactions on Information Theory*, vol. 48, no. 5, pp. 1135–1149, May 2002.

[72] D. Tse, "Variable-rate lossy compression and its effects on communication networks," 1994, Ph.D. Dissertation, Massachusetts Institute of Technology, Cambridge, MA.

[73] Qiang Li and C. N. Georghiades, "On the distortion delay tradeoff for Gaussian source transmitted over fading channel," in *IEEE ISIT*, Seattle, WA, July 2006, pp. 21–25.

[74] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, New York: Wiley-InterScience, 1991.

[75] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications, 2nd ed.*, Stochastic Modelling and Applied Probability. Springer, 1998.

[76] B. Hochwald and K. Zeger, "Tradeoff between source and channel coding for video transmission," *IEEE Transactions on Information Theory*, vol. 43, pp. 1412–1424, September 1997.

[77] Z. Wang and J. B. Giannakis, "Outage mutual information of space-time mimo channels," *IEEE Transactions on Information Theory*, vol. 50, pp. 657–662, April 2004.

[78] G. Carie, G. Taricco and E. Biglieri, "Optimum power control over fading channels," *IEEE Transactions on Information Theory*, vol. 45, pp. 1468–1489, July 1999.

[79] B. Hochwald, T. Marzetta and V. Tarokh, "Multiple-antenna channel hardening and its implications for rate feedback and scheduling," *IEEE Transactions on Information Theory*, vol. 50, pp. 1893–1909, September 2004.

[80] E. Biglieri and G. Caire and G. Taricco, "Limiting performance of block-fading channels with mulitple antennas," *IEEE Transactions on Information Theory*, vol. 47, no. 4, pp. 1273–1289, May 2001.

APPENDIX A

PROOF OF THEOREM 1 OF CHAPTER IV

Proof. For SIMO case, $\mathbf{H}$ is a column vector.

$$\begin{aligned}
\hat{x} &= \mathbf{W}_{MMSE} \cdot \mathbf{Y} \\
&= \mathbf{H}^H(\mathbf{H}\mathbf{H}^H + \hat{\mathbf{R}}_{II})^{-1}\mathbf{H} \cdot x + \mathbf{H}^H(\mathbf{H}\mathbf{H}^H + \hat{\mathbf{R}}_{II})^{-1} \cdot \mathbf{I}_{inter} \\
&\overset{(a)}{=} \mathbf{H}^H\left(\hat{\mathbf{R}}_{II}^{-1} - \frac{\hat{\mathbf{R}}_{II}^{-1}\mathbf{H}\mathbf{H}^H\hat{\mathbf{R}}_{II}^{-1}}{1+\mathbf{H}^H\hat{\mathbf{R}}_{II}^{-1}\mathbf{H}}\right)\mathbf{H} \cdot x + \mathbf{H}^H\left(\hat{\mathbf{R}}_{II}^{-1} - \frac{\hat{\mathbf{R}}_{II}^{-1}\mathbf{H}\mathbf{H}^H\hat{\mathbf{R}}_{II}^{-1}}{1+\mathbf{H}^H\hat{\mathbf{R}}_{II}^{-1}\mathbf{H}}\right) \cdot \mathbf{I}_{inter} \\
&= \frac{\mathbf{H}^H\hat{\mathbf{R}}_{II}^{-1}\mathbf{H}}{1+\mathbf{H}^H\hat{\mathbf{R}}_{II}^{-1}\mathbf{H}} \cdot x + \frac{\mathbf{H}^H\hat{\mathbf{R}}_{II}^{-1} \cdot \mathbf{I}_{inter}}{1+\mathbf{H}^H\hat{\mathbf{R}}_{II}^{-1}\mathbf{H}} \ , \qquad \text{(A.1)}
\end{aligned}$$

where (a) has used the matrix inverse lemma. Hence,

$$\hat{x} - x = \frac{-1}{1+\mathbf{H}^H\hat{\mathbf{R}}_{II}^{-1}\mathbf{H}} \cdot x + \frac{\mathbf{H}^H\hat{\mathbf{R}}_{II}^{-1} \cdot \mathbf{I}_{inter}}{1+\mathbf{H}^H\hat{\mathbf{R}}_{II}^{-1}\mathbf{H}} \qquad \text{(A.2)}$$

$$\begin{aligned}
MSE = E(\hat{x}-x)^2 &= \frac{1}{(1+\mathbf{H}^H\hat{\mathbf{R}}^{-1}\mathbf{H})^2} + \frac{\mathbf{H}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{H}}{(1+\mathbf{H}^H\hat{\mathbf{R}}^{-1}\mathbf{H})^2} \\
&= \frac{1}{(1+\mathbf{H}^H\hat{\mathbf{R}}^{-1}\mathbf{H})} + \frac{\mathbf{H}^H\hat{\mathbf{R}}^{-1}\Delta\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{H}}{(1+\mathbf{H}^H\hat{\mathbf{R}}^{-1}\mathbf{H})^2} \\
&= \hat{MSE} + \hat{MSE}^2 \cdot \mathbf{H}^H\hat{\mathbf{R}}^{-1}\Delta\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{H} \\
&= \hat{MSE} + \hat{MSE}^2 \cdot \mathrm{trace}\big(\mathbf{H}^H\hat{\mathbf{R}}^{-1}\Delta\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{H}\big) \qquad \text{(A.3)}
\end{aligned}$$

where $\Delta\mathbf{R} = \mathbf{R} - \hat{\mathbf{R}}$. Let $\hat{\mathbf{R}}^{-1} = \mathbf{U}^H\mathbf{U}, \quad \Delta\mathbf{R} = \mathbf{T}^H\mathbf{T}$.

From (A.3), we have following inequality:

$$MSE \le \hat{MSE} + \hat{MSE}^2 \cdot \|\mathbf{U}\mathbf{T}^H\mathbf{T}\mathbf{U}^H\|_F\|\mathbf{U}\mathbf{H}\mathbf{H}^H\mathbf{U}^H\|_F \ . \qquad \text{(A.4)}$$

wherein,

$$\begin{aligned}
\|\mathbf{U}\mathbf{H}\mathbf{H}^H\mathbf{U}^H\|_F &= \sqrt{\text{trace}(\mathbf{H}\mathbf{H}^H\mathbf{U}^H\mathbf{U}\mathbf{H}\mathbf{H}^H\mathbf{U}^H\mathbf{U})} \\
&= \sqrt{\mathbf{H}^H\hat{\mathbf{R}}^{-1}\mathbf{H}\mathbf{H}^H\hat{\mathbf{R}}^{-1}\mathbf{H}} \\
&= \frac{1}{\hat{MSE}} - 1 \ . \\
\|\mathbf{U}\mathbf{T}^H\mathbf{T}\mathbf{U}^H\|_F &= \sqrt{\text{trace}(\mathbf{U}\mathbf{T}^H\mathbf{T}\mathbf{U}^H\mathbf{U}\mathbf{T}^H\mathbf{T}\mathbf{U}^H)} \\
&= \sqrt{\text{trace}(\Delta\mathbf{R}\hat{\mathbf{R}}^{-1}\Delta\mathbf{R}\hat{\mathbf{R}}^{-1})} \\
&= \|\hat{\mathbf{R}}^{-1}\Delta\mathbf{R}\|_F \\
&\leq \|\Delta\mathbf{R}\|_F \cdot \lambda_{max}(\hat{\mathbf{R}}^{-1}) \ . \qquad \text{(A.5)}
\end{aligned}$$

where we have used the Cauchy-Schwarz inequality $\text{trace}(\mathbf{AB}) \leq \|\mathbf{A}\|_F \cdot \|\mathbf{B}\|_F$, and $\|\mathbf{AB}\|_F \leq \|\mathbf{B}\|_F\lambda_{max}(\mathbf{A})$. Therefore,

$$MSE \leq \hat{MSE} + \left(\hat{MSE} - \hat{MSE}^2\right) \cdot \|\Delta\mathbf{R}_{II}\|_F \cdot \lambda_{max}(\hat{\mathbf{R}}_{II}^{-1}). \qquad \text{(A.6)}$$

□

APPENDIX B

PROOF OF THEOREM 2 OF CHAPTER VI

Proof. From Eqn. (7.19) of Corollary, we have

$$\begin{aligned}\mathcal{D}(\lambda) &\leq \left[\rho^{-\lambda}\exp\Big(\frac{1}{\rho}\Big)\gamma\Big(1-\lambda,\frac{1}{\rho}\Big)\right]^{\frac{1}{\lambda}} \\ &= \left[\frac{1}{\lambda-1}\frac{1}{\rho}\,{}_1F_1\Big(1;2-\lambda;\frac{1}{\rho}\Big)+\Gamma(1-\lambda)\Big(\frac{1}{\rho}\Big)^{\lambda}\exp\Big(\frac{1}{\rho}\Big)\right]^{\frac{1}{\lambda}} \qquad \text{(B.1)}\end{aligned}$$

Since $\frac{1}{\lambda-1} < 0$ as $\lambda \to 0$, we first lower-bound the confluent hypergeometric function.

$$\begin{aligned}{}_1F_1(1;2-\lambda;x) &= \sum_{k=0}^{\infty}\frac{(1)_k}{(2-\lambda)_k}\frac{x^k}{k!} \\ &\geq \sum_{k=0}^{\infty}\frac{(1)_k}{(2)_k}\frac{x^k}{k!} = \frac{1}{x}(e^x-1)\ , \qquad \text{(B.2)}\end{aligned}$$

where $(a)_k \triangleq a\cdot(a+1)\cdots(a+k-1)$. For $\lambda \to 0$ this lower bound is asymptotically tight. Next we upper-bound the $\Gamma(1-\lambda)$.

$$\begin{aligned}\Gamma(1-\lambda) &= -\lambda\cdot\Gamma(-\lambda) = \frac{-\lambda}{\frac{1}{\Gamma(-\lambda)}} \\ &= \frac{-\lambda}{-\lambda+\xi(-\lambda)^2+\phi(-\lambda)^3+\delta(-\lambda)^4+O((-\lambda)^5)} \\ &\leq \frac{1}{1-\xi\lambda+\phi\lambda^2-\delta\lambda^3}\ , \qquad \text{(B.3)}\end{aligned}$$

where $\xi = 0.577215$, $\phi = \frac{1}{12}(6\xi^2-\pi^2)$ and δ is some constant. Hence replacing (B.2) and (B.3) in (6.21) we have the following upper bound

$$\mathcal{D}(\lambda)\dot{\leq}\left[\frac{1}{\lambda-1}\big(e^{\frac{1}{\rho}}-1\big)+\frac{1}{1-\xi\lambda+\phi\lambda^2}\rho^{-\lambda}e^{\frac{1}{\rho}}\right]^{\frac{1}{\lambda}}\ , \qquad \text{(B.4)}$$

where we have omitted $O(\lambda^3)$ term, which will not affect the result as $\lambda \to 0$. Using Taylor expansion for the first term and second term, and dropping the $O(\lambda^3)$, we

obtain the following asymptotic approximation,

$$\begin{aligned}\mathcal{D}_{upper}(\lambda) &\doteqdot [1+a\lambda+b\lambda^2]^{\frac{1}{\lambda}} \\ &= \exp(a)\exp\left((b-\frac{a^2}{2})\lambda\right) ,\end{aligned} \tag{B.5}$$

where we have used the identity $\lim_{x\to 0}(1+x)^{\frac{1}{x}} = e$, and

$$\begin{aligned}a &\triangleq 1 - e^{\frac{1}{\rho}} + \xi e^{\frac{1}{\rho}} - \ln\rho e^{\frac{1}{\rho}} \\ b &\triangleq 1 - e^{\frac{1}{\rho}} + (\xi^2-\phi)e^{\frac{1}{\rho}} - \xi\ln\rho e^{\frac{1}{\rho}} + \ln^2\rho\ .\end{aligned}$$

In order to show $\mathcal{D}_{upper}(\lambda) \to \mathcal{D}(\infty)$ in (6.17), in other word $(B.5) \to (6.17)$, we want to show that

$$F \triangleq 1 - e^{-\frac{1}{\rho}} - \xi + \ln\rho \to E_1(1/\rho)\ . \tag{B.6}$$

$E_1(\cdot)$ is a special function, and don't have simple expression. Instead we use numerical method to illustrate the convergence. We have plotted these two values in Figure 6. We can observe for most SNR these two values match perfectly. Hence we conclude that the upper bound converges and the convergent rate is exponential. □

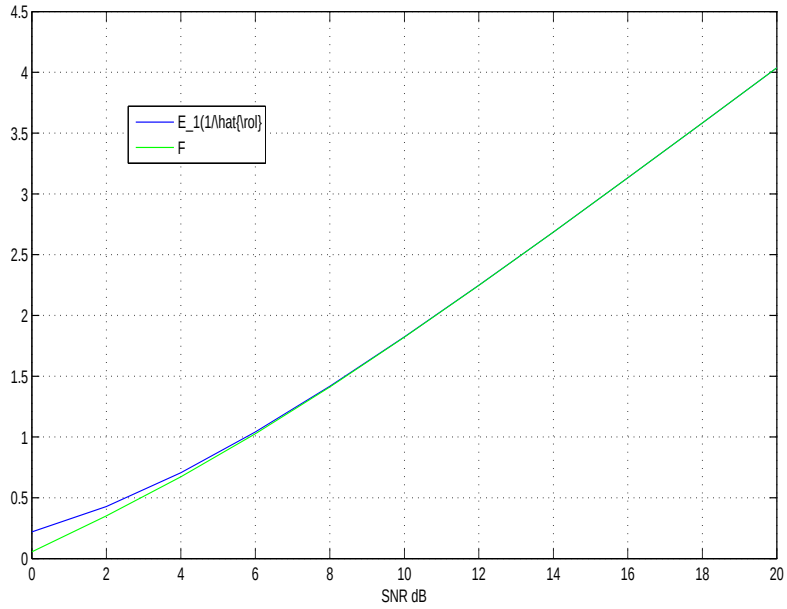

Fig. 36. Illustration (A-8) for different SNR values

APPENDIX C

PROOF OF THEOREM 4

Proof. We will follow the technique used in [2]. Assume without loss of generality that $M_t = M_* \leq M_r$ (the case $M_t > M - r$ is a simple extension). We start from the distortion delay function (7.17)

$$\mathcal{D}(\rho) = \left\{ \int_0^\infty \prod \left(1 + \frac{\rho}{M_t}\lambda_i\right)^{-\theta K \eta} f(\boldsymbol{\lambda}) d\boldsymbol{\lambda} \right\}^{\frac{1}{\theta K}} , \tag{C.1}$$

where $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{M_t}$ are the ordered eigenvalues of $\mathbf{H}\mathbf{H}^H$. We make the change of variable: $\alpha_i = -\log(\lambda_i)/\log(\rho)$, for all $i = 1, \cdots, M_t$, The joint pdf $\boldsymbol{\alpha} = [\alpha_1, \cdots, \alpha_{M_t}]$, where $\alpha_1 \geq \cdots \geq \alpha_{M_t}$, is given by

$$f(\boldsymbol{\alpha}) = K_{M_t, M_r}^{-1} \big(\log \rho\big)^{M_t} \prod_{i=1}^{M_t} \rho^{-(M_r - M_t + 1)\alpha_i} \prod_{i<j} \left(\rho^{-\alpha_i} - \rho^{-\alpha_j}\right)^2 \exp\left(\sum_i \rho^{-\alpha_i}\right) . \tag{C.2}$$

Replace $\boldsymbol{\lambda}$ with $\boldsymbol{\alpha}$, (C.1) yields

$$\mathcal{D}(\rho) = \left\{ \int_{\mathcal{A}} \prod_{i=1}^{M_t} (1 + \frac{1}{M_t}\rho^{1-\alpha_i})^{-\theta K \eta} f(\boldsymbol{\alpha}) d\boldsymbol{\alpha} \right\}^{\frac{1}{\theta K}} , \tag{C.3}$$

where

$$\mathcal{A} = \left\{ \boldsymbol{\alpha} \in \mathbb{R}^{M_t} \quad : \quad \alpha_1 \geq \cdots \geq \alpha_{M_t} \right\} .$$

Neglecting all terms that irrelevant to the SNR exponent, we obtain (C.1) yields

$$\begin{aligned}\mathcal{D}(\rho) \dot{\geq} & \left\{ \int_{\mathcal{A} \bigcap \mathbb{R}^{M_t}+} \left(\prod_{i=1}^{M_t} (1+\frac{1}{M_t}\rho^{1-\alpha_i})^{-\theta K \eta} \right) \prod_{i=1}^{M_t} \rho^{-(2i-1+M_r-M_t)\alpha_i} d\boldsymbol{\alpha} \right\}^{\frac{1}{\theta K}} \\ \dot{=} & \left\{ \int_{\mathcal{A} \bigcap \mathbb{R}^{M_t}+} \prod_{i=1}^{M_t} \rho^{-\theta K \eta (1-\alpha_i)^+} \prod_{i=1}^{M_t} \rho^{-(2i-1+M_r-M_t)\alpha_i} d\boldsymbol{\alpha} \right\}^{\frac{1}{\theta K}} \\ \dot{=} & \left\{ \int_{\mathcal{A} \bigcap \mathbb{R}^{M_t}+} \prod_{i=1}^{M_t} \rho^{-(\theta K \eta (1-\alpha_i)^+ + (2i-1+M_r-M_t)\alpha_i)} \right\}^{\frac{1}{\theta K}} \\ \dot{=} & \rho^{\alpha(\eta)\frac{1}{\theta K}} \end{aligned} \tag{C.4}$$

where we have used

$$(1+\frac{1}{M_t}\rho^{1-\alpha_i})^{-\theta K \eta} \dot{=} \rho^{-\theta K \eta [1-\alpha_i]^+} \ .$$

And

$$\alpha(\eta) = \inf_{\boldsymbol{\alpha} \in \mathcal{A} \bigcap \mathbb{R}^{M_t}+} \sum_{i=1}^{M_r} (2i-1+M_r-M-t)\alpha_i + \theta K \eta (1-\alpha_i)^+ \ .$$

We can minimizing individual term of the summation separately by set $\alpha_i = 0$ or 1. We also notice that $\theta K = \tau_n$, the buffer delay, hence we can obtain the SNR exponent of the buffered transmission is

$$\alpha(\eta) = \tau_n \min \left\{ \frac{\eta}{\tau_n}, 2i-1+M_r-M_t \right\} . \tag{C.5}$$

□

VITA

Qiang Li was born in Dongyang, Zhejiang, China. He studied six years at Shanghai Jiaotong University(SJTU) for his Bachelor's degree majoring in biomedical engineering and Master's degree in electrical engineering. After graduation from SJTU in 2000, he spent more than one year with Lucent Technologies in Shanghai. He came to College Station, Texas, in the fall of 2002 to study for his Ph.D in the Electrical Engineering Department majoring in wireless communications. He also worked as a research intern for six months in Intel's radio communication lab in 2006. He received his Ph.D. from Texas A&M University in December 2007.