UNIVERSITY OF CALIFORNIA

Los Angeles

Coding Techniques for High Data Rates in Wireless Multiple-Input Multiple-Output Communications

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Electrical Engineering

by

Adina Matache

2004

The dissertation of Adina Matache is approved.

Michael Fitz

Kung Yao

Chris Anderson

Richard D. Wesel, Committee Chair

University of California, Los Angeles

2004

To the memory of my grandmother,

Floarea Mârlogeanu

Contents

List of Figures

List of Tables

List of Acronyms

AWGN	Additive white Gaussian noise
BICM	Bit-interleaved coded modulation
BER	Bit-error rate
BPSK	Binary phase shift keying
CSI	Channel state information
FER	Frame-error rate
GDFE	Generalized decision feedback equalizer
LAN	Local area network
LDPC	Low-density parity-check
LLR	Log-likelihood ratio
MAP	Maximum a posteriori
MI	Mutual information
MIMO	Multiple-input multiple-output
ML	Maximum-likelihood
MLC	Multi-level coding
MMSE	Minimum mean-square error
MMSE-SIC	Minimum mean-square error soft interference cancellation
MRC	Maximal ratio combining

MRC-HIC	Maximal ratio combining hard interference cancellation
MRC-SIC	Maximal ratio combining soft interference cancellation
MSE	Mean-square error
OFDM	Orthogonal frequency division multiplexing
QAM	Quadrature amplitude modulation
QPSK	Quadrature phase-shift keying
PSK	Phase-shift keying
PSP	Per-survivor processing
RED	Residual Euclidean distance
SISO	Single-input single-output
SNR	Signal-to-noise ratio
TFB	Transfer function bound
TST	Threaded space-time
ZF	Zero-forcing
WSTC	Wrapped space-time code

ACKNOWLEDGEMENTS

It has been a pleasure to work for Professor Rick Wesel and I feel very fortunate to have had him as my adviser. I thank him for his guidance, his teachings, his help, and support. His passion and excitement for research is very inspiring. He is a great source of motivation and encouragement, and he praises his students for a job well done. I also thank him for his patience and support during my slow and painful recovery from the two accidents (one mountain biking, the other one skiing, one year later), which resulted in two fractured shoulders and a setback in my research progress.

I thank the committee members for their time and their valuable suggestions. A special thanks is in order for Professor Mike Fitz for his suggestion to explore Viterbi decoding with "not just-in-time" traceback in the D-BLAST scheme in Chapter 2. I benefitted from the discussions we had regarding my research and the integration of the LDPC and MIMO detectors software with the MIMO OFDM C-chain. I would also like to thank Gerard Foschini for his instrumental support of the D-BLAST related work and for his insightful comments and suggestions.

I cannot thank enough my colleague Chris Jones, for providing not only the LDPC codes used in this dissertation, but also some of the tools developed by him, including the simulation software for LDPC coded MIMO systems with MAP detection. Many of the findings presented in Chapter 4 are the result of our collaboration. I also thank Chris for the many interesting talks we had about research, work, and life in general.

I would like to extend special thanks to my research group members, Aditya Ra-

mamoorthy, Cenk Köse, Jun Shi, Wen-Yen Weng, Xueting Liu, Andres Vila, Mark Shane, Wei Shi, and Tom Sun for their friendship and support. In particular, I would like to thank my cube neighbor, Jun, for his kindness and his great sense of humor. Also, I thank Cenk for being there whenever I needed someone to talk to. I have always enjoyed our technical and not-so-technical discussions over coffee.

Some very special friendships have developed during my time at UCLA. I met Katherine Comanor during my second year and we have been close friends ever since. Her friendship has brightened my days and helped me survive graduate school. We have had great fun talking about anything and everything, going to movies, concerts, and dinners together, and also hiking and skiing on occasions.

I thank Brien Alkire for his friendship, compassion, and kindness, and for a great time skiing together in Colorado. I thank Christian Oberli for making my first year a little more enjoyable and fun. I also thank him for taking me to the ER room when I fractured my shoulder mountain biking and for waiting with me patiently for 5-6 hours.

To my dear friend Gabriela Alexandru, whom I have known since high-school years, many thanks for her wonderful friendship, for sharing with me many happy and sad moments, and for being my traveling companion in an unforgettable adventure in Patagonia, Chile.

Last but not least, I thank from the bottom of my hart my dearest parents, Dorina and George, and my brother Florin. I could not have completed this work without their support, encouragement, endless care and affection. They have always been there for

me and provided unconditional love and support. I would especially like to thank my mother, for taking care of me not only when I was sick, but also on a day-by-day basis. My parents are the reason why I made it so far. I thank them for sacrificing everything so I can follow my dream.

VITA

Dec. 28, 1971	Born, Bucharest, Romania
1995	B.S. in Electrical Engineering University of Washington Seatlle, Washington Magna cum Laude
1995-1997	Graduate Teaching Assistant Department of Electrical Engineering University of Washington
1997	M.S. in Electrical Engineering University of Washington Seattle, Washington
1997-2002	Engineer Jet Propulsion Laboratory Pasadena, California
2000-2004	Graduate Student Researcher Department of Electrical Engineering University of California, Los Angeles

PUBLICATIONS

Jones, C., Matache, A., Tian, T., Villasenor, J., and Wesel, R. D., "The Universality of LDPC Codes on Wireless Channels," *MilCom 2003*, Boston, MA, Oct. 2003.

Jones, C., Tian, T., Matache A., Wesel, R. D., and Villasenor, J., "Robustness of LDPC Codes on Periodic Fading Channels," *IEEE Globecom*, Taipei, Taiwan, Nov. 2003.

Kiely, A., Dolinar, S., Klimesh, M., and Matache, A., "Error Containment in Com-

pressed Data using Sync Markers," *IEEE International Symposium on Information Theory*, June 2000.

Kiely, A., Dolinar, S., Klimesh, M., and Matache, A., "Synchronization Markers for Error Containment in Compressed Data," *JPL TMO Progress Report 42-136*, Feb. 1999.

Manduchi, R., Dolinar, S., Pollara, F., and Matache, A., "Onboard Science Processing and Buffer Management for Intelligent Deep Space Communications," *Aerospace Conference Proceedings*, March 2000.

Manduchi, R., Dolinar, S., Pollara, F., and Matache, A., "Semantic Progressive Transmission for Deep Space Communications," *Data Compression Conference*, March 2000.

Matache, A., Dolinar, S., and Pollara, F., "Stopping Rules for Turbo Decoders," *JPL TMO Progress Report 42-142*, Aug. 2000.

Matache, A., and Ritcey, J. A., "Optimum Code Rates for Noncoherent MFSK with Errors and Erasures Decoding over Rayleigh Fading Channels," *Proceedings 1997 Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove CA, Nov. 1997.

Matache, A. and Wesel, R. D., "Trellis Coding for Layered Space-Time Systems," *Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 2001.

Matache, A. and Wesel, R. D., "Universal Trellis Codes for Diagonally Layered Space-Time Systems," *IEEE Transactions on Signal Processing*, Nov. 2003, vol. 51, no. 11, pp 2773-2783, Special Issue on MIMO Wireless.

Matache, A., Wesel, R. D., and Shi, J., "Trellis Coding for Diagonally Layered Space-Time Systems," *ICC 2002*, New York, April 2002.

ABSTRACT OF THE DISSERTATION

Coding Techniques for High Data Rates in Wireless Multiple-Input Multiple-Output Communications

by

Adina Matache

Doctor of Philosophy in Electrical Engineering

University of California, Los Angeles, 2004

Professor Richard D. Wesel, Chair

The ever-increasing demand for reliable high-data-rate communication over the wireless channel has lead to the development of efficient modulation and coding schemes. Systems employing multiple antennas for both transmitting and receiving promise a substantial increase in capacity over single antenna systems.

In this dissertation we propose several coding techniques for transmission over multiple-input multiple-output (MIMO) channels. One contribution of this work is a simple coding technique, which uses a single trellis code with finite-traceback Viterbi decoding for Foschini's diagonally layered space-time transmission system known as D-BLAST. We examine the performance of universal trellis codes designed to have a distance structure that is matched to the periodic signal-to-noise ratio variation of the

channel created by D-BLAST, under the assumption that the channel is static during one burst but may change from burst to burst. We prove that D-BLAST under a minimum mean-square error (MMSE) interference suppression criterion is theoretically optimal and achieves the MIMO channel capacity. Under a constrained input scenario and a Gaussian approximation on the MMSE filter output, the parallel channels created by D-BLAST have an aggregate capacity that is approximately equal to the constrained MIMO capacity. For large antenna configurations, computing the constrained ergodic capacity is numerically intensive; however, the D-BLAST technique can be used to drastically reduce the complexity of such calculations from exponential to linear in the number of antennas.

Another contribution of this work is the application of low-density parity-check (LDPC) codes for multiple-antenna wireless channels. We introduce a scheme using a single LDPC code spatially multiplexed on multiple antennas and an iterative detection and decoding receiver. We discuss several reduced-complexity soft demodulation techniques suitable for systems that achieve very high spectral efficiencies. We compare the different detection schemes in terms of their performance and complexity. We also demonstrate that properly designed LDPC codes behave very much like universal codes in the sense that their performance lies in close proximity to the Root and Varaiya compound channel capacity for the linear Gaussian vector channels for all but a few channels.

Chapter 1

Introduction

The ever-increasing demand for reliable high-data-rate communication over the wireless channel has lead to the development of efficient modulation and coding schemes. The wireless channel suffers from time-varying impairments like multi-path fading, interference, and noise. Diversity techniques (time, frequency, space, polarization) are an effective method to combat the fading in wireless channels. Time and frequency diversity yield a loss in bandwidth efficiency. However, by employing multiple antennas at the transmitter and/or receiver, spatial diversity is achieved without sacrificing the precious bandwidth resource. Moreover, by employing multiple antennas for both the transmitting and receiving, a substantial increase in capacity is possible over the single antenna systems. The basic information theory results reported by Foschini and Gans [13] and independently by Telatar [40] have indeed promised extremely high spectral efficiencies through multiple element antenna arrays at both the transmitter and

receiver.

In this dissertation we propose several coding techniques for transmission over multiple-input multiple-output (MIMO) channels and present analytical and simulation results for the coded MIMO system under an idealized propagation scenario. One contribution of this work is a simple coding technique which uses a single trellis code with finite-traceback Viterbi decoding for Foschini's diagonally layered space-time transmission system (better known as D-BLAST). In his original work [12], Foschini proposed D-BLAST as a receiver architecture designed for uncoded transmission over MIMO systems. With zero-forcing (ZF) interference suppression, this architecture has an "operational" capacity equal to a lower bound on the Shannon channel capacity. However, D-BLAST is theoretically capacity-achieving if zero-forcing is replaced by minimum mean-square error (MMSE) interference suppression. In essence, MMSE D-BLAST converts a linear MIMO channel into a periodic scalar channel with the same capacity.

In Chapter 2 we examine the performance of universal trellis codes for the periodic fading channel in conjunction with *full-overhead* and *reduced-overhead* versions of the diagonally layered space-time receiver architecture. A universal trellis code can (to the extent its blocklength and complexity will permit) approach the performance predicted by the compound channel information theory for linear Gaussian channels. Root and Variaya [36] proved that for a given rate and input distribution there exists a code that will reliably transmit that rate over every linear Gaussian channel for which the input

distribution induces a mutual information higher than the code rate. In other words, such a code works on every channel that it possibly could, and on each channel it works as if it were designed specifically for that channel. In [48] Wesel *et al.* constructed universal trellis codes for periodic erasure channels, which we show to be universal for more general periodic fading. We also show that these 64-state universal trellis codes have frame error rates within 1.8 - 3.6 dB of the quasistatic Rayleigh fading outage capacity of the periodic fading channel created by D-BLAST. An additional loss of 0.6 - 1.0 dB is incurred due to the overhead penalty associated with the diagonal layering.

At the same time as our initial results in [33], Caire and Colavolpe [9] proposed a similar space-time coding technique for the quasistatic multiple antenna channel, called the "wrapped" space-time coding (WSTC) scheme. The main difference in this work is that the decoding of a "wrapped" space-time code is via Viterbi's algorithm through the use of per-survivor processing (PSP) and that the component codes for the WSTC scheme are off-the-shelf convolutional codes, not designed for the periodic channel created by D-BLAST.

Recently, El Gamal and Hammons [17] proposed an iterative coding technique for the layered space-time systems which uses n constituent encoders, one for each layer. The main advantage of this scheme, which is referred to as the threaded space-time (TST) architecture, is that it does not incur the overhead penalty inherent in the initial pilots required by diagonal layering. One potential drawback of the TST architecture is the complexity of its iterative receiver which combines a SISO multi-user detector mod-

ule and n separate SISO channel decoders for each of the component channel codes. The TST results reported in [17] are within 2 - 3 dB of the outage capacity. For higher number of transmit/receive antennas, the TST architecture of El Gamal and Hammons has 1.4 - 2.0 dB better average performance in quasistatic Rayleigh fading than our scheme. The iterative MMSE receiver architecture of TST provides better detection and has no overhead penalty.

Another factor that contributes to a better average performance in the TST architecture is the use of smaller QPSK constellations. However, our larger 16-QAM constellations permit the code to operate on at least some channels that are singular. Such singular channels would force uncoded performance for QPSK systems such as that of El Gamal and Hammons. Another recent work includes the multilayered space-time architecture proposed by Tarokh, Naguib, Seshadri, and Calderbank [1]. The performance obtained with this scheme is 6 dB or more from the outage capacity.

In Chapter 3 we prove that D-BLAST under an MMSE interference suppression criterion is theoretically optimal and achieves the Shannon channel capacity. This is an alternative, simpler proof than the one given by Ariyavisitakul in [7]. Under a constrained input scenario and a Gaussian approximation on the MMSE filter output, the parallel channels created by D-BLAST have an aggregate capacity that is approximately equal to the constrained modulation MIMO capacity. For large antenna configurations, computing the constrained capacity is numerically intensive; however, the D-BLAST technique can be used to drastically reduce the complexity of such calculations from

exponential to linear in the number of antennas.

In Chapter 4, we introduce a scheme using a single low-density parity-check (LDPC) code spatially multiplexed on multiple antennas and an iterative detection and decoding receiver. Recent research has shown that combining iterative signal processing and decoding techniques can achieve near-capacity performance on multiple-input multiple-output channels. This iterative detection and decoding method is based on the turbo decoding principle and is often referred to as Turbo-BLAST [37], [38], [43]. Several techniques for low-complexity, high-performance iterative detection and decoding receivers are available in the literature. In [37] and [38] the authors introduce a sub-optimal receiver based on a minimum mean-square error (MMSE) soft interference cancellation detector. In [41] a "list" sphere decoder is used to iteratively detect and decode either simple convolutional or more powerful turbo codes. In [5] the authors present an iterative-greedy demodulation-decoding technique for turbo codes based on a greedy detection method for multiuser communications. More recently in [28] the authors present an LDPC coded MIMO OFDM system using either the optimal soft maximum *a posteriori* (MAP) demodulator or the low complexity minimum mean-square error soft interference cancellation (MMSE-SIC) demodulator. The results reported in [28] show that a system based on the MMSE-SIC detector suffers a performance degradation in comparison to a system based on the MAP detector. In our work however, we have observed very little difference in the performance of these two detection schemes for systems with a small number of transmit and receive antennas and QPSK

modulation.

In this research we focus on an LDPC coded MIMO scheme which uses a high-performance irregular LDPC code designed in our previous work [21], [20]. We discuss several reduced-complexity soft demodulation techniques suitable for systems that achieve very high spectral efficiencies. In particular, we consider the standard soft MAP detector, the MMSE suppression and MMSE-SIC detectors, and we introduce a simple maximal ratio combining (MRC) detector with either soft or hard interference cancellation. We compare the different detection schemes in terms of their performance and complexity. We also provide comparisons with the Shannon capacity limits for ergodic multiple-antenna channels. Chapter 4 also provides results that show that properly designed LDPC codes behave very much like universal space-time codes and that their performance lies in close proximity to the Root and Variaya capacity for the linear Gaussian vector channels for all but a few channels.

Chapter 2

Trellis Coded D-BLAST Systems

2.1 System Model

The system under consideration is a MIMO system with n transmit and n receive antennas. In a complex baseband representation, the basic vector equation describing the channel operating on the signal in a quasistatic scenario is

$$\mathbf{y}(t) = \mathbf{H}\,\mathbf{x}(t) + \mathbf{n}(t), \quad t = 1, \ldots, N_f \tag{2.1}$$

where N_f is the size of a frame, $\mathbf{x}(t) = [x_1(t)\, x_2(t)\ \ldots x_n(t)]^{\mathrm{T}}$ is the complex vector of modulation symbols transmitted in parallel by the n transmit antennas at symbol time t, $\mathbf{n}(t) = [n_1(t)\, n_2(t) \ldots n_n(t)]^{\mathrm{T}}$ is the vector of independent Gaussian noise samples with zero mean and variance $\sigma^2 = N_0/2$ per dimension, $\mathbf{y}(t) = [y_1(t)\, y_2(t) \ldots y_n(t)]^{\mathrm{T}}$

is the corresponding vector of received signal samples at the output of the receive antennas, and $\mathbf{H} = [\mathbf{h}_1\, \mathbf{h}_2\, \ldots \mathbf{h}_n]$ is the channel matrix.

We assume that the power emitted from each transmit antenna is proportional to $1/n$ so that the total radiated power is constant and independent of the number of transmit antennas. The average signal-to-noise ratio (SNR) at each receive antenna is denoted by ρ and is also independent of n. The path gains in the channel matrix $\mathbf{H}$ are modeled as samples of independent complex Gaussian random variables with zero mean and variance 0.5 per dimension. This is equivalent to the assumption that each path between a transmit and a receive antenna has frequency-flat independent Rayleigh fading. This models a channel with enough physical separation within the transmitting and receiving antennas to achieve independence in the entries of $\mathbf{H}$.

We consider a communication scenario in which the channel characteristics are essentially unchanged during a frame, but change randomly from one frame to another (quasistatic fading). We further assume that the receiver knows $\mathbf{H}$ perfectly, while the transmitter has no knowledge of $\mathbf{H}$. In practice, the assumption of perfect channel state information (CSI) at the receiver holds approximately when the channel varies slowly with respect to the duration of a frame. This is a realistic assumption in several wireless settings where mobility is limited or absent (e.g. indoor wireless local-area networks, wireless local loops, etc.). Furthermore, the assumption of no CSI at the transmitter is a reasonable assumption in situations where feedback or time-division duplexing cannot be exploited. However, the transmitter may use knowledge of the channel statistics

(such as average SNR) to select a transmission rate even though it has no knowledge of the particular realization of the channel matrix.

2.2 Overview of D-BLAST

The transmission process in Foschini's layered structure with n transmit and n receive antennas can be summarized as follows. The data stream is demultiplexed into n data streams of equal rate, and each substream is modulated using the same constellation. The n substreams are transmitted simultaneously using n antenna elements. Then at the receiver, the n receive antennas are used to decouple and detect the transmitted signals.

First, the transmitted signal from antenna n is treated as the desired signal, while the signals from the other transmit antennas are treated as interference. Linear processing on the received vector is then used to suppress the interference from antennas $1, \ \ldots, \ n-1$. This yields a diversity order of one. Once the signal transmitted from antenna n is correctly detected, it is canceled from the received vector. Then the signal from antenna $n-1$ is treated as the desired signal, while the signals from antennas $1, \ldots, n-2$ are treated as interference. Linear processing is used to suppress the interference from the remaining $n-2$ signals, providing a diversity order of two. This process is continued until all the transmitted signals are detected. With this scheme, the diversity order varies across the transmitted signal components from a diversity gain

of one for the signal transmitted from antenna n to a diversity gain of n for the signal transmitted from antenna one.

To summarize, the detection of transmitted signal components involves two major steps:

1. *interference canceling*: interference from already detected symbols is subtracted out.

2. *interference suppression*: interference from yet to be detected symbols is suppressed by linear operations.

The interference canceling step subtracts out from the received vector the already detected signal components, while the interference suppression step removes the interference stemming from the as yet undecoded components using linear operations optimized under a ZF or an MMSE criterion.

The interference suppression alternatives (ZF and MMSE) are described in detail next. In the remainder of the discussion, we drop the time index t for brevity. Assuming that the receiver has correctly detected the last $n - i$ signal components of $\mathbf{x}$, we can cancel the interference from these components by subtracting them out from the received vector $\mathbf{y}$. Then, the resulting vector $\mathbf{y}_i$ is

$$\mathbf{y}_i = \mathbf{y} - \sum_{k=i+1}^{n} x_k \mathbf{h}_k = \sum_{k=1}^{i} x_k \mathbf{h}_k + \mathbf{n}. \tag{2.2}$$

2.2.1 ZF Interference Suppression

Assuming the last $n - i$ signal components have been correctly detected and canceled out, the interference stemming from the simultaneous transmission of $x_1, \ldots, x_{i-1}$ is nulled out by projecting $\mathbf{y}_i$ onto the nullspace of the vectors $\mathbf{h}_1, \ldots, \mathbf{h}_{i-1}$. Let $\mathbf{H}_{[1,i-1]}$ denote the vector space spanned by the column vectors $\mathbf{h}_1, \ldots, \mathbf{h}_{i-1}$, and $\tilde{\mathbf{y}}_i$ the result of the projection

$$\tilde{\mathbf{y}}_i = \mathbf{y}_i - \mathbf{H}_{[1,i-1]} \left(\mathbf{H}^{\dagger}_{[1,i-1]} \mathbf{H}_{[1,i-1]} \right)^{-1} \mathbf{H}^{\dagger}_{[1,i-1]} \mathbf{y}_i, \tag{2.3}$$

where † is the complex conjugate operator. Note that the matrix $\mathbf{H}^{\dagger}_{[1,i-1]} \mathbf{H}_{[1,i-1]}$ is invertible under the assumption that the vectors $\mathbf{h}_1, \ldots, \mathbf{h}_{i-1}$ are linearly independent. The above expression can be further simplified to obtain

$$\tilde{\mathbf{y}}_i = x_i \tilde{\mathbf{h}}_i + \tilde{\mathbf{n}}, \tag{2.4}$$

where $\tilde{\mathbf{h}}_i$ and $\tilde{\mathbf{n}}$ are the projection of $\mathbf{h}_i$ and $\mathbf{n}$ respectively onto the nullspace of $\mathbf{H}_{[1,i-1]}$. Using standard maximum ratio combining, the decision on the ith component of $\mathbf{x}$ is given by

$$\hat{x}_i = \frac{1}{\|\tilde{\mathbf{h}}_i\|^2} \tilde{\mathbf{h}}_i^{\dagger} \tilde{\mathbf{y}}_i = x_i + \frac{1}{\|\tilde{\mathbf{h}}_i\|^2} \tilde{\mathbf{h}}_i^{\dagger} \tilde{\mathbf{n}}. \tag{2.5}$$

The mean-square error (MSE) between x_i and $\hat{x}_i$ is given by

$$\text{MSE}_i = \frac{1}{\|\tilde{\mathbf{h}}_i\|^4} \tilde{\mathbf{h}}_i^\dagger \boldsymbol{\Gamma} \tilde{\mathbf{h}}_i, \tag{2.6}$$

where $\boldsymbol{\Gamma} = \mathsf{E}\left[\tilde{\mathbf{n}}\,\tilde{\mathbf{n}}^\dagger\right]$ is the covariance matrix of the noise vector $\tilde{\mathbf{n}}$. Let

$$\mathbf{A} = \mathbf{H}_{[1,i-1]} \left(\mathbf{H}_{[1,i-1]}^\dagger \mathbf{H}_{[1,i-1]}\right)^{-1} \mathbf{H}_{[1,i-1]}^\dagger, \tag{2.7}$$

then it is easy to see that $\mathbf{A} = \mathbf{A}^\dagger$, $\mathbf{A}\mathbf{A}^\dagger = \mathbf{A}$ and $\tilde{\mathbf{n}} = (\mathbf{I}_n - \mathbf{A})\mathbf{n}$. It follows that

$$\begin{aligned} \boldsymbol{\Gamma} &= (\mathbf{I}_n - \mathbf{A})\mathsf{E}\left[\mathbf{n}\,\mathbf{n}^\dagger\right](\mathbf{I}_n - \mathbf{A}^\dagger) \\ &= (\mathbf{I}_n - \mathbf{A}) N_0 \mathbf{I}_n (\mathbf{I}_n - \mathbf{A}^\dagger) \\ &= N_0(\mathbf{I}_n - \mathbf{A}). \end{aligned} \tag{2.8}$$

This simplifies the MSE for the ith channel to

$$\begin{aligned} \text{MSE}_i &= \frac{N_0}{\|\tilde{\mathbf{h}}_i\|^4} \tilde{\mathbf{h}}_i^\dagger (\mathbf{I}_n - \mathbf{A}) \tilde{\mathbf{h}}_i \\ &= \frac{N_0}{\|\tilde{\mathbf{h}}_i\|^4} \tilde{\mathbf{h}}_i^\dagger (\mathbf{I}_n - \mathbf{A})(\mathbf{I}_n - \mathbf{A}) \mathbf{h}_i \\ &= \frac{N_0}{\|\tilde{\mathbf{h}}_i\|^4} \tilde{\mathbf{h}}_i^\dagger (\mathbf{I}_n - \mathbf{A}) \mathbf{h}_i \\ &= \frac{N_0}{\|\tilde{\mathbf{h}}_i\|^4} \tilde{\mathbf{h}}_i^\dagger \tilde{\mathbf{h}}_i = \frac{N_0}{\|\tilde{\mathbf{h}}_i\|^2} \end{aligned} \tag{2.9}$$

Then, the resulting signal-to-noise ratio for the ith channel is

$$\gamma_i^{\mathrm{ZF}} = \frac{E_x}{N_0}\|\tilde{\mathbf{h}}_i\|^2 = \frac{\rho}{n}\|\tilde{\mathbf{h}}_i\|^2, \tag{2.10}$$

where E_x is the average constellation energy and $\|\tilde{\mathbf{h}}_i\|^2$ is a chi-square random variable with $2(n-i+1)$ degrees of freedom. Since the entries of the channel matrix $\mathbf{H}$ are zero-mean, unit-variance complex Gaussians, the mean of the chi-square variate is $n-i+1$.

Under the assumption of correct decisions being canceled out, interference suppression under the ZF criterion decomposes the MIMO channel of (2.1) into n parallel channels with SNRs $\gamma_1^{\mathrm{ZF}} \ldots \gamma_n^{\mathrm{ZF}}$. This is equivalent to a single-input single-output (SISO) periodically time-varying channel of period n whose gains are chi-square random variables with $2n, \ldots, 2$ degrees of freedom.

2.2.2 MMSE Interference Suppression

Again, assuming the last $n-i$ signal components have been correctly detected and canceled out, the interference from components $x_1, \ldots, x_{i-1}$ is suppressed by minimizing the mean-square error between x_i and $\hat{x}_i$. The decision on the ith component of $\mathbf{x}$ is given by $\hat{x}_i = \mathbf{w}_i^\dagger \mathbf{y}_i$, where $\mathbf{w}_i$ is an n-dimensional column vector chosen to minimize the mean-square error. The MMSE solution for $\mathbf{w}_i$ satisfies [31]

$$\mathbf{w}_i = \begin{cases} \frac{\mathbf{h}_1}{\frac{N_0}{E_x}\|\mathbf{h}_1\|^2}, & i = 1 \\ \frac{\left(\frac{N_0}{E_x}\mathbf{I}_n + \mathbf{H}_{[1,i-1]}\mathbf{H}_{[1,i-1]}^\dagger\right)^{-1}\mathbf{h}_i}{1+\mathbf{h}_i^\dagger\left(\frac{N_0}{E_x}\mathbf{I}_n + \mathbf{H}_{[1,i-1]}\mathbf{H}_{[1,i-1]}^\dagger\right)^{-1}\mathbf{h}_i}, & i = 2, \ldots, n \end{cases} \tag{2.11}$$

This yields a minimum mean-square error (MMSE) on the ith signal component given by

$$\begin{aligned} \text{MMSE}_i &= E_x\left(1-\mathbf{w}_i^{\dagger}\mathbf{h}_i\right) && (2.12)\\ &= \begin{cases} \frac{E_x}{1+\frac{E_x}{N_0}\|\mathbf{h}_1\|^2}, & i=1 \\ \frac{E_x}{1+\mathbf{h}_i^{\dagger}\left(\frac{N_0}{E_x}\mathbf{I}_n+\mathbf{H}_{[1,i-1]}\mathbf{H}_{[1,i-1]}^{\dagger}\right)^{-1}\mathbf{h}_i}, & i=2,\ldots,n \end{cases} && (2.13) \end{aligned}$$

and a corresponding signal-to-noise and interference ratio given by

$$\gamma_i^{\text{MMSE}} = \begin{cases} \frac{E_x}{N_0}\|\mathbf{h}_1\|^2, & i=1 \\ \mathbf{h}_i^{\dagger}\left(\frac{N_0}{E_x}\mathbf{I}_n+\mathbf{H}_{[1,i-1]}\mathbf{H}_{[1,i-1]}^{\dagger}\right)^{-1}\mathbf{h}_i, & i=2,\ldots,n \end{cases} \quad (2.14)$$

Note that γ_1^{MMSE} has a chi-square distribution with $2n$ degrees of freedom. However, $\gamma_i^{\text{MMSE}}(2 \le i \le n)$ has a distribution which is different from chi-square.

It can be shown that γ_i^{MMSE} is a random variable given by

$$\gamma_i^{\text{MMSE}} = \frac{E_x}{N_0}\|\mathbf{D}^{1/2}\mathbf{p}_i\|^2, \quad (2.15)$$

where

$$\mathbf{D}^{1/2} = \text{Diag}\left(\frac{1}{\sqrt{1+\lambda_1\frac{E_x}{N_0}}},\ldots,\frac{1}{\sqrt{1+\lambda_n\frac{E_x}{N_0}}}\right) \quad (2.16)$$

and $\mathbf{p}_i = \mathbf{U}^\dagger \mathbf{h}_i$ is the projection of the complex Gaussian vector $\mathbf{h}_i$ onto the space spanned by the columns of the unitary matrix $\mathbf{U}$. This is a direct consequence of the symmetric eigendecomposition of the Hermitian matrix

$$\mathbf{H}_{[1,i-1]}\mathbf{H}_{[1,i-1]}^\dagger = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\dagger, \tag{2.17}$$

where $\mathbf{\Lambda}$ is a diagonal matrix whose elements are the eigenvalues $\lambda_1, \ldots, \lambda_n$ of the Hermitian matrix in (2.17).

Observe that under the same assumptions as before, the MMSE interference suppression scheme also creates an equivalent time-varying channel with a periodic SNR variation given by $\gamma_1^{\text{MMSE}} \ldots \gamma_n^{\text{MMSE}}$. These SNR factors do not have a closed-form distribution. Figure 2.1 shows the empirical distribution of the 4 channels created in a 4×4 D-BLAST system with MMSE interference suppression. For comparison, on the same plot we show the chi-square distributions with different means corresponding to the 4 channels created in a D-BLAST system with ZF interference suppression.

2.3 Space-Time Layering for D-BLAST

2.3.1 Full-Overhead

In the following we describe a trellis coding technique for the layered space-time architecture of D-BLAST. We assume a single trellis code producing codewords of

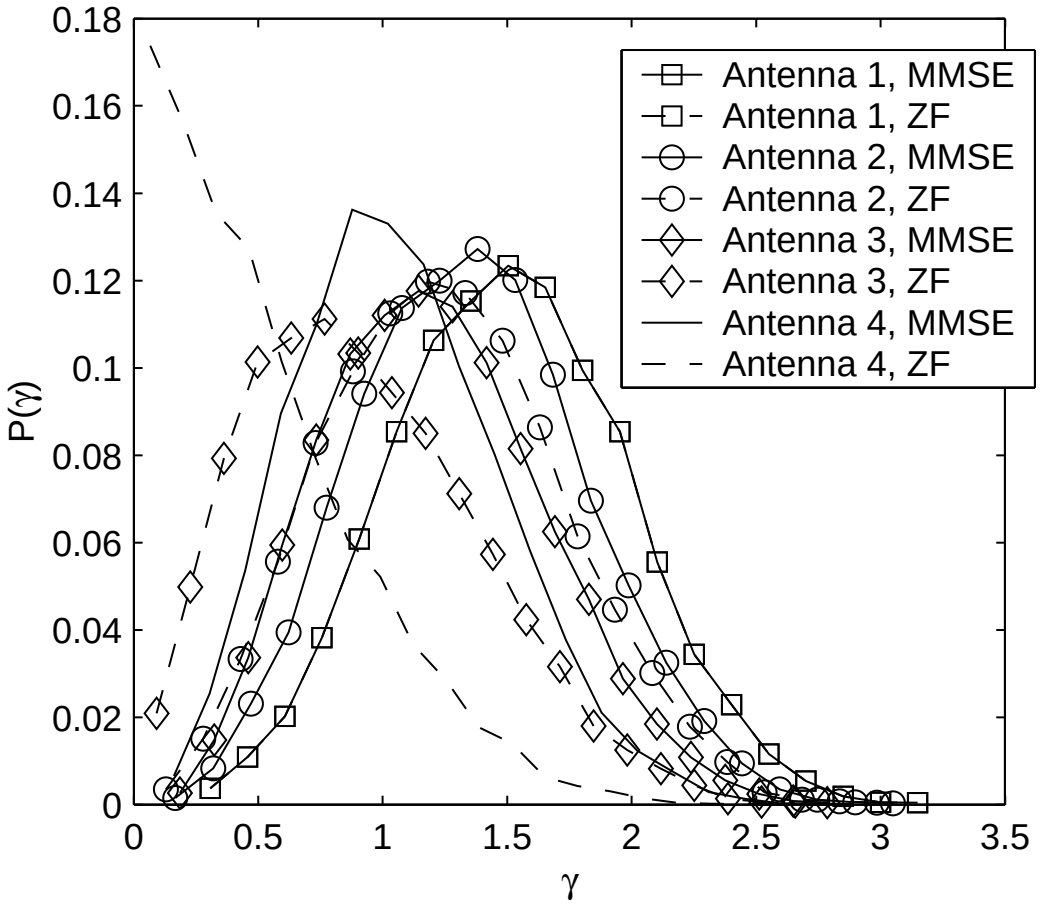


Figure 2.1: Distribution of periodic SNRs γ_i^{MMSE} and γ_i^{ZF} for 4×4 MMSE and ZF D - BLAST, $E_x/N_0 = 3$ dB, Rayleigh fading.

length N, including ν tail bits to drive the encoder back to state zero (where ν is the number of memory elements of the code). We further assume that N is a multiple of the number of antennas n. Then, the symbols in a codeword vector $\mathbf{c}$ are placed moving left-to-right in diagonal layers of width d, in order to form the $n \times N_f$ codeword matrix $\mathbf{X} = f(\mathbf{c})$, with $N_f = N/n + (n-1)d$. The mapping f sends the codeword vector $\mathbf{c}$ to a $n \times N_f$ complex-valued matrix $\mathbf{X}$ whose (i,t)th entry is equal to $x_i(t)$, i.e. $\mathbf{X}$ is the baseband version of the codeword $\mathbf{c}$ as transmitted across the channel. The mapping f is defined by

$$\mathbf{X}(i,j) = \begin{cases} c_k & \text{if } 1 \leq k \leq N \\ 0 & \text{otherwise} \end{cases} \tag{2.18}$$

where

$$k = n[j - 1 - (i-1)d] + i, \tag{2.19}$$

for $1 \leq i \leq n$ and $1 \leq j \leq N_f$.

Figure 2.2(a) illustrates the symbol layering for a codeword of length 16 in a 2×2 system with diagonal layers of width $d = 4$. Each square represents a symbol transmitted from a single antenna at a single symbol time. The numbers in the squares represent the processing order of the Viterbi decoder. The empty squares represent overhead symbols where nothing is transmitted.

Notice that this scheme has an inherent rate loss of approximately $(n-1)d/N_f$ because of the lower and upper triangles of overhead (zero) symbols in the codeword matrix. This ignores the small rate loss due to the tail bits used for trellis termination. Assuming a trellis code of rate R and a modulation mapping each M_c-tuple of binary symbols into a complex-valued constellation point, the spectral efficiency of the above layering scheme is

$$\eta = nRM_c - \frac{n(n-1)d}{N_f} RM_c. \tag{2.20}$$

In this formulation efficiency is taken to mean the number of information bits per vector channel use.

In Fig. 2.2(a) we have used a codeword matrix that is conceptually appealing. The order in which the columns of the codeword matrix are transmitted is, in fact, arbitrary so that we could re-arrange them as in Fig. 2.2(c) to avoid any delay in processing. In the sequel we will refer to this scheme as the *full-overhead* system in order to differentiate it from a variation of the scheme that has only half of the overhead penalty. The new *reduced-overhead* scheme is described next.

2.3.2 Reduced-Overhead

The reduced-overhead scheme is a variation of the full-overhead system that cuts the overhead penalty in half. There are many possible symbol arrangements which

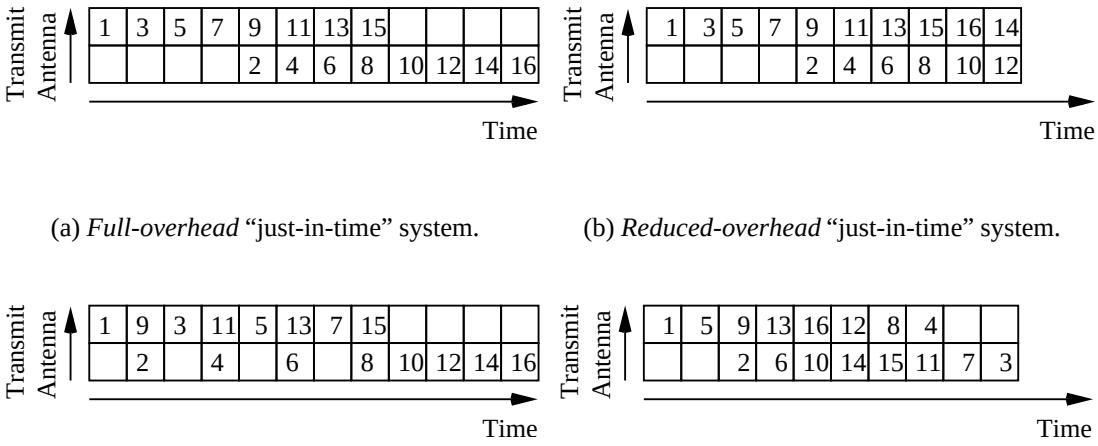
yield a reduced-overhead system. The simplest one is shown in Fig. 2.2(b), where the only difference from the full-overhead system is the placement of symbols $16, 14$ above $10, 12$. All four of these symbols are intended to be decoded using suppression. Thus, in this particular arrangement we do not require empty squares of pilot symbols at the end of the block. Then, the overhead penalty is cut in half and the spectral efficiency of the reduced-overhead system is increased to

$$\eta = nRM_c - \frac{n(n-1)d}{2N_f} RM_c. \tag{2.21}$$

A slightly different approach is illustrated in Fig. 2.2(d), which preserves the diagonal structure of the block, but has half the diagonal width. Observe that in this scheme we arrange the symbols moving from both ends of the frame toward the middle. While the first version is conceptually more transparent, this approach has the appeal of a diagonal block structure. This is in fact the version implemented in all our simulations. However, we believe that the difference in performance between the two reduced-overhead schemes should be negligible.

2.4 Finite-Traceback Viterbi Decoding

In this section we present both an optimal "just-in-time" and a "not just-in-time" Viterbi decoding method for the layered space-time systems discussed in the previous section. Note that we make a Gaussian approximation on the MMSE filter output,



(a) *Full-overhead* "just-in-time" system.

(b) *Reduced-overhead* "just-in-time" system.

(c) Reordered *full-overhead* "just-in-time" system.

(d) *Reduced-overhead* "just-in-time" system.

Figure 2.2: Symbol placement for full-overhead and reduced-overhead "just-in-time" systems, $L_D^* = 7$, blocklength 16 data symbols.

so that the Viterbi algorithm can be employed to decode the layered systems, provided that the appropriate variance is used in the metric calculation. For ease of exposition we consider 2×2 systems, however the decoding method extends to larger $n \times n$ systems in a straight forward manner.

2.4.1 Optimal "Just-in-Time" Traceback

First, consider the full-overhead system in Fig. 2.2(a). The diagonal layering in this example is well matched for finite-traceback Viterbi decoding [46] with traceback depth

$L_D^* = 7$. Linear processing as described in Section 2.2 is used to detect the received vectors. Thus, each odd-numbered symbol is detected by canceling the even-numbered symbol below it, while each even-numbered symbol is detected by suppressing the symbol above it.

Observe that in this diagonal layering there are four pilot symbols at the beginning of the block. With this arrangement, symbols $1, 3, 5, 7$ are detected by simply canceling the known pilots below them, while symbols $2, 4, 6$ are detected by suppressing the interference from the symbols above them. Thus, the first seven symbols can be detected without any Viterbi decoding. This is enough to permit the first traceback. Thereafter, the bottom even-numbered symbols are decoded "just-in-time" to provide the cancellation required for the next Viterbi traceback. For example, after symbol 8 is detected by suppression, Viterbi traceback is performed to decode symbol 2. This symbol is decoded "just-in-time" to be canceled out from the received vector so that symbol 9 can be detected. This process continues until all the symbols in the block have been decoded.

With this decoding method, the basic idea is to choose a traceback depth L_D^* (perhaps chosen according to [6] or through empirical study) and a diagonal layering which together permit the decoding of a symbol "just-in-time" to be subtracted out from the received vectors. In general, in order to decode a symbol "just-in-time" the step-size of the diagonal layering must be

$$d = \frac{L_D^* + 1}{n}. \tag{2.22}$$

Similar "just-in-time" decoding is supported by the reduced-overhead system in Fig. 2.2(b). The only difference is in the detection of symbols $10, 12, 14, 16$, which must be received by suppressing the interfering symbols above or below them (in the full-overhead system these symbols might also be detected by suppression; however, one can do better by canceling the known pilot symbols instead).

2.4.2 "Not Just-in-Time" Traceback

For a given traceback depth L_D, one might consider lowering the overhead penalty by selecting a layer width d smaller than what is required for "just-in-time" decoding. For illustration, consider using the traceback depth of 7 with only two initial pilots (zero symbols) as illustrated in Fig. 2.3. With only two initial pilots, the first four space-time symbols are detected in the same manner as before. However, the odd-numbered space-time symbols 5 and 7 have the undecoded symbols 2 and 4 below them, therefore they must be detected by suppressing the symbols 2 and 4 respectively. In general, the first Viterbi traceback requires the last $L_D - nd + 1$ symbols to be detected by suppression only. After a traceback, the decoded even-numbered symbols must be canceled from an earlier received vector, which implies that the last $L_D - nd + 1$ stages of the trellis must be updated for future tracebacks.

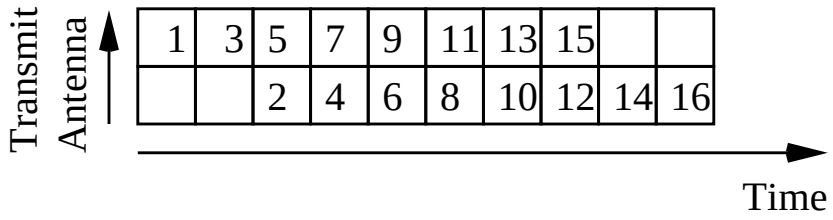


Figure 2.3: Symbol placement for "not just-in-time" traceback, $L_D^* = 7$, blocklength 16 data symbols.

While more general, this scheme suffers from the added complexity of updating the last $L_D - nd + 1$ stages of the trellis whenever a decoded symbol needs to be canceled out. In Section 2.6 we take a closer look at this decoding method and show that the added complexity is not justified by the gain in performance. In fact, our results show that the "just-in-time" decoding method provides the best performance (measured as the SNR gap to outage probability) for a fixed traceback depth L_D.

2.5 Trellis Codes as Universal Space-Time Codes

2.5.1 Excess MI versus SNR Gap

Mutual information (MI) identifies the fundamental information-carrying potential of a channel for a specific input distribution. Root and Variaya [36] proved the existence of codes that can communicate reliably over any linear Gaussian channel for which the

MI exceeds the information rate of the code. A code has *universal* property if it has consistently good proximity to capacity (to the extent its blocklength and decoding complexity permit) over a class of channels. Thus a universal code should provide performance that is consistent in terms of required excess mutual information. The excess MI is defined as the capacity margin between the operational channel MI and the information transmission rate.

The common way to plot bit-error rate (BER) performance is versus channel signal-to-noise ratio (SNR). Since capacity on the additive white Gaussian noise (AWGN) channel is a monotonic (and almost linear) function of SNR in dB, $C_{\text{AWGN}} = \log_2(1 + \text{SNR})$, this representation is essentially equivalent to plotting BER against MI. However, when assessing the performance of a code over a variety of linear Gaussian channels, considering SNR performance or SNR gap is problematic because the monotonic relationship between SNR in dB and MI is different for different channel eigenvalue skews.

To better understand the previous statement, consider a 2×2 linear Gaussian vector channel $\mathbf{H}$ and its per antenna capacity [40]

$$\begin{aligned} C(\mathbf{H}) &= \frac{1}{2}\log_2 \det\left(\mathbf{I}_2 + \rho\mathbf{H}\mathbf{H}^\dagger\right) \\ &= \frac{1}{2}\log_2 \left(1+\rho\lambda_1\right)\left(1+\rho\lambda_2\right) \\ &= \frac{1}{2}\log_2 \left(1+\rho\lambda_1\right)\left(1+\rho\kappa\lambda_1\right), \end{aligned} \tag{2.23}$$

where λ_1 and λ_2 are the eigenvalues of $\mathbf{H}\mathbf{H}^\dagger$, $\kappa = \lambda_2/\lambda_1$ is the eigenvalue skew and ρ is the SNR per transmit symbol. For a fixed transmission rate R, the SNR gap in dB $\Delta\text{SNR}_{\text{dB}}$ is defined as the difference between the ρ_0 (in dB) required to achieve the desired BER and the ρ (in dB) at which the channel capacity in (2.23) is equal to R bits per transmit antenna.

Without loss of generality, let $\lambda_1 = 1$ and find ρ at which the channel capacity is equal to R bits per transmit antenna:

$$2^{2R} = \kappa\rho^2 + (\kappa + 1)\rho + 1, \tag{2.24}$$

which yields

$$\rho = \frac{-(\kappa + 1) + \sqrt{(\kappa - 1)^2 + 4^{R+1}}}{2\kappa}. \tag{2.25}$$

Then, the SNR gap in dB is given by

$$\Delta\text{SNR}_{\text{dB}} = 10\log_{10}\rho_0 - 10\log_{10}\rho, \tag{2.26}$$

from which the SNR ρ_0 can be expressed as

$$\rho_0 = 10^{0.1(\Delta\text{SNR}_{\text{dB}} + 10\log_{10}\rho)} = \rho 10^{0.1\Delta\text{SNR}_{\text{dB}}}. \tag{2.27}$$

The excess MI per transmit antenna ΔMI is given by the difference between the channel capacity in (2.23) at the SNR ρ_0 and the information rate R. It can be written in terms of the SNR gap in dB as follows

$$\begin{aligned} \Delta\text{MI} &= \frac{1}{2}\log_2(1+\rho_0)(1+\kappa\rho_0) - R \qquad (2.28) \\ &= \frac{1}{2}\log_2\left(1+\rho 10^{0.1\Delta\text{SNR}_{\text{dB}}}\right)\left(1+\kappa\rho 10^{0.1\Delta\text{SNR}_{\text{dB}}}\right) - R, \qquad (2.29) \end{aligned}$$

where ρ is given in (2.25) above.

Figure 2.4 illustrates the excess MI per antenna as a function of the SNR gap in dB for $R = 1$ bit/transmit antenna and different eigenvalue skews κ. Note that the excess MI curves are approximately linear functions of the SNR gap in dB, however the slope depends on the eigenvalue spread (eigenskew) of the channel. In other words, a constant level of excess MI is achieved by differing excess SNR levels (depending on the eigenskew of the channel). The MI available in the channel is the absolute measure for performance, while an excess SNR measure depends both on the MI level and the specific channel realization (eigenskew).

2.5.2 Universal Code Design for Periodic Erasures

In Section 2.2 we saw that the layered architecture with linear processing at the receiver creates a periodic fading channel, therefore we require a trellis code that is effective on such periodic variations in SNR. A special case of the results of Root and

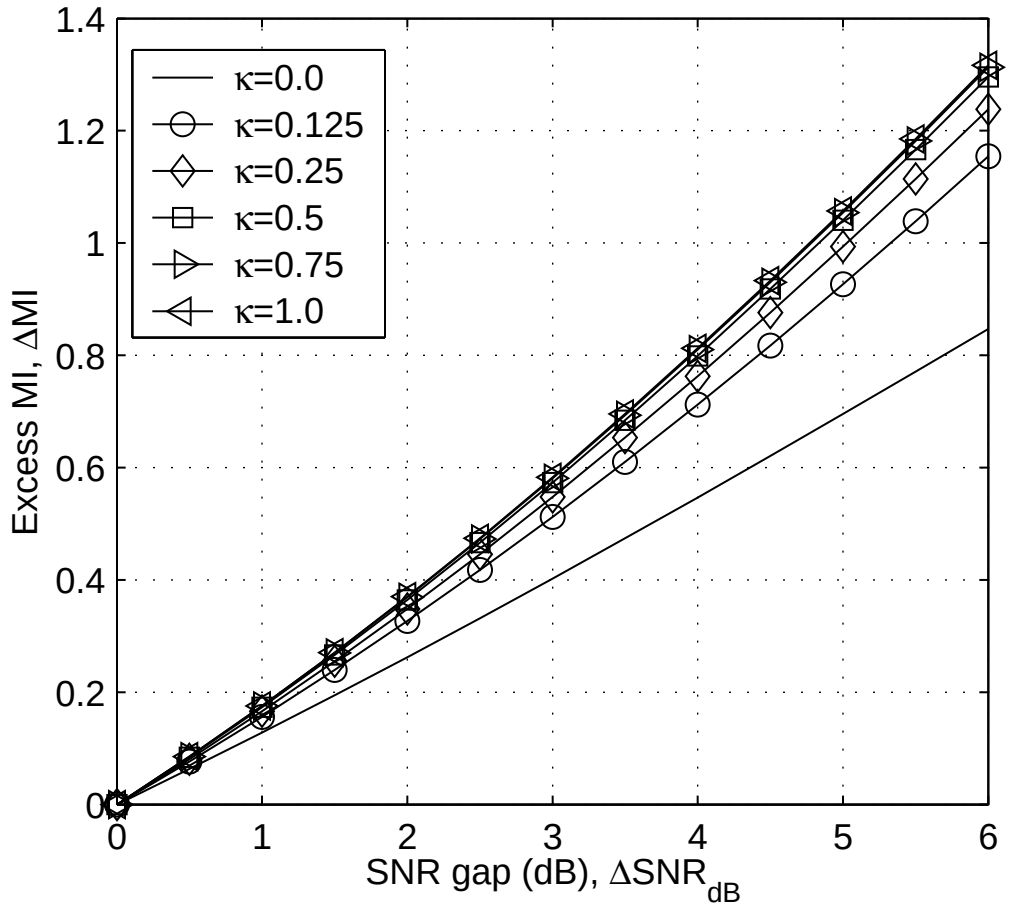


Figure 2.4: Excess MI per antenna versus SNR gap for 2×2 matrix channel, $R = 1$ bit per transmit antenna, for eigenvalue skews (top to bottom) $\kappa = \{1, 0.75, 0.5, 0.25, 0.125, 0\}$.

Varaiya [36] indicates that a single code can guarantee a rate R over every periodic Gaussian channel that has a mutual information greater than the rate R, where

$$\text{MI} = \frac{1}{n}\sum_{i=1}^{n} \log_2\left(1 + |\tilde{a}_i|^2 \frac{E_x}{N_0}\right), \tag{2.30}$$

and $\tilde{a}_i$ are the periodic scale factors. One implication of this result is that good error performance over one particular channel does not have to come at the expense of significant performance degradation over others.

The design of trellis codes for channels that are characterized by additive white Gaussian noise with a distinct periodic variation in signal-to-noise ratio is investigated in [48], [47], [14], [23], and [32]. The case of periodic Rayleigh fading is treated in [47], [23], [32]. In [48] Wesel *et. al* construct universal trellis codes for periodic erasure channels. A family of trellis codes for periods 2 - 5 are designed to provide robust performance over all periodic erasure patterns for which the number of unerased coded bits per period is at least equal to the number of information bits per period. A brief review of the design approach for these trellis codes is presented next.

Let the n-element vector $\tilde{\mathbf{a}} = [\tilde{a}_1, \ldots, \tilde{a}_n]$ contain the periodic scale factors with $\tilde{\mathbf{a}}^2 = [|\tilde{a}_1|^2, \ldots, |\tilde{a}_n|^2]$. Also let the normalized symbol-wise squared Euclidean distance between two constellation points be $d_i^2(x \rightarrow \hat{x}) = |\hat{x}_i - x_i|^2/E_x$, where x_i and $\hat{x}_i$ are the correct and incorrect constellation points associated with the ith symbols of a trellis error event $x \rightarrow \hat{x}$. The periodic erasure of symbols scales distances with the

same index modulo n by the same binary scale factor $\tilde{a}_i$. Define the periodic squared distance $\tilde{d}_i^2(x \rightarrow \hat{x})$ for $i \in \{1, 2, \ldots, n\}$ as the sum of the square of the distances that are scaled by the same factor $\tilde{a}_i$ for a given error event: $\tilde{d}_i^2(x \rightarrow \hat{x}) = \sum_{m=0}^{\infty} d_{i+mn}^2$. The n values $\tilde{d}_1^2, \ldots, \tilde{d}_n^2$ form the periodic distance vector

$$\tilde{\mathbf{d}}^2 = [\tilde{d}_1^2, \ldots, \tilde{d}_n^2]. \tag{2.31}$$

The minimum distance of valid error events under a periodic erasure pattern is referred to as the squared residual Euclidean distance (RED) for a specified periodic attenuation vector $\tilde{\mathbf{a}}$

$$\text{RED}(\tilde{\mathbf{a}}) = \min_{\tilde{\mathbf{d}}^2} \langle \tilde{\mathbf{a}}^2, \tilde{\mathbf{d}}^2 \rangle, \tag{2.32}$$

where the expression $\langle \tilde{\mathbf{a}}^2, \tilde{\mathbf{d}}^2 \rangle$ represents an inner product. Trellis code design for all possible erasure patterns is a multi-criterion problem since we have to minimize bit-error rate (BER) (essentially maximize RED) at all erasure patterns simultaneously.

The primary approach described in [48] is to seek to maintain the same required excess mutual information for all erasure patterns, which is motivated by the desire of having similar error performance over all erasure patterns with the same MI. The search for trellis codes that maintain a similar level of excess mutual information over all erasure patterns produces a design criterion based on maximizing J_{MI},

$$J_{\text{MI}} = \sum_{j} (n - q_j/n) \log_2(\text{RED}^2(\tilde{\mathbf{a}}_j)), \tag{2.33}$$

where q_j is the number of elements of the jth erasure pattern $\tilde{\mathbf{a}}_j$ equal to zero.

The codes in [48] were designed to be universal only for periodic erasures channels. However, we show that code $\#1$ designed in [48] using J_{MI} as the objective function is a universal trellis code for general periodic fading and hence D-BLAST. First, consider the performance of the 64-state, rate-$1/3$, 8-PSK code $\#1$ over periodic erasures. For a BER of 10^{-5}, this code requires an excess mutual information of 0.64 bits on the channel where every other symbol is an erasure ($\tilde{\mathbf{a}} = [0, 1]$), and an excess mutual information of 0.87 bits on the channel with no erasures ($\tilde{\mathbf{a}} = [1, 1]$). Figure 2.7 shows an extension of the results presented in [48], by considering the excess mutual information required for BER $= 10^{-5}$ on a periodic channel $\tilde{\mathbf{a}} = [\tilde{a}_1, \tilde{a}_2]$ when the ratio of squared scale factors $\alpha = |\tilde{a}_2|^2/|\tilde{a}_1|^2$ varies from 0 to 1. In this plot the curve labeled Period-2 TFB is generated using a transfer function bound to obtain the SNR necessary to achieve BER $= 10^{-5}$ for values of α between 0 and 1. Observe that as α increases from 0 to 1 the excess mutual information is monotonically increasing from 0.64 to 0.87, which correspond to the excess MI figures on the two erasure channels. For comparison, the best Euclidean distance 64-state trellis code for transmission of one bit per symbol (rate-$1/2$ maximum Hamming distance convolutional code used with Gray-labeled QPSK) requires an SNR of 4.3 dB to achieve BER $= 10^{-5}$ in additive white

Gaussian noise. At this SNR, the capacity of the AWGN channel is 1.88 bits/symbol and thus this code requires an excess MI of 0.88 bits. Therefore, our universal code has less excess MI on every period-2 channel than the standard amount of excess MI required by a 64-state code at this rate for the AWGN channel.

Practical limitations on the number of memory elements and the constellation size make the design of universal trellis codes for periodic fading channels with large periods a very challenging task. This is why the family of codes found in [48] are for periods no larger than five. However, there exists a class of trellis codes for period-8 designed in [47]. These codes are not universal for all period-8 channels, but they are still robust codes and reasonable candidates for larger 8×8 full-overhead systems or 4×4 reduced-overhead systems.

2.5.3 Universal Performance of Trellis Coded D-BLAST Systems

Next consider the performance of an MMSE full-overhead D-BLAST system using code $\#1$ over a set of linear Gaussian channels obtained by sampling the continuum of 2×2 channels. For that it is sufficient to consider channels of the form

$$\mathbf{H} = \sqrt{\lambda_1} \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{\kappa} \end{bmatrix} \begin{bmatrix} \cos(\phi) & \sin(\phi)e^{j\theta} \\ -\sin(\phi)e^{-j\theta} & \cos(\phi) \end{bmatrix} \tag{2.34}$$

where λ_1, λ_2 are the eigenvalues of the Hermitian matrix $\mathbf{H}\mathbf{H}^\dagger$ with $\lambda_2 \leq \lambda_1$, $\kappa = \lambda_2/\lambda_1$ is the eigenvalue skew, $\phi \in [0, 2\pi]$, and $\theta \in [0, 2\pi]$. We sample the above

matrix via the parameters $\lambda_1 = 1$, $\kappa = \{0, 0.125, 0.25, 0.5, 0.75, 1\}$, $\phi = \{0, \pi/3, \pi/2\}$, and $\theta = \pi/4$. The periodic SNRs γ_1^{MMSE} and γ_2^{MMSE} turn out to be independent of θ, therefore we can choose θ arbitrarily. This is easily shown given the above formulation for the channel matrix $\mathbf{H}$. The periodic SNRs take the following form:

$$\gamma_1^{\text{MMSE}} = \frac{E_x}{N_0}(\cos^2\phi + \kappa \sin^2\phi) \tag{2.35}$$

$$\gamma_2^{\text{MMSE}} = \frac{\sin^2\phi + \kappa\cos^2\phi + \frac{E_x}{N_0}\kappa}{\kappa\sin^2\phi + \cos^2\phi + \left(\frac{E_x}{N_0}\right)^{-1}}. \tag{2.36}$$

One immediate observation is that for $\kappa = 1$, the above equations yield $\gamma_1^{\text{MMSE}} = \gamma_2^{\text{MMSE}} = E_x/N_0$. Let α be the ratio of periodic SNRs induced by the MMSE processing in D-BLAST:

$$\alpha = \gamma_2^{\text{MMSE}}/\gamma_1^{\text{MMSE}}. \tag{2.37}$$

Assuming $\gamma_1^{\text{MMSE}} > \gamma_2^{\text{MMSE}}$, it follows that for unitary channels, $\alpha = 1$ regardless of the angle ϕ and hence the particular sampling of the channel. A plot of the ratio of periodic SNRs α as a function of the angle ϕ is shown in Fig. 2.5 for the different eigenvalue skews considered and under fixed mutual information, MI $= 1.5$ bits/antenna. Observe that when the channel is singular ($\kappa = 0$) and ϕ is a multiple of $\pi/2$, the SNR ratio α is zero (which corresponds to a periodic erasure channel); however, α is one when ϕ is approximately a multiple of $\pi/3$. The same results are shown in Fig. 2.6, where

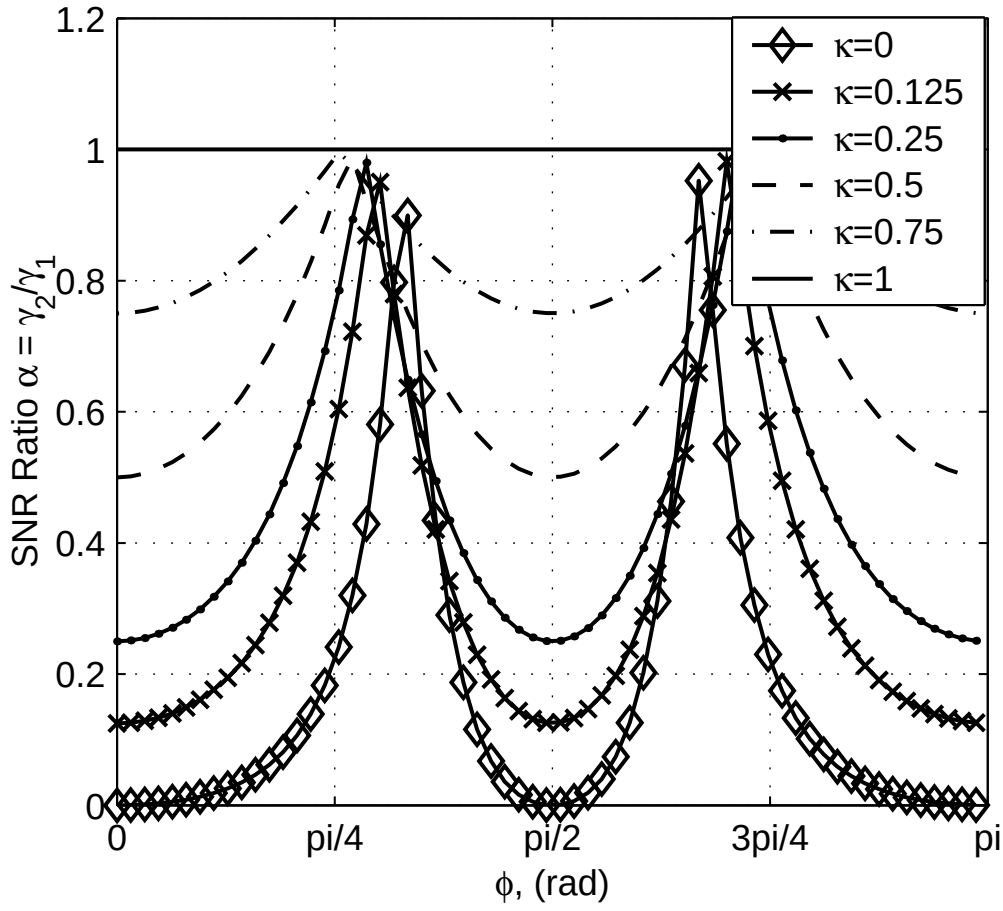


Figure 2.5: The ratio of periodic SNRs α versus the angle ϕ for eigenvalue skews $\kappa = \{1, 0.75, 0.5, 0.25, 0.125, 0\}$ under constant mutual information (MI=1.5 bits/antenna).

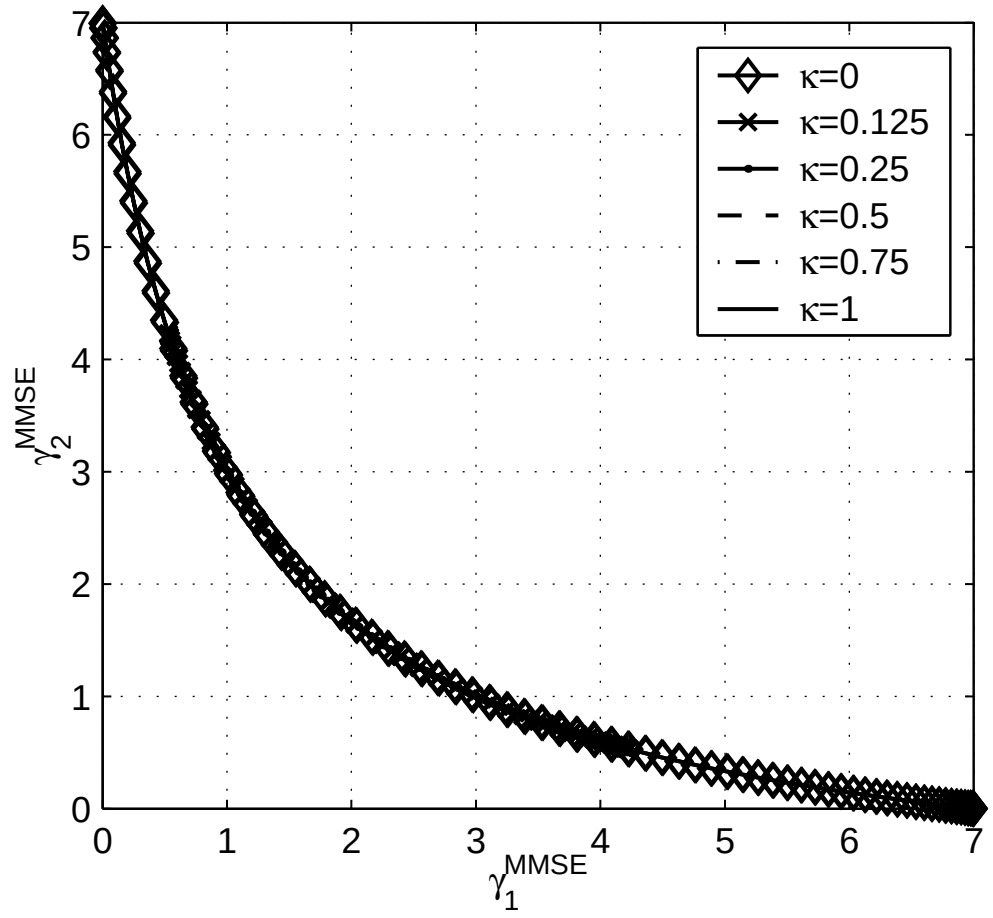


Figure 2.6: γ_2^{MMSE} versus γ_1^{MMSE} for eigenvalue skews $\kappa = \{1, 0.75, 0.5, 0.25, 0.125, 0\}$ and underlying angle $\phi \in (0, \pi/2)$ under constant mutual information (MI=1.5 bits/antenna).

instead we plot the achievable γ_1^{MMSE} - γ_2^{MMSE} pairs. Since the mutual information is constant, the different channel samplings yield the same achievable region, although the individual pairs in the achievable region are different under the same value of the angle ϕ.

Performance on each channel in this sampling is measured by the mutual information in excess of the transmission rate η required to achieve a BER of 10^{-5}. Figure 2.7 plots this required excess mutual information per transmit antenna as individual points against the corresponding ratio of periodic SNRs α. The points labeled with circles correspond to a blocklength of 1016 symbols and an overall rate $\eta = 1.9538$, while the points labeled with squares correspond to a blocklength of 127 symbols and an overall rate $\eta = 1.6821$. Observe that with the larger blocklength where the overhead penalty is negligible, the excess MI points obtained on the individual channels in the sampling set are very close to the excess MI curve for the periodic channel. In the case of the smaller blocklength, the loss in excess MI is relatively constant for all channels in the set and around 0.15 bits per antenna.

For both blocklengths, there is a single point that has noticeably poor performance. Both points correspond to $\kappa = 0$ and $\phi = \pi/3$. The poor performance is due to error propagation, as shown by the solid points which simulate decoding on this channel with the cancellation performed using the correct symbols. Note that simulations for $(\kappa = 0, \phi = 0)$ and $(\kappa = 0, \phi = \pi/2)$ showed no performance degradation due to error propagation. For the blocklength 1016 case the ratio of periodic SNRs α for

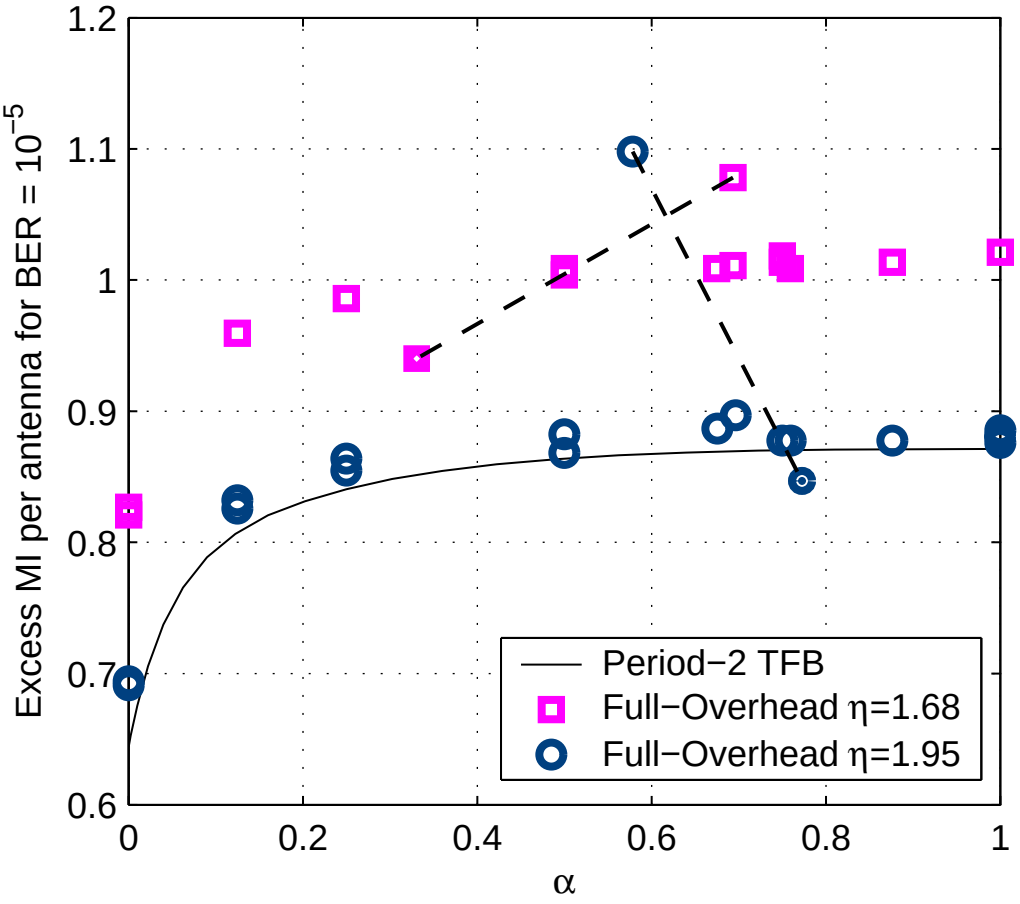


Figure 2.7: Excess MI per antenna required by rate-$1/3$, 8-PSK trellis code #1 to achieve BER $= 10^{-5}$ versus the ratio of periodic SNRs $\alpha = \gamma_2^{\text{MMSE}}/\gamma_1^{\text{MMSE}}$.

$(\kappa = 0, \phi = \pi/3)$ changes when error propagation is removed because α depends on the operating SNR.

It is not surprising that we observe error propagation effects in a decision directed cancellation algorithm like D-BLAST. In fact, perhaps it is surprising that error propagation is only observed under the particular channel condition where $\kappa = 0$ and ϕ is particularly unfavorable. Such a channel is essentially the scalar channel $y = \cos(\phi)x_1 + \sin(\phi)x_2 + n$. On such a channel, when $\cos(\phi) = 0$ or $\sin(\phi) = 0$, we have a periodic erasure channel which our trellis code handles quite well. However, this scalar channel with $\cos(\phi)$ and $\sin(\phi)$ of similar magnitude presents the toughest challenge for the suppression algorithm, often leading to error propagation. From a practical point of view, the channels with this degenerate singular value structure ($\kappa = 0$) are line-of-sight channels and "keyhole" channels [4]. Under Rayleigh statistics, the $\kappa \approx 0$ channels occur frequently enough, thus error propagation can be a problem. One possible remedy is to use a precoder similar to the Tomlinson-Harashima precoder [42], [19], since the processing of BLAST can be viewed as a generalized decision feedback equalizer (GDFE) [18]. However, the use of a precoder requires knowledge of the channel statistics at the transmitter site.

The above results show that code $\#1$ is a universal space-time code for D-BLAST, with consistently performance over the set of 2×2 matrix channels, with the exception of the channel $(\kappa = 0, \phi = \pi/3)$. For large blocklengths and ignoring error propagation, this code requires an excess MI of around 0.7 bits per antenna under the most

favorable channel and around 0.9 bits per antenna under the least favorable channel. This is in agreement with the results in [24] which show that in a more general setting, the performance of universal trellis codes obtained by an exhaustive minimax search on a 2×2 compound channel falls within the same range of excess mutual information requirement. One immediate implication of the universal property over the set of linear Gaussian channels is a good proximity to capacity on any fading channel, regardless of the channel fading statistics. Thus, this code may be deployed for any 2×2 transmission scenario, not just Rayleigh fading.

2.6 Performance Results in Rayleigh Fading

In this section we present performance results for the coded layered space-time architecture of D-BLAST. As performance measure we consider both the frame-error rate (FER) and the bit-error rate (BER) under the assumption of an ideal Rayleigh channel with perfect CSI at the receiver and a burst-mode communication scenario. The results are obtained through Monte Carlo simulation. The FER is averaged over enough channel realizations so that 1000 erroneous frames are accumulated in each simulation.

In order to compare the performance results to channel capacity, we also provide the theoretical outage probability curves. In the quasistatic case, it is assumed that the channel matrix $\mathbf{H}$ is chosen randomly at the beginning of a frame and held fixed for

the entire duration of the frame. The Shannon capacity of the channel is zero. No matter how small we make the rate η at which we try to communicate at, there is a non-zero probability that the channel realization is incapable of supporting that rate regardless of how long we make the code length. In this case a channel outage is said to have occurred and the channel is considered to be in the OUT state. Thus, under the quasistatic channel model and in the limit of large block lengths, the best possible achievable FER of the MIMO channel is given by the outage probability defined by

$$P_{\text{out}} = \mathsf{P}(C(\mathbf{H}) \leq \eta). \tag{2.38}$$

In the following examples we provide simulation results for several layered space-time coded systems under the optimal MMSE criterion, and whenever possible we make comparisons with the best performance of previously reported space-time systems. In general, the ZF criterion yields a performance that is 2 - 3 dB worse than MMSE. Figure 2.8 illustrates the FER performance of code $\#1$ in [48] for period-2 erasures in a full-overhead system under the ZF criterion together with the outage probability curves at the overall rate which takes the overhead penalty into account and at the rate which neglects the overhead penalty. On the same plot we show the FER performance of code $\#1$ over an equivalent periodic fading channel whose gains are given by the chi-squared random variables in (2.10) and observe that it matches closely the performance of the ZF full-overhead system.

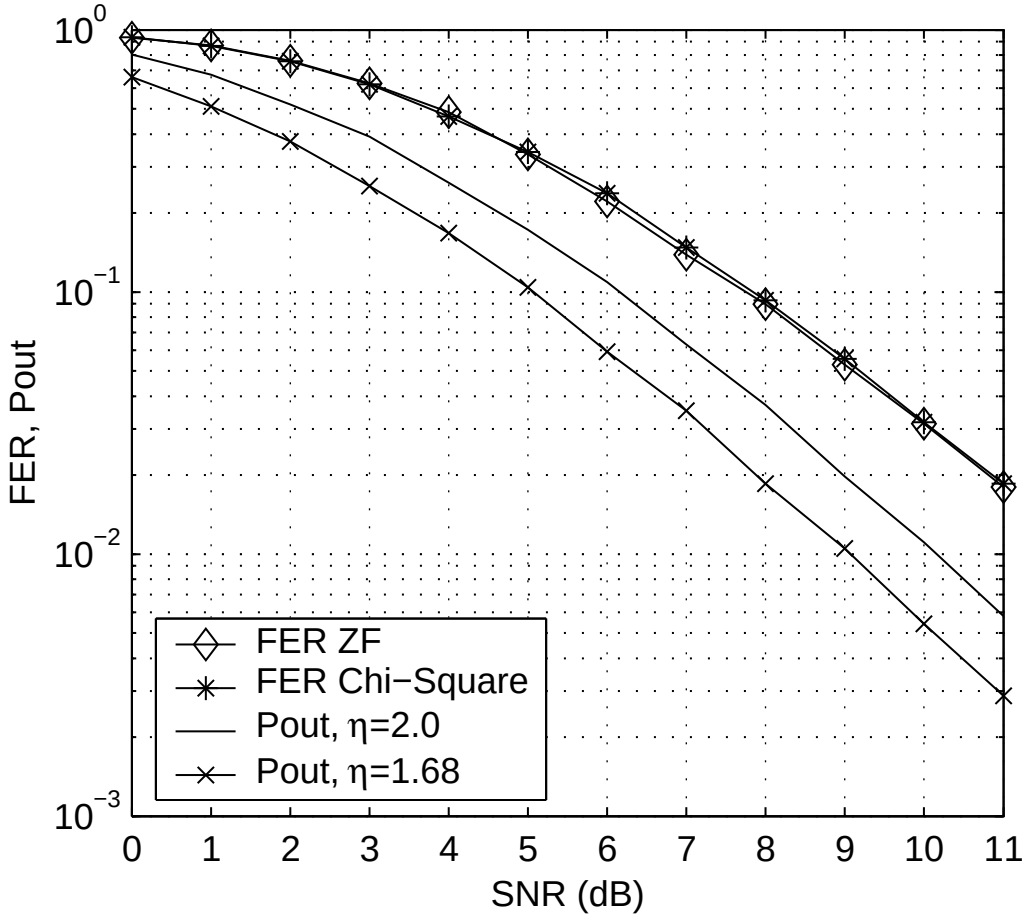


Figure 2.8: Outage probability and FER for 2×2 ZF full-overhead system coded with rate-1/3, 8-PSK trellis code, blocklength 127 information bits per transmit antenna.

Example 1: Here we consider a 2×2 MMSE full-overhead layered space-time coded system in conjunction with code $\#1$ in [48]. This is a rate-$1/3$, 64-state, 8-PSK trellis code with octal generators $(G_1, G_2, G_3) = (173, 62, 115)$. The 8-PSK labeling is the same as in [48], i.e. $0, 2, 3, 1, 5, 7, 6, 4$ in octal going around the circle. The underlying symbol layering is for a codeword (with trellis termination) of length 254 information bits (127 information bits per transmit antenna). The traceback depth is $L_D^* = 41$, which determines the "just-in-time" step-size of the diagonal layering to be $d = 21$. The spectral efficiency of this system is $\eta = 1.6821$ bits/s/Hz.

Figure 2.9 provides the FER and BER performance with the horizontal axis showing the average SNR per receive antenna. For comparison we also provide the FER performance of the 64-state space-time trellis code with similar complexity and same blocklength designed and simulated by Tarokh in [39]. The same or better performance is achieved by the space-time parallel concatenated full-diversity turbo code in [26]. Even though our code was designed to be robust on all periodic erasure channels, and not specifically matched to the Rayleigh statistics, our performance in Rayleigh fading is essentially the same as Tarokh's. Furthermore, the Tarokh code performs poorly on channels where an eigenvalue is close to zero, because for such a channel the code effectively has no redundancy.

The previous paragraph ignores the overhead penalty required by D-BLAST to create the periodic channel. For demonstrating the performance of the trellis code, it is reasonable to neglect the overhead and look at how closely each code comes to the

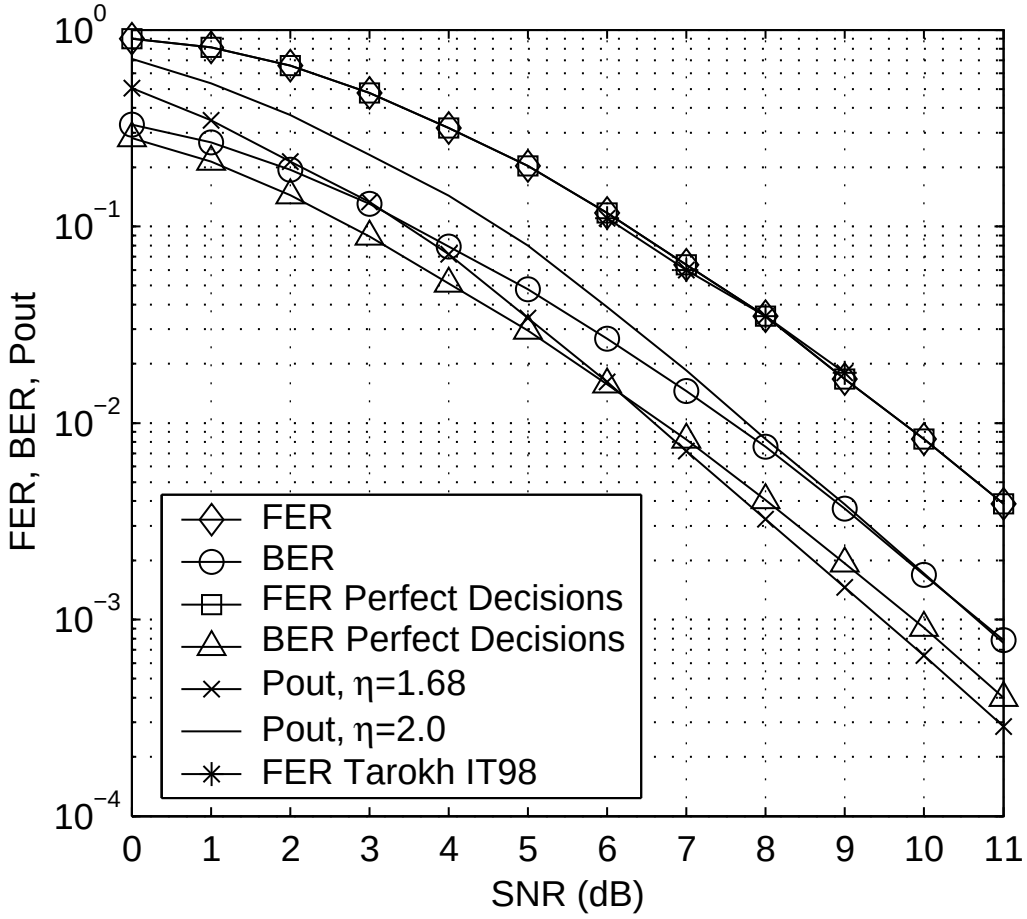


Figure 2.9: Outage probability, FER, and BER for 2×2 MMSE full-overhead system coded with rate-1/3, 8-PSK trellis code, blocklength 127 information bits per transmit antenna.

theoretical outage curve that applies to the code itself (the rate $\eta = 2$ outage curve in Fig. 2.9 for both Tarokh's code and the universal code on the periodic channel created by D-BLAST). However, to analyze the overall system performance, the overhead penalty must be considered. At 10% outage the performance of our universal code is within 1.8 dB from the theoretical outage curve neglecting overhead and within 3 dB from the theoretical outage curve which takes into account the overhead penalty required by D-BLAST. So the code loses 1.8 dB from ideal performance and the overhead penalty incurs an additional 1.2 dB of loss.

Figure 2.9 also provides the performance of the 2×2 system when perfect decisions are used in the cancellation step. As seen from the plot, the error propagation due to non-perfect decisions has no effect on the frame-error rate performance. This fact is also observed in [10], where the authors conclude that the PSP-based decoder is not affected by error propagation provided that $nd - 1$ (which in our case coincides with the traceback depth L_D^*) is large enough in order to have a very high probability of path merge at delay less than $nd-1$. However, as expected, the BER performance is affected by error propagation and a loss of about 1 dB is observed due to non-perfect decisions.

Example 2: Here we consider again 2 transmit and 2 receive antennas, however a reduced-overhead layered space-time coded system is used to cut in half the overhead penalty. We use the same code as in the preceding example, with traceback depth $L_D^* = 39$ and a corresponding "just-in-time" diagonal width of $d = 10$. The spectral efficiency of this system is $\eta = 1.8142$ bits/s/Hz. Figure 2.10 provides the FER and

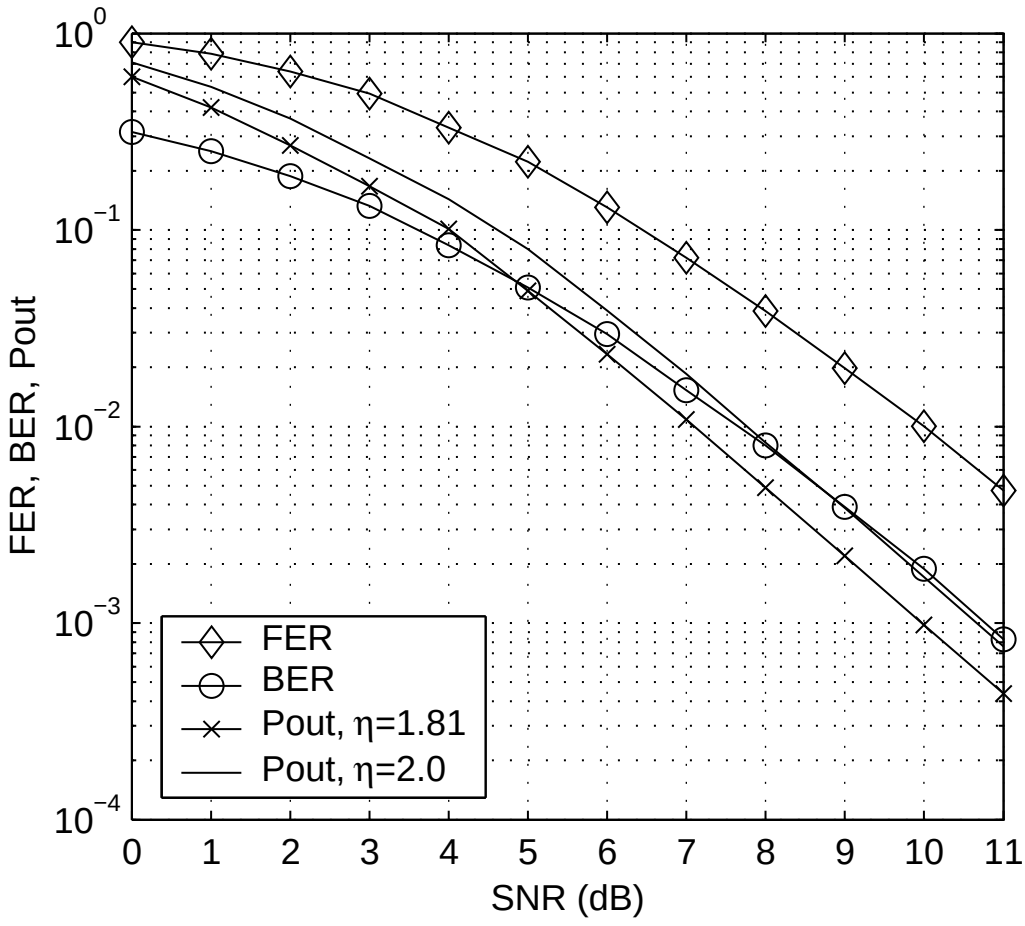


Figure 2.10: Outage probability, FER, and BER for 2×2 MMSE reduced-overhead system coded with rate-1/3, 8-PSK trellis code, blocklength 127 information bits per transmit antenna.

BER performance versus the average SNR per receive antenna. For comparison we also provide the outage probability curves for a 2×2 MIMO system at a transmitted rate of 2 bits/s/Hz and 1.8142 bits/s/Hz respectively. The absolute FER and BER performance of this system is very close to the performance of the full-overhead system, however because the overhead is cut in half, the gap to outage is decreased. For example, at 10% outage, the FER performance is around 2.4 dB from the outage curve that takes into account the overhead penalty.

Example 3: In this example we consider the 2×2 reduced-overhead layered space-time system when the diagonal width is varied and the traceback depth is fixed to $L_D = 41$. Figure 2.11 shows the FER performance for 10 different layerings corresponding to diagonal widths $d = 1, 2 \ldots 10$. As expected, performance degrades with smaller diagonal widths since the Viterbi decoder has to trace through many *bad* symbols - symbols detected by suppression with a diversity gain of one, and only a few *good* symbols - symbols detected by canceling and suppression with a diversity gain of two. However, the rate loss incurred by overhead is directly proportional to the diagonal width, therefore the better performance results come at the expense of a higher rate penalty. In order to correctly assess the performance with various diagonal widths, we consider the SNR gap to the outage curve associated with each layering and its overall transmission rate. Figure 2.12 shows these SNR gaps at 10% outage versus the corresponding diagonal width d. This plot illustrates the trade-off between performance relative to outage (including the overhead penalty). It demonstrates that a layering for

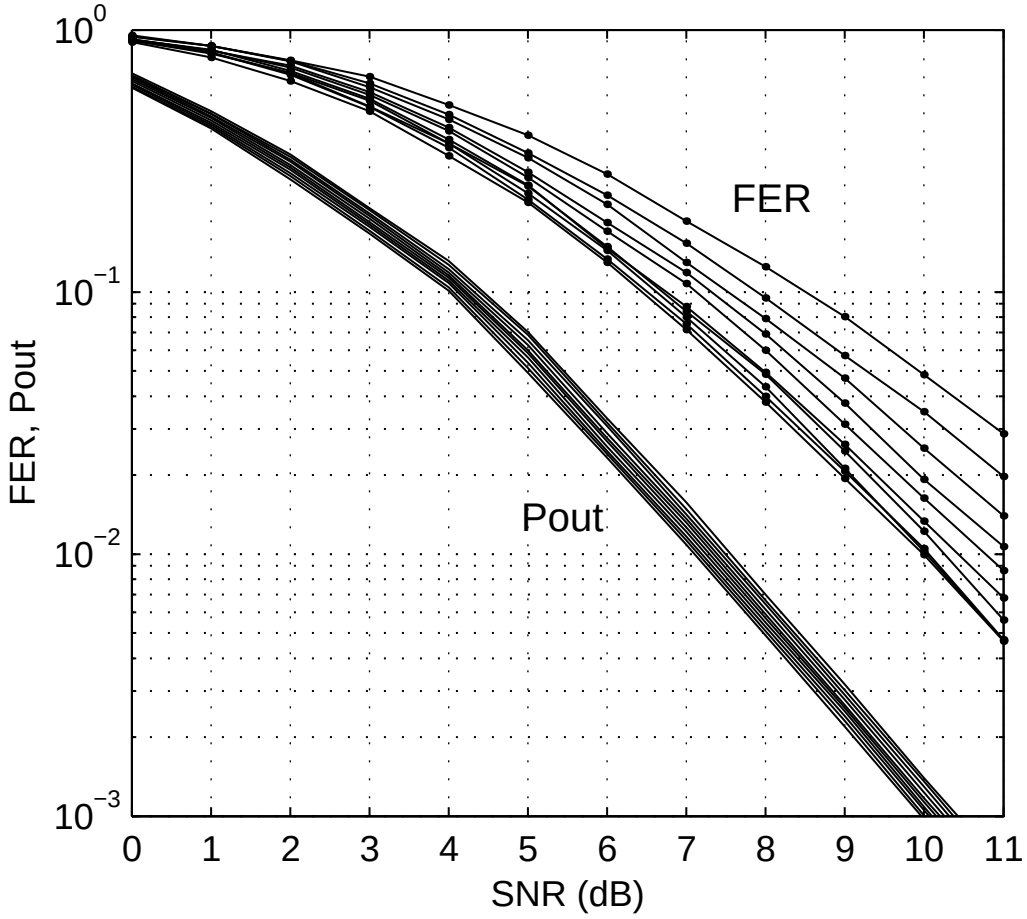


Figure 2.11: Outage probability and FER for 2×2 MMSE reduced-overhead system coded with rate-1/3, 8-PSK trellis code and diagonal widths $d = 1, 2, \ldots 10$ (top to bottom).

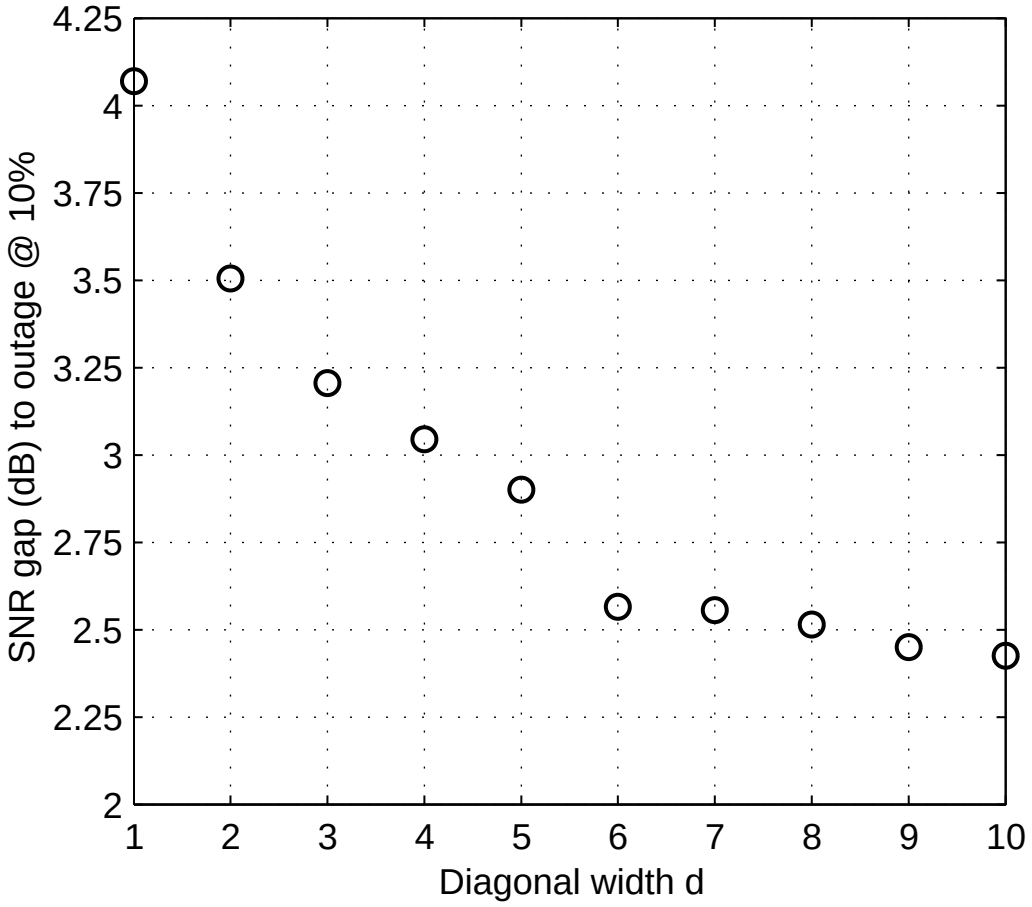


Figure 2.12: SNR gap to outage for 2×2 MMSE reduced-overhead system coded with rate-1/3, 8-PSK trellis code and different diagonal widths d.

continuous Viterbi decoding with the "just-in-time" diagonal width discussed in Section 2.4.1 (which corresponds to the point for $d = 10$ in Fig. 2.12) is in fact the optimal choice for a given traceback depth L_D.

Example 4: Here we consider 4 transmit and 4 receive antennas and a reduced-overhead layered space-time coded system. For this system we use code $\#7$ in [47] designed for period-8 Rayleigh fading. This is a rate-1/4, 64-state, 16-QAM trellis code with octal generators $(G_1, G_2, G_3, G_4) = (63, 175, 135, 103)$. The labeling used for the 16-QAM constellation is the same as in [47], i.e. $3, 1, 5, 7, 2, 0, 4, 6, a, 8, c, e, b, 9, d, f$ in hexadecimal in raster order. The symbol layering is for a codeword of 468 information bits (127 information bits per transmit antenna). The traceback depth is $L_D^* = 39$ and the corresponding "just-in-time" diagonal width is $d = 5$. The resulting spectral efficiency of this system is $\eta = 3.5276$ bits/s/Hz. Figure 2.13 shows the FER and BER performance of the 4×4 layered space-time system together with the two outage curves at a rate of 4.0 bits/s/Hz and 3.5276 bits/s/Hz. At 10% outage, the FER performance of this system is around 3 dB from the theoretical outage curve neglecting overhead, and 3.9 dB from the outage curve at the overall rate of 3.5276 bits/s/Hz, which takes the overhead penalty into account. Figure 2.13 also provides the performance of the 4×4 system when perfect decisions are used in the cancellation step. Again, the FER performance is unaffected by error propagation, whereas the BER performance degrades by 0.5 dB or more due to nonperfect decisions.

Including the loss due to the overhead penalty incurred by D-BLAST, the perfor-

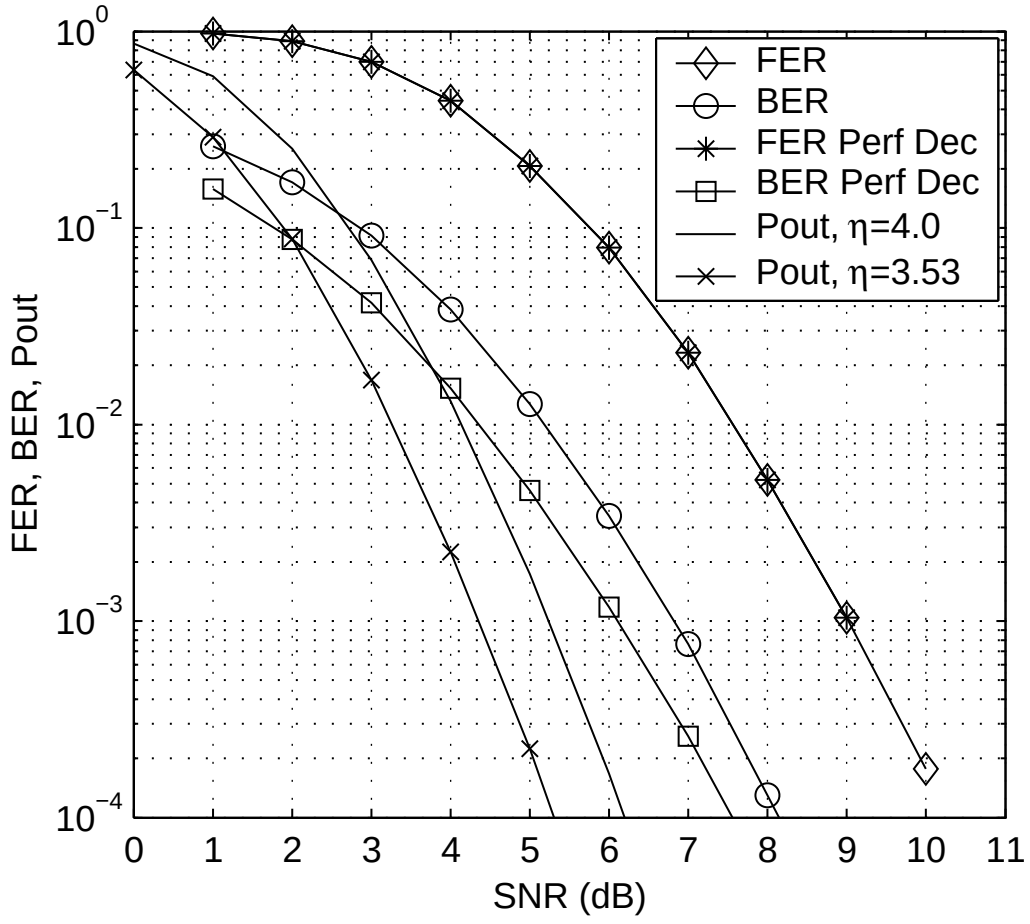


Figure 2.13: Outage probability, FER, and BER for 4×4 MMSE reduced-overhead system coded with rate-1/4, 16-QAM trellis code, blocklength 127 information bits per transmit antenna.

mance of this code is approximately 1.4 dB worse than the performance obtained by El Gamal and Hammons in [17] with the TST architecture. At 10% outage the TST scheme with rate-1/2 component codes, QPSK modulation, and 4 bits/s/Hz spectral efficiency, performs around 2.5 dB from the outage capacity. The performance gain of the TST scheme can be attributed to two factors: the iterative MMSE receiver and a smaller QPSK constellation. However, the TST scheme with rate-1/2 QPSK component codes performs poorly when two of the channel eigenvalues are zero, since it has no remaining redundancy in that case and the best it can achieve is an uncoded performance. Our 16-QAM trellis code can potentially work in this scenario since it still has redundancy left.

Example 5: In this example we consider an 8×8 full-overhead layered space-time system coded with code $\#7$ in [47]. The underlying symbol layering is for a codeword of 936 information bits (127 information bits per transmit antenna) with a traceback depth $L_D^* = 39$. The "just-in-time" step-size of the diagonal layering is $d = 5$ and the resulting overall spectral efficiency is $\eta = 6.2328$ bits/s/Hz. Figure 2.14 provides the FER and BER performance of the 8×8 full-overhead layered space-time system together with the theoretical outage curves. At 10% outage, this system has a FER performance of 3.6 dB from the outage curve neglecting overhead (the $\eta = 4.0$ curve). An additional loss of 1.7 dB is incurred due to the inherent overhead penalty. However, from our experience we predict that a reduced-overhead system would yield very similar performance to the full-overhead system. Moreover, with a reduced-overhead

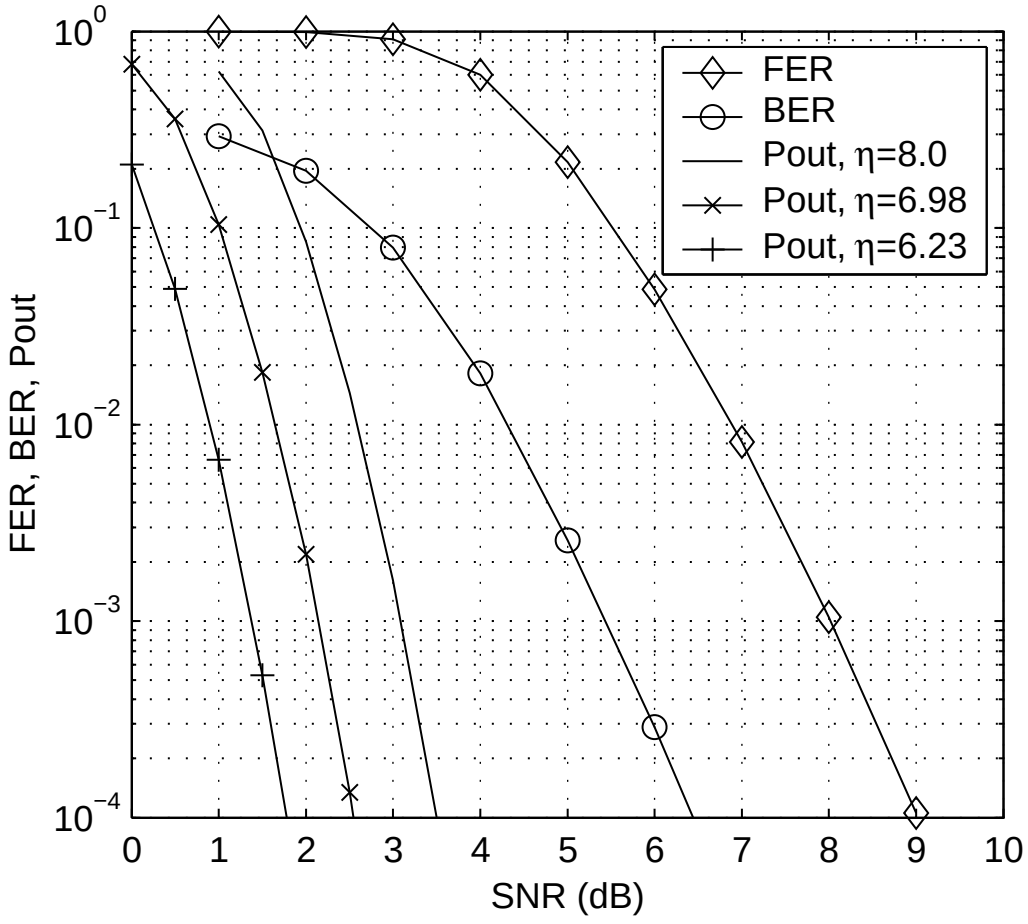


Figure 2.14: Outage probability, FER, and BER for 8×8 MMSE full-overhead system coded with rate-1/4, 16-QAM trellis code, blocklength 127 information bits per transmit antenna.

system the overall rate is increased to 6.9828 bits/s/Hz and the loss due to the overhead penalty is decreased to 0.9 dB. This can also be seen in Fig. 2.14 which provides the $\eta = 6.9828$ outage curve corresponding to a reduced-overhead system.

The TST architecture with rate-1/2 QPSK component codes and spectral efficiency 8 bits/s/Hz is again around 2.5 dB from the outage capacity, having a performance gain of 2 dB or more over a 16-QAM trellis coded D-BLAST system including the loss due the overhead penalty. As in the previous example, this gain in performance is attributed to the superiority of the iterative MMSE receiver and the use of a smaller constellation.

2.7 Conclusion

In this chapter we have considered a simple trellis coding technique for the layered space-time architecture of D-BLAST which employs a single trellis code and finite-traceback Viterbi decoding with traceback depth L_D. Several results using universal trellis codes are presented for different antenna configurations. We show that a 2×2 reduced-overhead trellis coded D-BLAST system performs within 2.5 dB of the outage curve which takes the overhead penalty into account and is competitive with the best available space-time trellis/turbo codes.

One advantage of the coded D-BLAST system is the fact that it uses a single trellis code and an MMSE detector that permits decoding via Viterbi's algorithm, without the need for iterative decoding. Also, this system scales nicely with the number of trans-

mit/receive antennas. The complexity of the MMSE detector is dominated by n times the complexity of computing the suppression vectors $\mathbf{w}_i$ in (2.11). This complexity is polynomial in n, in comparison with multiuser detection methods with m^n complexity (m is the signal constellation size). One drawback of D-BLAST is the inherent overhead penalty due to the diagonal layering. However, one way to eliminate the inherent rate loss is by concatenating consecutive blocks such that the leading lower triangle of overhead symbols in a block is filled by the tailing lower triangle of data symbols from the previous block. Another solution is simply to use longer blocks. This solution does not work well if low raw FER is the design objective because longer blocklengths lead to higher raw FER. However, if low BER is the design objective or if outer coding can correct a fixed percentage of bit errors as frame length is increased, then D-BLAST becomes a competitive solution with both good performance and a simple decoder.

Chapter 3

Gaussian Input and Constrained Modulation Capacity of D-BLAST Systems

3.1 Gaussian Input

The results of Ariyavisitakul [7] show that D-BLAST under the MMSE criterion is optimal in the sense that it achieves the Gaussian input capacity for MIMO channels. Below we give an alternative proof of this result. We have learned that this proof is also contained in soon-to-be published [2].

Theorem 1 *Space-time processing based on successive cancellations and MMSE in-*

terference suppressions converts a $n \times n$ MIMO channel into an equivalent period-n scalar channel with the same channel capacity per n-symbol period as the original $n \times n$ channel,

$$C(\mathbf{H}) = \log_2 \det\left(\mathbf{I}_n + \frac{E_x}{N_0}\mathbf{H}\mathbf{H}^\dagger\right) = \sum_{i=1}^{n} \log_2\left(1 + \gamma_i^{\text{MMSE}}\right). \tag{3.1}$$

Proof. For simplicity assume that $E_x/N_0 = 1$. Using the determinant identity $\det(\mathbf{I}+\mathbf{A}\mathbf{B}) = \det(\mathbf{I}+\mathbf{B}\mathbf{A})$, we can rewrite the capacity equation as

$$C(\mathbf{H}) = \log_2 \det\left(\mathbf{I}_n + \mathbf{H}^\dagger\mathbf{H}\right). \tag{3.2}$$

Let $\mathbf{H} = [\mathbf{H}_{[1,n-1]}\ \mathbf{h}_n]$. Then, it follows that

$$\mathbf{I}_n + \mathbf{H}^\dagger\mathbf{H} = \begin{bmatrix} \mathbf{I}_{n-1} + \mathbf{H}_{[1,n-1]}^\dagger\mathbf{H}_{[1,n-1]} & \mathbf{H}_{[1,n-1]}^\dagger\mathbf{h}_n \\ \mathbf{h}_n^\dagger\mathbf{H}_{[1,n-1]} & 1 + \mathbf{h}_n^\dagger\mathbf{h}_n \end{bmatrix}. \tag{3.3}$$

Using the fact that

$$\det\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} = \det(\mathbf{A})\det(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}) \tag{3.4}$$

when $\mathbf{A}$ is nonsingular, we obtain

$$C(\mathbf{H}) = \log_2\left[\det\left(\mathbf{I}_{n-1} + \mathbf{H}_{[1,n-1]}^\dagger\mathbf{H}_{[1,n-1]}\right)\left(1 + \mathbf{h}_n^\dagger\mathbf{W}\mathbf{h}_n\right)\right], \tag{3.5}$$

where

$$\mathbf{W} = \mathbf{I}_n - \mathbf{H}_{[1,n-1]} \left(\mathbf{I}_{n-1} + \mathbf{H}_{[1,n-1]}^{\dagger} \mathbf{H}_{[1,n-1]} \right)^{-1} \mathbf{H}_{[1,n-1]}^{\dagger}. \tag{3.6}$$

The Sherman-Morrison-Woodbury formula [15, p.50] provides a convenient simplification for (3.6), namely

$$\mathbf{W} = \left(\mathbf{I}_n + \mathbf{H}_{[1,n-1]} \mathbf{H}_{[1,n-1]}^{\dagger} \right)^{-1}. \tag{3.7}$$

This in turn reduces equation (3.5) to

$$\begin{aligned} C(\mathbf{H}) = & \log_2 \left[\det \left(\mathbf{I}_{n-1} + \mathbf{H}_{[1,n-1]}^{\dagger} \mathbf{H}_{[1,n-1]} \right) \times \right. \\ & \left. \left(1 + \mathbf{h}_n^{\dagger} \left(\mathbf{I}_n + \mathbf{H}_{[1,n-1]} \mathbf{H}_{[1,n-1]}^{\dagger} \right)^{-1} \mathbf{h}_n \right) \right] \\ = & \log_2 \left[\det \left(\mathbf{I}_{n-1} + \mathbf{H}_{[1,n-1]}^{\dagger} \mathbf{H}_{[1,n-1]} \right) \times \right. \\ & \left. \left(1 + \gamma_n^{\text{MMSE}} \right) \right], \end{aligned} \tag{3.8}$$

where the expression for γ_n^{MMSE} is given in (2.14). Repeated application of the above procedure to further factor $\det \left(\mathbf{I}_{n-1} + \mathbf{H}_{[1,n-1]}^{\dagger} \mathbf{H}_{[1,n-1]} \right)$, yields

$$C(\mathbf{H}) = \log_2 \left[\left(1 + \gamma_1^{\text{MMSE}} \right) \ldots \left(1 + \gamma_n^{\text{MMSE}} \right) \right]. \tag{3.9}$$

Therefore,

$$C(\mathbf{H}) = \sum_{i=1}^{n} \log_2 \left(1 + \gamma_i^{\text{MMSE}} \right), \tag{3.10}$$

and the proof of the theorem is complete.

The above theorem proves the optimality of the layered space-time system under the assumption of perfect cancellations, i.e. no decision errors affect the detection in subsequent layers. This proof also provides some insight into how D-BLAST achieves the MIMO channel capacity. In particular, observe that the scalar periodic SNRs produced under MMSE suppression are in general not equal to the eigenvalues of the matrix $\mathbf{H}\mathbf{H}^{\dagger}$.

Under the ZF criterion, the operational capacity of D-BLAST is a lower bound on the true MIMO channel capacity, as shown in [13],

$$C(\mathbf{H}) > C_{\text{ZF-DBLAST}}(\mathbf{H}) = \sum_{i}^{n} \log_2 \left(1 + \gamma_i^{\text{ZF}}\right). \tag{3.11}$$

Figure 3.1 shows the MIMO capacity in (3.1) together with its lower bound in (3.11) under Rayleigh fading for 2×2, 4×4, and 8×8 systems. These curves were evaluated numerically using Monte Carlo methods (sampling of the channel matrix). Note that the capacity lower bound (operational capacity of ZF D-BLAST) becomes looser for larger antenna configurations. Also note that the MMSE and ZF solutions for D-BLAST are equivalent in the limit of large SNR.

3.2 Constrained Modulation Input

Here we assume that the transmit vector $\mathbf{x}$ has entries chosen from some complex constellation with $2^{M_c} (M_c \geq 1)$ possible signal points. We restrict our attention to PSK

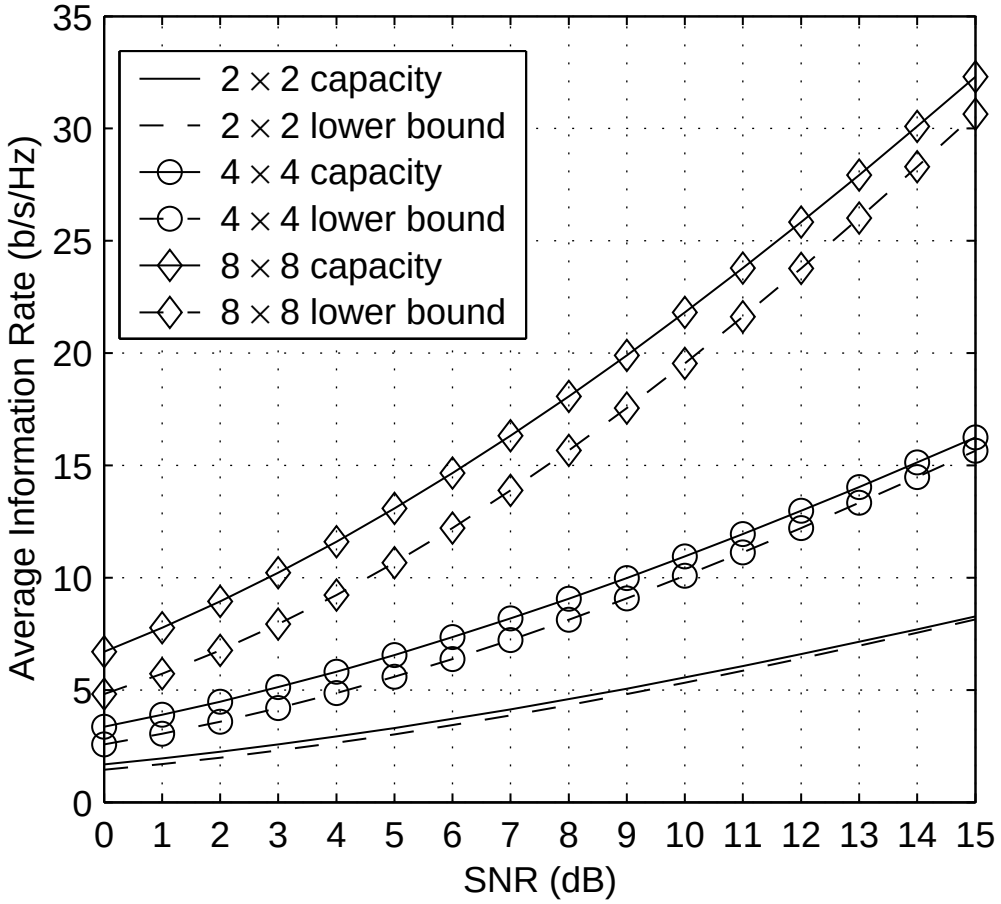


Figure 3.1: Ergodic capacity and its lower bound, Gaussian input, Rayleigh fading.

or QAM constellations. To see the effect of a constellation on the maximum achievable rate, we compute the mutual information between the output $\mathbf{y}$ and the input $\mathbf{x}$ assuming the entries of $\mathbf{x}$ are chosen independently and equally likely from the constellation. The mutual information is computed as

$$\mathcal{I}(\mathbf{x};\mathbf{y}|\mathbf{H}) = \mathcal{H}(\mathbf{y}|\mathbf{H}) - \mathcal{H}(\mathbf{y}|\mathbf{x},\mathbf{H}), \tag{3.12}$$

where $\mathcal{H}(\cdot) = -\mathsf{E}\log_2 p(\cdot)$ is the entropy function. It is easy to show that

$$\mathcal{H}(\mathbf{y}|\mathbf{x},\mathbf{H}) = \mathcal{H}(\mathbf{n}) = n\log_2(2\pi e\sigma^2) \tag{3.13}$$

for our channel model with any constellation. The term $\mathcal{H}(\mathbf{y}|\mathbf{H})$ in (3.12) is in general more difficult to compute since the expectation in $\mathcal{H}(\mathbf{y}|\mathbf{H}) = -\mathsf{E}\log_2 p(\mathbf{y}|\mathbf{H})$ is over two sources of randomness in the choices of $\mathbf{x}$ and $\mathbf{n}$. The expectation over $\mathbf{x}$ can be computed using a sum,

$$\mathcal{H}(\mathbf{y}|\mathbf{H}) = -\mathsf{E}\log_2\left(\frac{1}{2^{nM_c}(2\pi\sigma^2)^n}\sum_{\mathbf{x}}\exp\left[-\frac{1}{2\sigma^2}\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2\right]\right), \tag{3.14}$$

where the sum over $\mathbf{x}$ runs for all 2^{nM_c} possible combinations of the constellation values. The above expectation is now over the noise vector $\mathbf{n}$. The expectation over $\mathbf{n}$ can be evaluated via numerical integration methods. Note, however, that if an $L \times L$ grid is used to quantize each complex received dimension (antenna) in $\mathbf{y}$ then L^{2n} points exist

in the PMF approximation of $p(\mathbf{y}|\mathbf{H})$ and numerical evaluation becomes computationally intensive.

Instead, we note that under the assumption of correct decisions being canceled out, D-BLAST with MMSE interference suppression decomposes the MIMO channel into n parallel complex constant fading channels with SNRs $\gamma_1^{\text{MMSE}}, \gamma_2^{\text{MMSE}} \ldots \gamma_n^{\text{MMSE}}$ given by (2.14), such that the aggregate capacity is the same as the original capacity. In a constrained modulation setting, the residual interference plus noise at the output of the MMSE filter is neither Gaussian nor *i.i.d.* Let $\hat{x}_k$ be the detection statistic for the kth antenna at the output of the MMSE suppression step,

$$\hat{x}_k = \mathbf{w}_k^\dagger \mathbf{y}_k \tag{3.15}$$

$$= \underbrace{(\mathbf{w}_k^\dagger \mathbf{h}_k) x_k}_{\text{desired term}} + \underbrace{\sum_{j=1}^{k-1} (\mathbf{w}_j^\dagger \mathbf{h}_j) x_j}_{\text{co-antenna interference}} + \underbrace{\mathbf{w}_k^\dagger \mathbf{n}}_{\text{phase-rotated noise}} . \tag{3.16}$$

In the above expression, the first term represents the desired term, the second term is the residual interference from the suppressed $k-1$ antennas, and the last term is a phase-rotated noise term.

In the following, as done in [10], we make a Gaussian approximation for the MMSE filter output:

$$\hat{x}_k \approx x_k^{\text{Gauss}} = \mu_k x_k + z_k, \tag{3.17}$$

where

$$z_k \sim \mathcal{N}(0, \eta_k^2). \tag{3.18}$$

In other words, the Gaussian noise vector $\mathbf{z}$ is assumed to have independent entries with zero-mean and variance given by η_k^2, where

$$\begin{aligned} \eta_k^2 &= E_x \mu_k (1 - \mu_k), \quad \text{and} && (3.19) \\ \mu_k &= \mathbf{w}_k^\dagger \mathbf{h}_k. && (3.20) \end{aligned}$$

With the above Gaussian approximation the resulting mutual information is an approximation lower bound to the true mutual information:

$$\begin{aligned} \mathcal{I}(\mathbf{x}; \mathbf{y}|\mathbf{H}) &= \mathcal{I}(\mathbf{x}; \hat{\mathbf{x}}|\mathbf{H}) && (3.21) \\ &= \mathcal{H}(\hat{\mathbf{x}}|\mathbf{H}) - \mathcal{H}(\hat{\mathbf{x}}|\mathbf{x}, \mathbf{H}) && (3.22) \\ &\geq \mathcal{H}(\hat{\mathbf{x}}|\mathbf{H}) - \mathcal{H}(\mathbf{z}|\mathbf{H}) && (3.23) \\ &\approx \mathcal{H}(\mathbf{x}^{\text{Gauss}}|\mathbf{H}) - \mathcal{H}(\mathbf{z}|\mathbf{H}) && (3.24) \\ &= \mathcal{I}(\mathbf{x}; \mathbf{x}^{\text{Gauss}}|\mathbf{H}). && (3.25) \end{aligned}$$

where (3.23) follows from the fact that Gaussian distribution maximizes the entropy over all distributions with the same covariance and (3.24) follows from the Gaussian approximation in (3.17).

Then, the mutual information calculation is simplified since the expectation in (3.14) now reduces to the following expectation

$$\mathcal{H}(\mathbf{y}|\mathbf{H}) \approx -\sum_{k=1}^{n} \mathsf{E} \log_2 \left(\frac{1}{2^{M_c}} \sum_{x} \frac{1}{\pi \eta_k^2} \exp \left[-\frac{1}{\eta_k^2} \|x_k^{\text{Gauss}} - \mu_k x\|^2 \right] \right), \qquad (3.26)$$

where the summation over x is for all 2^{M_c} points in the constellation.

In essence, the D-BLAST approach with Gaussian approximation can be used to linearize the constrained channel capacity calculation in (3.12), thus significantly reducing the computational complexity for system configurations with more than 2 transmit/receive antennas. The expectation over $\mathbf{H}$ can be evaluated numerically using Monte-Carlo sampling of the channel matrix. From our experience, this technique yields a tight approximation for low and medium SNRs (within thousands of a bit). Figures 3.2, 3.3, and 3.4 show the average information rate obtained with the D-BLAST technique with Gaussian approximation for BPSK, QPSK, and 8-PSK constellation constrained input respectively, under Rayleigh fading and 2×2, 3×3, and 4×4 antenna configurations.

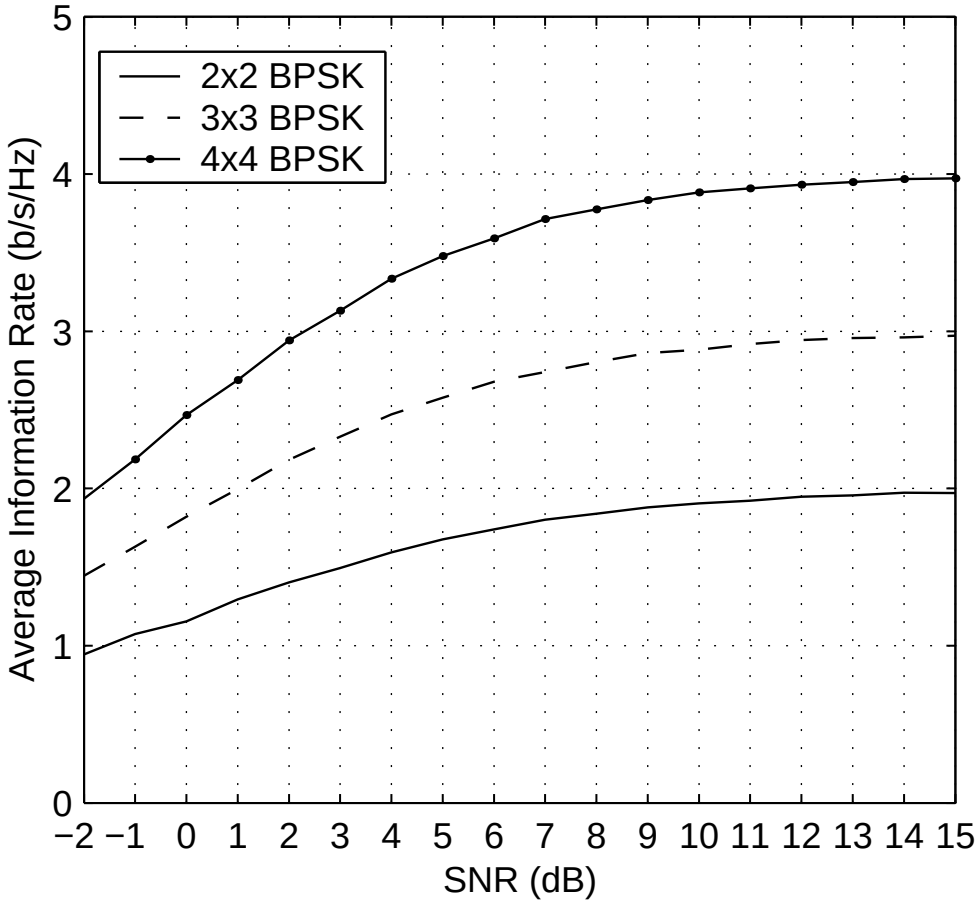


Figure 3.2: Information rates for 2×2, 3×3, and 4×4 BPSK constellation constrained input, Rayleigh fading.

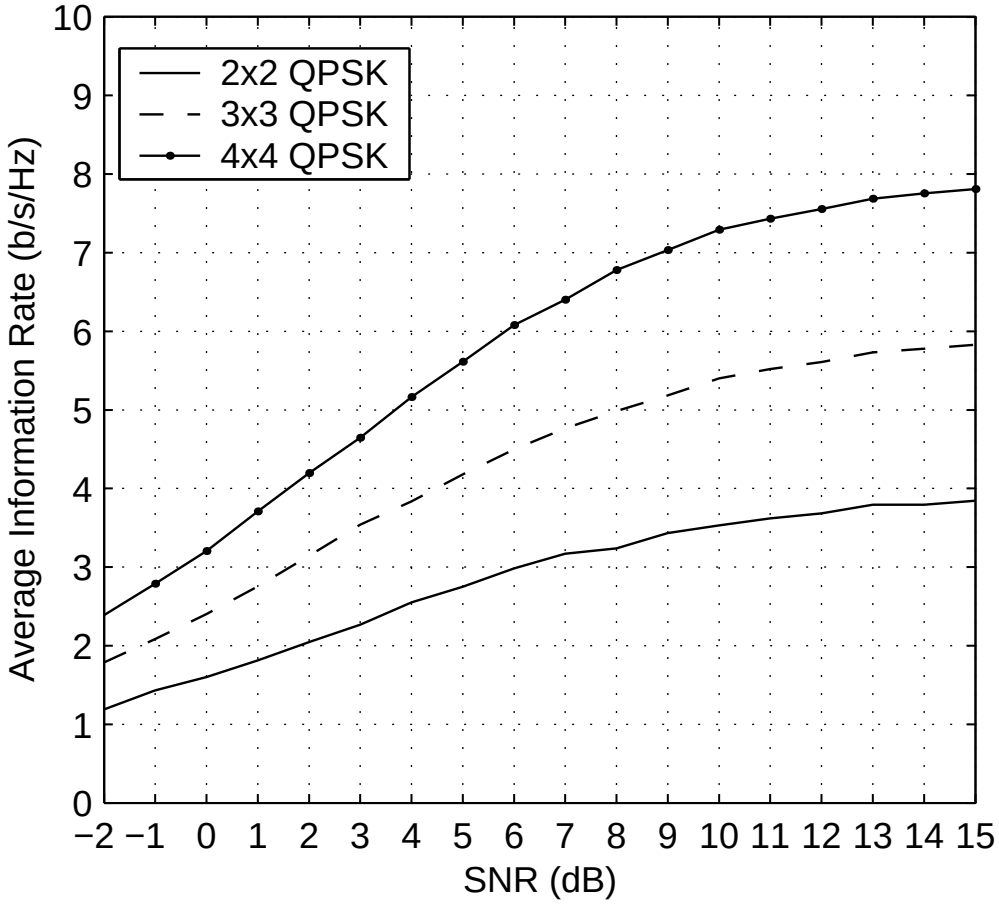


Figure 3.3: Information rates for 2×2, 3×3, and 4×4 QPSK constellation constrained input, Rayleigh fading.

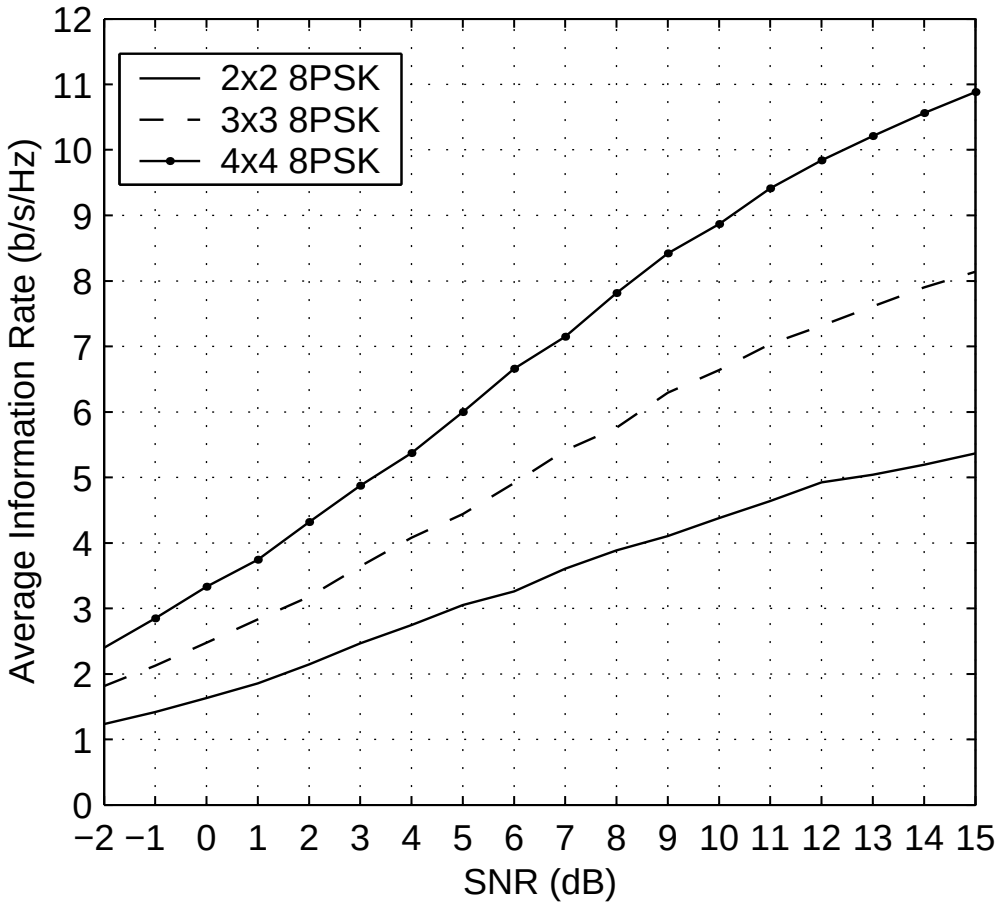


Figure 3.4: Information rates for 2×2, 3×3, and 4×4 8-PSK constellation constrained input, Rayleigh fading.

Chapter 4

Low-Density Parity-Check Coded MIMO Systems

4.1 LDPC Basics

Low-density parity-check (LDPC) codes were proposed by Gallager in the early 1960s [16]. The structure of Gallager's codes (uniform column and row weight) led them to be called regular LDPC codes. Gallager provided simulation results for codes with block lengths on the order of hundreds of bits. However, these codes were too short for the sphere packing bound to approach Shannon capacity, and the computational resources for longer random codes were decades away from being broadly accessible.

Following the groundbreaking demonstration by Berrou *et al.* [8] of the impressive

capacity-approaching capability of long random linear (turbo) codes, MacKay [30] re-established interest in LDPC codes during the mid to late 1990s. Luby *et al.* [29] formally showed that properly constructed irregular LDPC codes can approach capacity more closely than regular ones. Richardson, Shokrollahi and Urbanke [34] created a systematic method called density evolution to analyze and synthesize the degree distribution in asymptotically large random bipartite graphs under a wide range of channel realizations. The authors showed that it is possible to predict a noise threshold below which a code realized from a given ensemble can be expected to converge to zero errors with high probability.

In a nutshell, an LDPC code is a linear binary block code specified by a very *sparse* parity-check matrix. The parity-check matrix H of a *regular* (n, k, s, t) LDPC code is a $(n-k) \times n$ matrix, which has s ones in each column and $t > s$ ones in each row where $s << n$, and the ones are typically placed at random in the parity-check matrix. When the number of ones in every column is not the same, the code is known as an *irregular* LDPC code.

Associated with the code there is a simple bipartite graph representation which consists of two types of nodes: *variable* or *left* nodes and *check* or *right* nodes. Each code bit is a variable node while each parity check or each row of the parity-check matrix represents a check node. An edge in the graph is placed between variable node i and check node j if $H_{j,i} = 1$. In other words, each check node is connected to code bits whose sum modulo-2 should be zero. As an example, consider the code of length

$n = 10$ and $k = 5$, where

$$H = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}. \tag{4.1}$$

The bipartite graph representing the code is shown in Fig. 4.1. Note that each check node represents one linear constraint (one row of H). In this example, every variable node has degree three and each check node has degree six. Therefore, this code is called a $(3, 6)$ regular low-density parity-check code.

As shown in [29] the performance of an LDPC code can be improved considerably if we allow nodes of various degrees and if we choose the proportion of the various degrees carefully. An irregular LDPC code is specified by a degree distribution pair. In general, we call $\gamma(x)$ a *degree distribution* if $\gamma(x)$ is a real valued polynomial with non-negative coefficients and $\gamma(1) = 1$. Let d_l and d_r denote the maximum variable node and check node degrees, respectively. Also let

$$\lambda(x) = \sum_{i=1}^{d_l} \lambda_i x^{i-1}, \text{ and} \tag{4.2}$$

$$\rho(x) = \sum_{i=1}^{d_r} \rho_i x^{i-1} \tag{4.3}$$

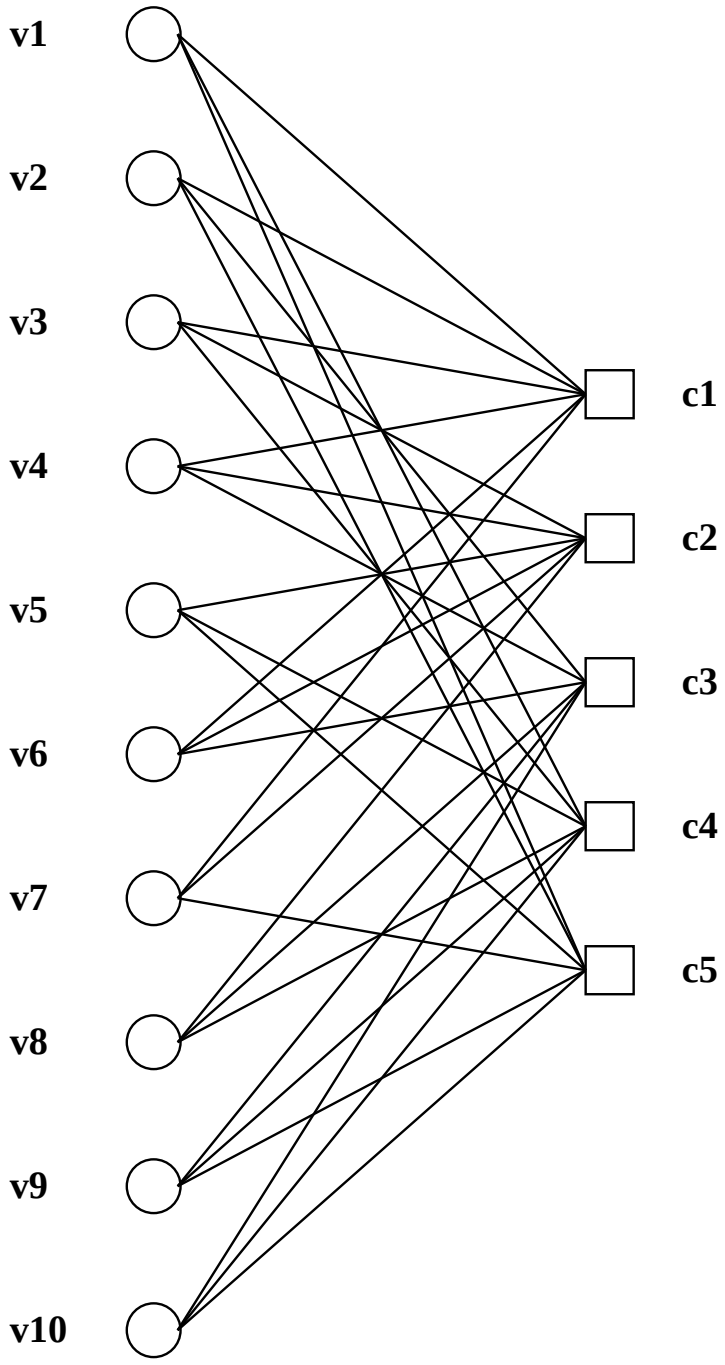


Figure 4.1: A $(3, 6)$ regular LDPC code of length 10.

be two degree distributions. The coefficients of $\lambda(x)(\rho(x))$ represent the fraction of *edges* emanating from variable (check) nodes of degree i. Note that the fraction of edges that emanate from a degree i node is the coefficient of x^{i-1}. Conversion between edge and node perspective is useful for defining the rate of the code in terms of the degree distribution pair (λ, ρ). Let E be the total number of edges in the graph, then $E\lambda_i/i$ denotes the number of variable nodes with degree i. It is easy to check that the design rate $R(\lambda, \rho)$ is given by

$$R(\lambda, \rho) \quad = \quad 1 - \frac{\sum \frac{\rho_i}{i}}{\sum \frac{\lambda_i}{i}} == 1 - \frac{\int_0^1 \rho(x)dx}{\int_0^1 \lambda(x)dx}. \tag{4.4}$$

4.2 System Model

The system model under consideration is an LDPC-coded MIMO system with n_t transmitter antennas and n_r receiver antennas, signaling through frequency-nonselective fading. The transmitter structure is illustrated in Fig. 4.2. The information data is first encoded by a rate-R LDPC code, modulated by a complex constellation with 2^{M_c} possible signal points and average energy E_x, and then distributed among the n_t antennas. Let $\mathbf{x}$ be an $n_t \times 1$ vector of transmitted symbols with components $x_1, x_2, \ldots, x_{n_t}$ and $\mathbf{y}$ an $n_r \times 1$ vector of received signals with components $y_1, y_2, \ldots, y_{n_r}$, related by

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{4.5}$$

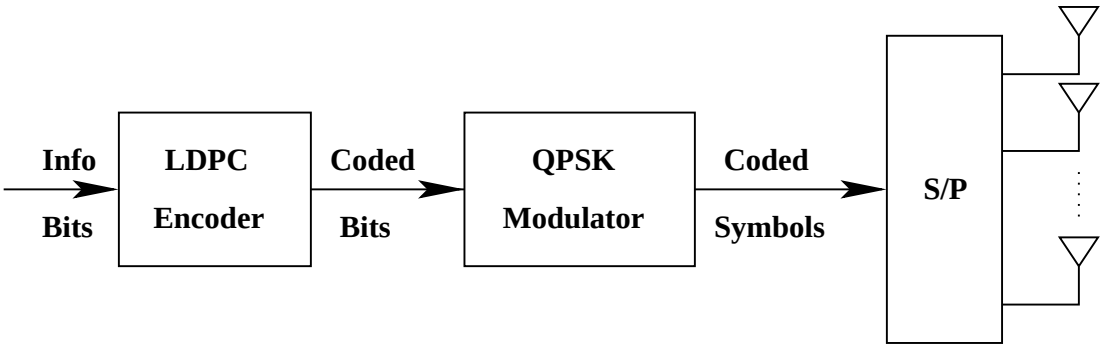


Figure 4.2: Transmitter structure of an LDPC coded MIMO system.

where $\mathbf{H} = [\mathbf{h}_1 \mathbf{h}_2 \ldots \mathbf{h}_{n_t}]$ is the $n_r \times n_t$ complex channel matrix known perfectly by the receiver, and $\mathbf{n}$ is a vector of independent zero-mean, complex Gaussian noise entries with variance $\sigma^2 = N_0/2$ per real component. We assume that the average signal-to-noise ratio at each receiver antenna, denoted by ρ, is independent of the number of transmitter antennas n_t. The system spectral efficiency is $RM_c n_t$ bits/channel use.

4.3 Iterative Detection and LDPC Decoding

With the system model described above, we can apply the turbo principle to decode the received signals. Iterative detection-and-decoding is used to approach the maximum-likelihood (ML) performance of joint MIMO detection and LDPC decoding. Figure 4.3 gives a flowchart of the iterative detector/decoder structure. In this structure, the soft MIMO detector incorporates extrinsic information provided by the LDPC decoder, and the LDPC decoder incorporates soft information provided by the

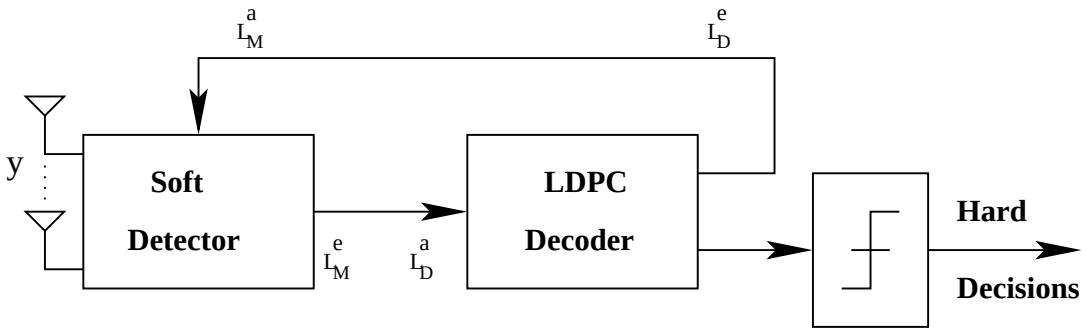


Figure 4.3: Turbo iterative detection and decoding receiver for an LDPC coded MIMO system.

MIMO detector. There are two kinds of iterations involved: one is the outer loop for both the detector and decoder, the other is the inner loop for the decoder only. For each outer-loop iteration, the soft detector accepts as its inputs the channel observation $\mathbf{y}$ and the *a priori* information L_M^a of the coded bits b and generates extrinsic information L_M^e. This extrinsic information becomes *a priori* input L_D^a for the LDPC decoder. The LDPC decoder carries out the message-passing (also known as belief-propagation) decoding algorithm [35] and produces the extrinsic output L_D^e. Then, L_D^e becomes the input L_M^a for the next outer-loop iteration. Extrinsic information between the detector and decoder is exchanged in this iterative fashion until an LDPC codeword is found or a maximum number of iterations is performed. With LDPC codes, convergence to a codeword is easy to detect since we need only verify that the parity checks are satisfied.

In the following, we describe in more detail how the soft extrinsic information is

computed by each component. We focus on the soft MIMO detector and discuss several reduced-complexity solutions along with the standard maximum *a posteriori* (MAP) detector. We examine the complexity of each scheme and select specific parameters for modulation cardinality and transmit-receive antenna multiplicity to facilitate a numerical comparison of complexity.

4.3.1 MAP Detector

In the soft MAP detector, the received vector $\mathbf{y}$ is demapped by a log-likelihood ratio (LLR) calculation for each of the $n_t M_c$ coded bits included in the transmit vector $\mathbf{x}$. The extrinsic information provided by the MAP detector is the difference of the soft-input and soft-output LLR values on the coded bits. For the ith code bit b_i ($i = 1, \ldots, n_t M_c$) of the transmit vector $\mathbf{x}$, the extrinsic LLR value of the estimated bit is computed as

$$\begin{aligned} L_M^e(b_i) &= \log \frac{P(b_i = +1|\mathbf{y}, \mathbf{H})}{P(b_i = -1|\mathbf{y}, \mathbf{H})} - \log \frac{P(b_i = +1)}{P(b_i = -1)} \\ &= \log \frac{\sum_{\mathbf{x} \in \mathcal{X}_i^{+1}} P(\mathbf{y}|\mathbf{x}, \mathbf{H}) P(\mathbf{x})}{\sum_{\mathbf{x} \in \mathcal{X}_i^{-1}} P(\mathbf{y}|\mathbf{x}, \mathbf{H}) P(\mathbf{x})} - L_M^a(b_i) \end{aligned} \tag{4.6}$$

where the *a priori* information $L_M^a(b_i)$ is equal to the extrinsic information of the bit b_i computed by the LDPC decoder in the previous turbo iteration ($L_M^a(b_i) = 0$ at the first iteration) and $\mathcal{X}_i^{+1}$ is the set of $2^{n_t M_c - 1}$ vector hypotheses $\mathbf{x}$ having $b_i = +1$, ($\mathcal{X}_i^{-1}$ is similarly defined). Assuming the bits within $\mathbf{x}$ are statistically independent of one

another, the *a priori* probability $P(\mathbf{x})$ can be written as

$$P(\mathbf{x}) = \prod_{j=1}^{n_t M_c} P(b_j) = \prod_{j=1}^{n_t M_c} \left[1 + \exp(-\mathbf{x}^{b_j} L_M^a(b_j))\right]^{-1}, \tag{4.7}$$

where $\mathbf{x}^{b_j}$ corresponds to the value $(+1, -1)$ of the jth bit in the vector $\mathbf{x}$. In the above LLR value calculation, the likelihood function $P(\mathbf{y}|\mathbf{x}, \mathbf{H})$ is specified by a multivariate Gaussian density function.

Since the cardinality of the vector sets $\mathcal{X}_i^{+1}$ and $\mathcal{X}_i^{-1}$ in (4.6) equals $2^{n_t M_c - 1}$, the complexity of the soft MAP detector is exponential in the number of transmitter antennas and the number of bits per constellation symbol. At each iteration, the MAP detector has to compute the LLRs for $n_t M_c$ bits in each transmit symbol vector. With an equal number of transmit and receive antennas ($n_t = n_r = n$), evaluating (4.6) involves the following steps:

for $i = 1$ to 2^{nM_c}

1) Compute the *a priori* probability $P(\mathbf{x})$ in (4.7): $2nM_c$ flops

end

for $i = 1$ to nM_c

2) Compute the LLR value in (4.6): $2 \cdot 2^{nM_c}$ flops

end

We define a flop as a single addition, subtraction, multiplication or division between two complex numbers. Table lookups to $\exp$ and $\log$ functions are not included in this

complexity analysis because we make the simplification that single operand functions have no cost. Instead, we use an operand counting approach where the total number of "hashing" occurrences between different operands is counted. For instance, the cost of evaluating (4.7) is found using this method by counting the cost of the form $(1+x)^{-1}$ as a single flop (trivial additions with constants are not included since they are not true operands). This form is evaluated nM_c times and $nM_c - 1$ multiplications form the final product. To simplify our counting approach, we round up the multiplication cost to nM_c and the total cost is said to be $2nM_c$ flops.

The likelihood function $P(\mathbf{y}|\mathbf{x}, \mathbf{H})$ is precomputed for all 2^{nM_c} hypotheses at the beginning of the iterative process and has a cost of $2n^2+3n$ flops per vector hypothesis. Then, the (approximate) cost of the MAP detector per turbo iteration is $(4nM_c) \cdot 2^{nM_c}$ flops and the initial cost of precomputing the likelihood functions for all hypotheses is $(2n^2 + 3n) \cdot 2^{nM_c}$ flops.

4.3.2 MMSE-SIC Detector

The sub-optimal demodulator based on a minimum mean-square error soft interference cancellation (MMSE-SIC) criterion consists of a parallel interference canceler followed by an MMSE filter. It is analogous to a multiuser detector proposed in [45]. It is also described in [38] and [27]. Below we provide our review and comments on complexity.

The MMSE-SIC detector first forms soft estimates of the transmitted symbols by

computing the symbol mean $\bar{x}_j$ based on the available a priori information:

$$\bar{x}_j = \sum_{x \in \mathcal{A}} x P(x_j = x), \tag{4.8}$$

where $\mathcal{A}$ is the complex constellation set. The a priori probabilities are calculated assuming the bits within a symbol are statistically independent of one another:

$$P(x_j = x) = \prod_{l=1}^{M_c} [1 + \exp(-x^{b_l} L_M^a(b_{(j-1)M_c+l}))]^{-1}, \tag{4.9}$$

where x^{b_l} indicates the value of the lth bit of symbol x. At the beginning of the iterative process, all symbols are equally likely and their probability is 2^{-M_c}. It follows that for a complex symmetric constellation like QPSK, the soft estimates are equal to zero and in effect no cancellation is performed in the first iteration. In subsequent iterations, as the bit reliabilities provided by the LDPC decoder improve, the soft estimates become closer to their true value.

For the kth transmit antenna, the soft interference from the other $n_t - 1$ antennas is canceled to obtain

$$\mathbf{y}_k = \mathbf{y} - \sum_{j=1, j \neq k}^{n_t} \bar{x}_j \mathbf{h}_j \tag{4.10}$$

$$= x_k \mathbf{h}_k + \sum_{j=1, j \neq k}^{n_t} (x_j - \bar{x}_j) \mathbf{h}_j + \mathbf{n}. \tag{4.11}$$

A detection estimate $\hat{x}_k$ of the transmitted symbol on the kth antenna is obtained by applying a linear filter $\mathbf{w}_k$ to $\mathbf{y}_k$:

$$\hat{x}_k = \mathbf{w}_k^\dagger \mathbf{y}_k \tag{4.12}$$

$$= (\mathbf{w}_k^\dagger \mathbf{h}_k) x_k + \sum_{j=1, j\neq k}^{n_t} (\mathbf{w}_j^\dagger \mathbf{h}_j)(x_j - \bar{x}_j) + \mathbf{w}_k^\dagger \mathbf{n}. \tag{4.13}$$

In the above equation, the first term represents the desired term, the second term is the residual interference from the other transmitter antennas, and the last term is a phase rotated noise term.

The filter $\mathbf{w}_k$ is chosen to minimize the mean-square error between the transmit symbol x_k and the filter output $\hat{x}_k$, and depends on the variance of the symbols used in the cancellation step. It can be shown that the MMSE-SIC solution is given by

$$\mathbf{w}_k = \left(\frac{N_0}{E_x} \mathbf{I}_{n_r} + \mathbf{H} \mathbf{\Delta}_k \mathbf{H}^\dagger \right)^{-1} \mathbf{h}_k, \tag{4.14}$$

where the covariance matrix $\mathbf{\Delta_k}$ is

$$\mathbf{\Delta}_k = \text{diag} \left[\frac{\sigma_{x_1}^2}{E_x}, \ldots, \frac{\sigma_{x_{k-1}}^2}{E_x}, 1, \frac{\sigma_{x_{k+1}}^2}{E_x}, \ldots, \frac{\sigma_{x_{n_t}}^2}{E_x} \right], \tag{4.15}$$

and $\sigma_{x_i}^2$ is the variance of the ith antenna symbol computed as:

$$\sigma_{x_i}^2 = \sum_{x \in \mathcal{A}} |x - \bar{x}_i|^2 P(x_i = x). \tag{4.16}$$

Note that the MMSE-SIC filter adjusts its weights according to the quality of the soft canceled symbols through the covariance matrix $\mathbf{\Delta}_k$. In the two extreme cases, the MMSE-SIC filter reduces to a simple suppression filter or a maximal ratio combining (MRC) filter. The first case corresponds to the case in which the canceled symbols are all zero (i.e. no cancellations) and the symbol variances $\sigma_{x_i}^2$ are all equal to E_x. It follows that the covariance matrix $\mathbf{\Delta_k}$ becomes an identity matrix and the filter reduces to the well-known MMSE suppression filter:

$$\mathbf{w}_k = \left(\frac{N_0}{E_x} \mathbf{I}_{n_r} + \mathbf{H}\mathbf{H}^\dagger \right)^{-1} \mathbf{h}_k. \tag{4.17}$$

The second case corresponds to the case in which the canceled symbols are the true symbols (i.e. perfect cancellations) and the symbol variances $\sigma_{x_i}^2$ are all zero. It can be shown that the MMSE-SIC filter reduces to a filter of the form

$$\mathbf{w}_k = \left(\frac{N_0}{E_x} + \mathbf{h}_k^\dagger \mathbf{h}_k \right)^{-1} \mathbf{h}_k, \tag{4.18}$$

which in effect forms a maximal ratio combining with the corresponding column vector of the channel matrix.

Therefore, in general, the MMSE-SIC detector performs a combination of suppressions and cancellations. The amount of suppression done by the detector is determined by the quality of the canceled symbols, which ultimately dictates the performance of the MMSE-SIC detector.

As we saw before, the output of the MMSE-SIC filter includes the desired symbol, residual co-antenna interference, and noise. Note that under a Gaussian input, the output of the filter is also Gaussian. However, under a constrained input scenario in which the symbols belong to a complex constellation like QPSK, the filter output is neither Gaussian nor *i.i.d.* Despite this fact, following [45], we approximate $\hat{x}_k$ by the output of an equivalent AWGN channel with $\hat{x}_k = \mu_k x_k + z_k$, where

$$\mu_k = \mathbf{w}_k^{\dagger} \mathbf{h}_k, \tag{4.19}$$

and z_k is a zero-mean complex Gaussian variable with variance η_k^2 given by

$$\eta_k^2 = E_x(\mu_k - \mu_k^2). \tag{4.20}$$

Then, the extrinsic log-likelihood ratio computed by the MMSE-SIC detector for the lth bit ($l = 1, \ldots, M_c$) of the symbol x_k transmitted by the kth antenna is:

$$\begin{aligned} L_M^e(b_{(k-1)M_c+l}) &= \log \frac{P(b_{(k-1)M_c+l} = +1|\hat{x}_k)}{P(b_{(k-1)M_c+l} = -1|\hat{x}_k)} - \log \frac{P(b_{(k-1)M_c+l} = +1)}{P(b_{(k-1)M_c+l} = -1)} \\ &= \log \frac{\sum_{x \in \mathcal{A}_l^{+1}} P(\hat{x}_k|x)P(x)}{\sum_{x \in \mathcal{A}_l^{-1}} P(\hat{x}_k|x)P(x)} - L_M^a(b_{(k-1)M_c+l}) \end{aligned} \tag{4.21}$$

where $\mathcal{A}_l^{+1}$ is the set of 2^{M_c-1} hypotheses x for which the lth bit is $+1$ ($\mathcal{A}_l^{-1}$ is similarly defined). In the above calculation of the extrinsic LLR value, the *a priori* probability $P(x)$ is given by (4.9) and the likelihood function $P(\hat{x}_k|x)$ is approximated by

$$P(\hat{x}_k|x) \simeq \frac{1}{\pi\eta_k^2}\exp\left(-\frac{1}{\eta_k^2}|\hat{x}_k - \mu_k x|^2\right). \tag{4.22}$$

Note that for the LLR value calculation, only the term in the exponent is relevant; the constant factor outside the exponent can be dropped.

The MMSE-SIC detector has a lower complexity than the MAP detector. This can be seen from (4.21) where the extrinsic LLR is computed from the *scalar* output $\hat{x}_k$ of the MMSE filter, in contrast with (4.6) where the extrinsic LLR is computed from the received vector $\mathbf{y}$. With n transmit and n receive antennas, evaluating (4.21) involves the following steps:

for $k = 1$ to n

for $i = 1$ to 2^{M_c}

1) Compute the *a priori* probability $P(x)$ in (4.9): $2M_c$ flops

end

2) Evaluate the symbol mean $\bar{x}_i$ and variance $\sigma_{x_i}^2$: $5 \cdot 2^{M_c}$ flops

3) Cancel the soft estimates in (4.11): $2n^2$ flops.

4) Evaluate the matrix $\mathbf{G} = \left(\frac{N_0}{E_x}\mathbf{I}_{n_r} + \mathbf{H}\boldsymbol{\Delta}_k\mathbf{H}^\dagger\right)$: $2n^3 + n^2 + 3n$ flops

5) Solve $\mathbf{G}\mathbf{w}_k = \mathbf{h}_k$ for $\mathbf{w}_k$: $n^3/3$ flops using Choleski factorization [15]

6) Compute the detection estimate $\hat{x}_k$: $2n$ flops

7) Compute μ_k and η_k^2: $2n$ flops

for $l = 1$ to 2^{M_c}

8) Evaluate the likelihood function $P(\hat{x}_k|x)$ in (4.22): 4 flops

end

for $l = 1$ to M_c

9) Compute the LLR value in (4.21): $2 \cdot 2^{M_c}$ flops

end

end

Then, the (approximate) computational complexity of the MMSE-SIC detector is $7n^4/3+3n^3+7n^2+(4nM_c)\cdot 2^{M_c}+(9n)\cdot 2^{M_c}$ flops per turbo iteration.

4.3.3 MRC-SIC Detector

A sub-optimum solution to the MMSE-SIC detector is to replace the filter in (4.14) with the simple MRC filter in (4.18). This requires a scalar inversion only and eliminates the need for matrix inversion. The soft interference cancellation is still performed as above. The only difference is the use of the simple MRC filter at each iteration, except at the first iteration where the MMSE suppression filter is used instead, since no cancellations are being performed. Again, we make a Gaussian approximation for the MRC filter output, $\hat{x}_k \sim \mathcal{N}(\mu_k x_k, \eta_k^2)$. The corresponding mean and variance are re-computed to be

$$\mu_k = \mathbf{w}_k^{\dagger}\mathbf{h}_k \tag{4.23}$$

$$\eta_k^2 = \sum_{i=1, i\neq k}^{n_t} \sigma_{x_i}^2 \|\mathbf{w}_k^{\dagger}\mathbf{h}_i\|^2 + N_0\|\mathbf{w}_k\|^2. \tag{4.24}$$

The MRC-SIC detector offers an additional reduction through the simple MRC filter, which can be evaluated at the beginning of the iteration process. With n transmit and n receive antennas, evaluating the LLR values involves the following steps:

for $k = 1$ to n

for $i = 1$ to 2^{M_c}

1) Compute the *a priori* probability $P(x)$ in (4.9): $2M_c$ flops

end

2) Evaluate the symbol mean $\bar{x}_i$ and variance $\sigma_{x_i}^2$: $5 \cdot 2^{M_c}$ flops

3) Cancel the soft estimates in (4.11): $2n^2$ flops

4) Compute the detection estimate $\hat{x}_k$: $2n$ flops

5) Compute η_k^2: $2n$ flops

for $l = 1$ to 2^{M_c}

6) Evaluate the likelihood function $P(\hat{x}_k|x)$ in (4.22): 4 flops

end

for $l = 1$ to M_c

7) Compute the LLR value in (4.21): $2 \cdot 2^{M_c}$ flops

end

end

It follows that the complexity per turbo iteration is $2n^3+2n^2+(4nM_c)\cdot 2^{M_c}+(9n)\cdot 2^{M_c}$ flops. Moreover, the initial cost of evaluating the MRC and MMSE suppression filters is $7n^4/3+n^2+3n$ flops and the cost of evaluating μ_k and the vector norms appearing in the computation of η_k^2 is $2n^3+2n^2$ flops, yielding a total initial cost of $7n^4/3+2n^3+3n^2+3n$ flops.

4.3.4 MRC-HIC Detector

The MRC-HIC detector is similar in nature to the MRC-SIC detector, except that *hard* decisions are used in the cancellation step. Hard decision estimates on the LDPC code bits b_i ($i=1,\ldots,n_tM_c$) can be obtained from

$$\hat{b}_i = \text{sign}\left[L_D^a(b_i)+L_D^e(b_i)\right], \tag{4.25}$$

and a hard decision estimate on the kth antenna symbol can be formed as

$$\tilde{x}_k = f\left(\hat{b}_{(k-1)M_c+1}\hat{b}_{(k-1)M_c+2}\ldots\hat{b}_{kM_c}\right) \tag{4.26}$$

where f is a function that maps an input bit vector to a complex constellation point. Assuming these hard decision symbol estimates are correct, their cancellation would provide a better detection estimate for the antenna of interest. As in the case of the MRC-SIC detector, the MMSE suppression detector is used in the first iteration and

the simple MRC filter is applied whenever hard decisions are canceled to obtain the detection estimate $\hat{x}_k$.

Observe that the assumption of correct hard decisions does not hold very well in the early stages of the iterative process. In order to avoid error propagation due to incorrect hard decisions, it is very important that cancellations be performed only when the reliability of the canceled symbols is high according to some cancellation criterion. We experimented with different criteria for interference cancellation and found that the following two methods give the best results:

- **Average of LLRs**

  With this criterion, first the following average is computed and then compared to a predetermined threshold value:

$$\Phi_a = \frac{1}{n_t M_c} \sum_{i=1}^{n_t M_c} |L_D^e(b_i)| \geq \Theta_a. \tag{4.27}$$

  The threshold Θ_a is found experimentally as the threshold that yields the best bit-error rate performance. Note that a too low threshold value would introduce undesirable error propagation due to incorrect cancellations, while a too high threshold value would give the same performance as the MMSE suppression detector since no cancellations are performed in this case.

- **Probability of bit vector**

With this criterion, the probability of a bit vector is first computed and then compared to a threshold value:

$$\Phi_p = P(b_1 b_2 \ldots b_{n_t M_c}) = \prod_{i=1}^{n_t M_c} [1 + \exp(-|L_D^e(b_i)|)]^{-1} \geq \Theta_p. \qquad (4.28)$$

As before, the threshold Θ_p is found experimentally to optimize the BER performance.

The computational complexity of the MRC-HIC detector is approximately the same as the complexity of the MRC-SIC detector.

4.3.5 MMSE Suppression Detector

An even further reduction in complexity is obtained with a simple MMSE suppression detector. Since no interference cancellation is performed, the MMSE suppression filter $\mathbf{w}_k$ from (4.17) is evaluated only once, at the beginning of the turbo iterative process, and applied to the received vector $\mathbf{y}$ to obtain a detection estimate $\hat{x}_k$ for the kth antenna.

Assuming an equal number of transmit and receive antennas, evaluating the LLR values involves the following steps:

for $k = 1$ to n

for $i = 1$ to 2^{M_c}

1) Compute the *a priori* probability $P(x)$ in (4.9): $2M_c$ flops

end

for $l = 1$ to M_c

2) Compute the LLR value in (4.21): $2 \cdot 2^{M_c}$ flops

end

end

Then, the complexity per turbo iteration is $(4nM_c) \cdot 2^{M_c}$ flops. Moreover, the initial cost of evaluating the MMSE suppression filters $\mathbf{w}_k$, the detection estimates $\hat{x}_k$, and the likelihood functions $P(\hat{x}_k|x)$ is $7n^4/3 + 5n^2 + (4n) \cdot 2^{M_c}$ flops. Note that the MMSE suppression detector reduces the complexity per turbo iteration from $\mathcal{O}(n^4)$ (for the MMSE-SIC detector) to $\mathcal{O}(n)$. Of course, the overall complexity for the MMSE suppression scheme remains $\mathcal{O}(n^4)$ because of required initial processing. However, a more flexible allocation of computational resources becomes possible given the need to solve a system of equations to determine $\mathbf{w}_k$ only once rather than on a per iteration basis.

In Table 4.1 we give a complexity comparison example based on a flop count for the MAP, MMSE-SIC, MRC-SIC, and MMSE suppression detectors for $n \times n$ antenna configurations with $n = 2, 4, 8$, using QPSK modulation. For each detector, we provide the initial cost (which is the cost of computations that must be done only once for all of the iterations), the cost per turbo iteration, and the total computational cost assuming

	$n = 2$	$n = 4$	$n = 8$
MAP initial	224	1.13×10^4	9.96×10^6
MAP per iteration	256	8.19×10^3	4.19×10^6
MAP total (30 iterations)	7.90×10^3	2.57×10^5	1.36×10^8
MMSE-SIC initial	0	0	0
MMSE-SIC per iteration	225	1.17×10^3	1.21×10^4
MMSE-SIC total (30 iterations)	6.76×10^3	3.52×10^4	3.63×10^5
MRC-SIC initial	71	785	1.08×10^4
MRC-SIC per iteration	160	432	1.7×10^3
MRC-SIC total (30 iterations)	5.58×10^3	1.37×10^4	6.17×10^4
MMSE suppression initial	89	741	1.0×10^4
MMSE suppression per iteration	64	128	256
MMSE suppression total (30 iterations)	2.0×10^3	4.58×10^3	1.77×10^4

Table 4.1: Cost (in flops) of computing the LLRs for different MIMO detectors.

30 iterations.

4.3.6 Belief Propagation Decoding of LDPC Codes

The standard message-passing (also known as belief-propagation) decoding algorithm [35] is used to decode the LDPC code. In the bipartite graph of the LDPC code, the variable nodes are numbered from 1 to n, and the check nodes from 1 to $n - k$.

Denote by $\{e^b_{i,1}, e^b_{i,2}, \ldots, e^b_{i,d_v}\}$ the set of edges connected to the ith variable node and by $\{e^c_{i,1}, e^c_{i,2}, \ldots, e^c_{i,d_c}\}$ the set of edges connected to the ith check node. The message-passing algorithm of LDPC codes can be summarized as follows:

- *Iterate between variable node update and check node update*:
  - *Variable node update*: For each of the variable nodes, for every edge connected to the variable node, compute the extrinsic message passed from the variable node to the check node along the edge

$$L^e_B(e^b_{i,j}) = L^a_D(b_i) + \sum_{k=1,k\neq j}^{d_v} L^e_C(e^b_{i,k}), \tag{4.29}$$

    where L^e_C is set to zero at the first iteration.
  - *Check node update*: For each of the check nodes, for all edges connected to the check node, compute the extrinsic message passed from the check node to the variable node

$$L^e_C(e^c_{i,j}) = 2\tanh^{-1}\left[\prod_{k=1,k\neq j}^{d_c} \tanh\left(\frac{L^e_B(e^c_{i,k})}{2}\right)\right]. \tag{4.30}$$

- *Compute extrinsic messages passed back to the demodulator*:

$$L^e_D(b_i) = \sum_{k=1}^{d_v} L^e_C(e^b_{i,k}). \tag{4.31}$$

4.4 LDPC Codes as Universal Space-Time Codes

4.4.1 Mutual Information Performance under MAP Detection on 2×2 Channels

In this section the performance of two different rate-1/3 length-15,000 spatially-multiplexed LDPC codes modulating QPSK on parameterized 2×2 MIMO channels (for a net rate of 4/3 bits per transmission) using the MAP detector will be described.

Rather than computing average performance over randomly drawn channels, we examine performance on a representative set of specific 2×2 channels. Each of the 2×2 channels will be characterized by three parameters: two rotation parameters and the ratio of the eigenvalues in the system (or eigenskew), specifically,

$$\mathbf{H} = \sqrt{\lambda_1} \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{\kappa} \end{bmatrix} \begin{bmatrix} \cos(\phi) & \sin(\phi)e^{j\theta} \\ -\sin(\phi)e^{-j\theta} & \cos(\phi) \end{bmatrix} \tag{4.32}$$

where $\kappa = \lambda_2/\lambda_1$ is the eigenvalue skew of the Hermitian matrix $\mathbf{H}\mathbf{H}^\dagger$ (with $\lambda_2 \leq \lambda_1$), $\phi \in [0, \pi/2]$, and $\theta \in [0, 2\pi]$. We sample the above matrix via the parameters $\lambda_1 = 1$, $\kappa = \{0, 0.125, 0.25, 0.5, 0.75, 1\}$, $\phi = \{0, \pi/4\}$ and $\theta = \{\pi/4, 3\pi/4\}$.

The value of such a parameterized assessment is that it explicitly shows how excess MI performance varies across the set of channels and in fact allows *worst* and *best* channels for a given codeword to be identified. Furthermore, flatness of the excess mutual information measure versus channel skew becomes a criteria for comparing the

degree of robustness possessed by a given code. Again, in our terminology, a code is universal if reliable communication (for instance BER = 10^{-5}) occurs at the same (small) excess mutual information level across all channels. The absolute SNR required to achieve a given mutual information level will vary with the degree of fading imposed by the channel matrix.

Performance on each channel in the sampling is measured by the mutual information in excess of the transmission rate of $4/3$ bits/channel use required to achieve a BER of 10^{-5}. Note that the product of a diagonal eigenmatrix and a complex Givens rotation (the rightmost matrix in (4.32)) does not parameterize *every* 2×2 matrix, but does parameterize every interesting 2×2 matrix because matrices not described by (4.32) can be found through a symmetric reflection of the eigenmatrix. A simple example of this is that a $\kappa = 0$ channel erases all received symbols in even positions (all y_2 observations). The matrix could be reformulated to erase odd received symbols by swapping elements on the diagonal in the eigenmatrix.

Our LDPC codes were actually designed for periodic fading SISO channels. Note that a period-2 SISO fading channel, $y_k = a_{(k \bmod 2)} x_k + n_k$, with fading vector $\mathbf{a} = [\sqrt{\lambda_1}, \sqrt{\lambda_2}]$ is equivalent to a diagonal ($\phi = 0$) 2×2 matrix fading channel (but requires two channel uses to relay the same information). Following the work in [22], we are able to optimize LDPC degree distributions for period-p fading via an adaptation of the Gaussian approximation to density evolution. Specifically, at iteration l, degree i variable nodes have their mean values updated in correspondence to the periodic ini-

tial means given by $m_{a_j} = \frac{2a_j^2}{\sigma^2}$ (where a_j are known fading gains) and the means of messages arriving from check nodes (m_u):

$$m_{v,i}^{(l)}(j) = m_{a_j} + (i-1)m_u^{(l-1)}, j = \{0, \ldots, p-1\}. \tag{4.33}$$

Then, randomly selected edges emanating from variable nodes adhere to the following Gaussian mixture density,

$$f_v^{(l)} = \sum_{j=0}^{p-1} \sum_{i=2}^{d_l} \frac{\lambda_i}{p} \mathcal{N}\left(m_{v,i}^{(l)}(j), 2m_{v,i}^{(l)}(j)\right). \tag{4.34}$$

Using this kernel, codes were optimized for $\mathbf{a} = [1, 1]$ and $\mathbf{a} = [1, 0]$ period-2 fading channels. These codes were then tested across the parameterization of 2×2 channels. The resulting degree distributions are given in Table 4.2. In general, a code for periodic fading does not have to be a good code for MIMO channels. However, LDPC codes are random enough that our codes, designed for periodic fading, exhibit an exceptional degree of universality on MIMO channels.

4.4.2 Gaussian, Constellation Constrained, and Parallel Independent Decoding Mutual Information

We next describe the mutual information measures ($J(\cdot)$) that can be considered when analyzing the universal property of LDPC codes. Of course, the mutual information between transmitter and receiver using a Gaussian codebook is independent of

i	λ_i [1,1]	λ_i [1,0]	ρ_i [1,1]	ρ_i [1,0]
2	0.27603	0.354954	-	-
3	0.11195	-	-	-
4	0.17229	0.249982	-	-
5	0.01712	0.065503	-	0.5000
6	-	-	1.0	0.5000
15	0.42261	0.329558	-	-

Table 4.2: Degree distributions optimized using Guassian approximation to density evolution adapted to periodic fading. Columns labeled with $[1, 0]$ indicate the distribution resulting from optimization for the period-2 channel where half of all received symbols are erased. Columns labeled with $[1, 1]$ indicate a period-2 code optimized for AWGN.

Givens rotation parameters, θ and ϕ, and is given by:

$$J_{\text{Gauss}} = \log_2 \left(1 + \frac{E_x}{N_0}\lambda_1\right) \left(1 + \frac{E_x}{N_0}\lambda_2\right). \tag{4.35}$$

However, two other capacity measures are relevant to a system that modulates an LDPC code directly onto a channel and these measures do depend on the channel parameters θ and ϕ. The first is the QPSK constellation constrained (J_{QPSK}) capacity:

$$J_{\text{QPSK}} = \mathcal{I}(\mathbf{x}; \mathbf{y}, \mathbf{H}) = \sum_{i=1}^{n_t M_c} \mathcal{I}(\mathbf{x}^{b_i}; \mathbf{y}, \mathbf{H} | \mathbf{x}^{b_1}, \ldots, \mathbf{x}^{b_{i-1}}). \tag{4.36}$$

If the dependencies between the bit levels are neglected for decoding, we obtain the parallel independent decoding constellation constrained ($J_{\text{PID QPSK}}$) capacity [44], which has a lower mutual information:

$$J_{\text{PID QPSK}} = \sum_{i=1}^{n_t M_c} \mathcal{I}(\mathbf{x}^{b_i}; \mathbf{y}, \mathbf{H}) \leq J_{\text{QPSK}}. \tag{4.37}$$

For SISO channels it has been shown [11] that the parallel independent decoding constellation constrained capacity suffers only a negligible loss compared to modulation constrained capacity provided that a Gray labeling of signal points is possible.

Lampe *et al.* in [25] suggested that in fast fading channels, approaches that use multi-level coding (MLC) with multi-stage decoding (MSD) can achieve the constellation constrained capacity of a MIMO channel. In particular, they showed that systems

which applied separate codes on each antenna could approach the modulation constrained ergodic capacity of the fast fading MIMO channel.

Without channel interleaving, MLC/MSD approaches will not perform well under certain parameterizations of the static fading channel described above. For example, a channel with parameterization $\kappa = 0$ and $\phi = 0$ has $y_1 = x_1 + n_1$ and $y_2 = n_2$. In this scenario, the codesymbols from code(s) on the second transmit antenna are completely punctured at the receive array and the information associated with these codewords cannot be recovered. This work focuses on spatially multiplexed systems in which a single code is applied to the entire transmit array. Bit-interleaved coded modulation (BICM) systems such as this have lower complexity as compared to MLC/MSD approaches.

The robustness results for the length 15,000, $\mathbf{a} = [1, 0]$ optimized code applied to a 2×2 channel under spatial multiplexing are given in Fig. 4.4. The curves with circular and square markers show excess mutual information on the $\phi = \pi/4$ (worst case) channel when excess MI is measured from J_{QPSK} capacity (circle) or $J_{\text{PID QPSK}}$ capacity (square) (the raw underlying simulation data for these two curves are the same).

On the worst case channel ($\phi = \pi/4$) when $\kappa = 0$, the excess MI for the code to operate at BER=10^{-5} as measured from J_{QPSK} is 0.67 bits, while the gap to $J_{\text{PID QPSK}}$ is 0.09 bits. The difference between these two capacity measures at this operating SNR is therefore 0.58 bits. Figure 4.5 shows how these two measures differ at this operating SNR not only at $\phi = \pi/4$ (where the difference is 0.58 bits), but across the entire range of interest for this parameter. This figure also provides mutual information plots for the

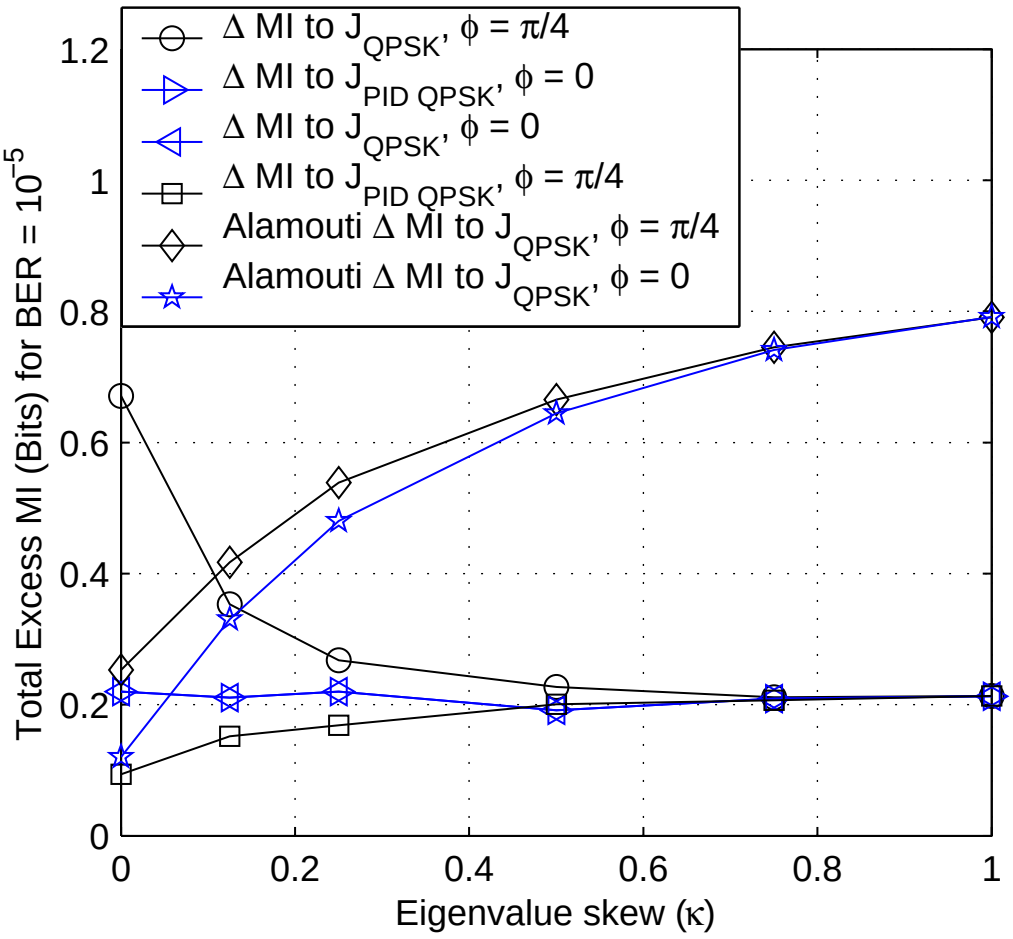


Figure 4.4: Excess mutual information performance of length 15,000, rate-1/3 $\mathbf{a} = [1, 0]$ optimized LDPC code at BER = 10^{-5} as measured from QPSK constellation constrained capacity (J_{QPSK}), and parallel independent decoding constellation constrained capacity ($J_{\text{PID QPSK}}$) across eigenskew and two distinct values of ϕ. Also plotted is the performance of a rate-2/3, length 7,500 LDPC code (of same maximum left degree as the rate-1/3 code) modulating QPSK into 2×2 Alamouti space-time block code.

SNR associated with BER=10^{-5} operation on the $\phi = 0$, $\kappa = 0$ channel (excess MI results for which are provided in Fig. 4.4).

Note that channels with $\phi = \pi/4, 3\pi/4$ maximize J_{QPSK} and minimize $J_{\text{PID QPSK}}$. At these angles, we also observe that the constellation constrained capacity approaches the Gaussian alphabet capacity closely (which is reasonable given the relatively low rate loading of the QPSK constellations). Though not shown in Fig. 4.5, as κ approaches unity, the constellation constrained and parallel independent decoding constellation constrained mutual informations do not drift from each other, nor do they vary absolutely across the parameterization of ϕ because the $\kappa = 1$ unitary channel is a generalization of a diagonal channel where J_{QPSK} and $J_{\text{PID QPSK}}$ are equivalent (within the loss of $J_{\text{PID QPSK}}$ to Gray labeling, which is negligible).

The opposing diamond curves in Fig. 4.4 measure the performance of the code on the $\phi = 0$ channel across κ (the raw underlying simulation data of these two curves are also the same). The flatness of the gaps to J_{QPSK} and $J_{\text{PID QPSK}}$ mutual information is indicative of the robustness of the code on this subset of channels (diagonal channels yield $J_{\text{PID QPSK}}$ that is approximately equal to J_{QPSK}). We note the observation that in all cases where $J_{\text{PID QPSK}}$ is flat across a given parameterization, so is the excess mutual information performance of the code.

We conjecture that a binary coded system (without feedback) can achieve J_{QPSK} capacity measures only in cases where $J_{\text{PID QPSK}}$ approaches J_{QPSK}. All other cases must depend on feedback between detector and decoder to push the system performance

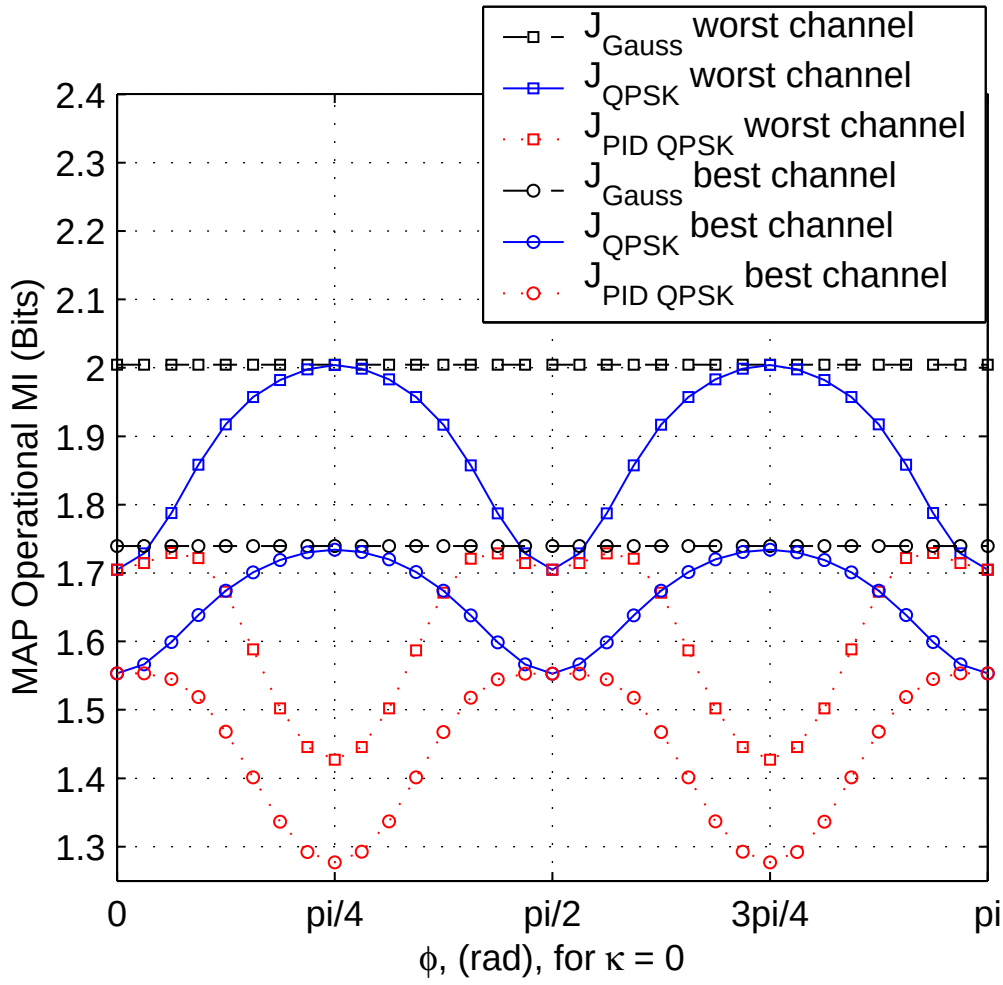


Figure 4.5: Channel mutual information versus channel matrix parameter ϕ and eigenskew. Gaussian input (J_{Gauss}), QPSK constellation constrained (J_{QPSK}), and parallel independent decoding constellation constrained ($J_{\text{PID QPSK}}$) capacities are shown at the SNR levels that yield BER=10^{-5} for $\kappa = 0$ on ϕ equal to 0 (the best) and $\pi/4$ (the worst) channels. The data on this plot can be used to understand the excess MI data presented for the $\kappa = 0$ channels in Fig. 4.4.

closer to the constellation constrained capacity. Even so, all of the results presented in Fig. 4.4 have applied feedback between detector and decoder and the remaining large gaps to J_{QPSK} capacity evidence the ineffectiveness of the simplest feedback approach. The existence of feedback schedules and techniques that cause the operational capacity to closely approach the net constellation constrained capacity of the system is an open research problem. Other solutions may include design of multidimensional mappings to flatten the parallel independent decoding constellation constrained capacity across all channel parameterizations, or symbolwise message passing over GF($2^{n_t M_c}$) so as to avoid bit marginalization of the received vector before decoding.

One might suspect that a better approach to achieving space-time diversity with a binary linear code is to concatenate a relatively high rate code with a space-time block code such as the code devised by Alamouti in [3]. The channel sampling approach provides a detailed basis for the comparison of direct bit-multiplexing and Alamouti concatenation. The diamond and pentagon curves of Fig. 4.4 describe the excess mutual information performance of a rate-2/3, length 7,500 LDPC code concatenated with a 2×2 Alamouti space-time block code. Note that the rate and block length of this code have been chosen so that each frame contains the same number of vector symbols and information bits as in the rate-1/3, length 15,000 bit-multiplexed case. The mutual information difference between these two Alamouti curves is entirely due to the difference in QPSK constellation constrained capacity at angle $\phi = 0$ and angle $\phi = \pi/4$ (the raw simulation data underlying both curves is the same).

The Alamouti concatenation performs well on the $\kappa = 0$ channels since the erasure of one transmit antenna at the receive array is perfectly absorbed by the Alamouti scheme and a complete set of observations remains for each rate 2/3 codeword. The penalty, however, for using an orthogonal space-time code is severe on unitary ($\kappa = 1$) channels. Specifically, the penalty for using Alamouti on a unitary channel is about 0.1 bits in excess of the penalty of using bit-multiplexing on a singular channel.

4.4.3 Mutual Information Performance under MMSE Detection on 2×2 Channels

In this section we examine the MI performance of the spatially-multiplexed LDPC code under MMSE detection by considering the set of 2×2 parameterized matrix channels previously defined.

The matrix parameterization makes the MMSE suppression detector easy to analyze and offers some insight to what can be expected in terms of universal performance with this simple detector. Let γ_1 and γ_2 be the SNRs at the output of the MMSE detector. It is easy to show that we obtain the following formulation for the SNRs at the output of the MMSE filter:

$$\gamma_1 = \mathbf{h}_1^\dagger \left(\frac{N_0}{E_x} \mathbf{I}_2 + \mathbf{h}_2 \mathbf{h}_2^\dagger \right)^{-1} \mathbf{h}_1 \tag{4.38}$$

$$= \frac{\kappa \sin^2(\phi) + \cos^2(\phi) + \kappa E_x / N_0}{N_0 / E_x + \kappa \cos^2(\phi) + \sin^2(\phi)}, \tag{4.39}$$

$$\gamma_2 = \mathbf{h}_2^\dagger \left(\frac{N_0}{E_x}\mathbf{I}_2 + \mathbf{h}_1\mathbf{h}_1^\dagger \right)^{-1} \mathbf{h}_2 \tag{4.40}$$

$$= \frac{\kappa\cos^2(\phi) + \sin^2(\phi) + \kappa E_x/N_0}{N_0/E_x + \kappa\sin^2(\phi) + \cos^2(\phi)}. \tag{4.41}$$

Note that the angle θ does not matter at all and that for unitary channels ($\kappa = 1$) the SNRs reduce to $\gamma_1 = \gamma_2 = E_x/N_0$. In the case of singular channels ($\kappa = 0$), the above SNRs become

$$\gamma_1 = \frac{\cos^2(\phi)}{N_0/E_x + \sin^2(\phi)}, \tag{4.42}$$

$$\gamma_2 = \frac{\sin^2(\phi)}{N_0/E_x + \cos^2(\phi)} \tag{4.43}$$

Observe that on such a channel, when $\cos(\phi) = 0$ or $\sin(\phi) = 0$, we have in effect a scalar period-2 erasure channel. However, when $\sin(\phi) = \cos(\phi)$ (i.e. $\phi = \pi/4$ or a multiple of $\pi/4$), we obtain $\gamma_1 = \gamma_2 = \frac{E_x/N_0}{2+E_x/N_0}$ and a periodic scalar channel of the form $y_1 = \frac{\sqrt{2}}{2}x_1 + \frac{\sqrt{2}}{2}x_2 + n_1$, $y_2 = n_2$. This represents the most challenging channel for our MMSE suppression detector.

In order to better understand the effect of the eigenvalue skew κ and the rotation angle ϕ on the performance with MMSE suppression, we need to consider the operational capacity of the MMSE detector, given by

$$C_{\text{MMSE}} = \log_2(1+\gamma_1)(1+\gamma_2). \tag{4.44}$$

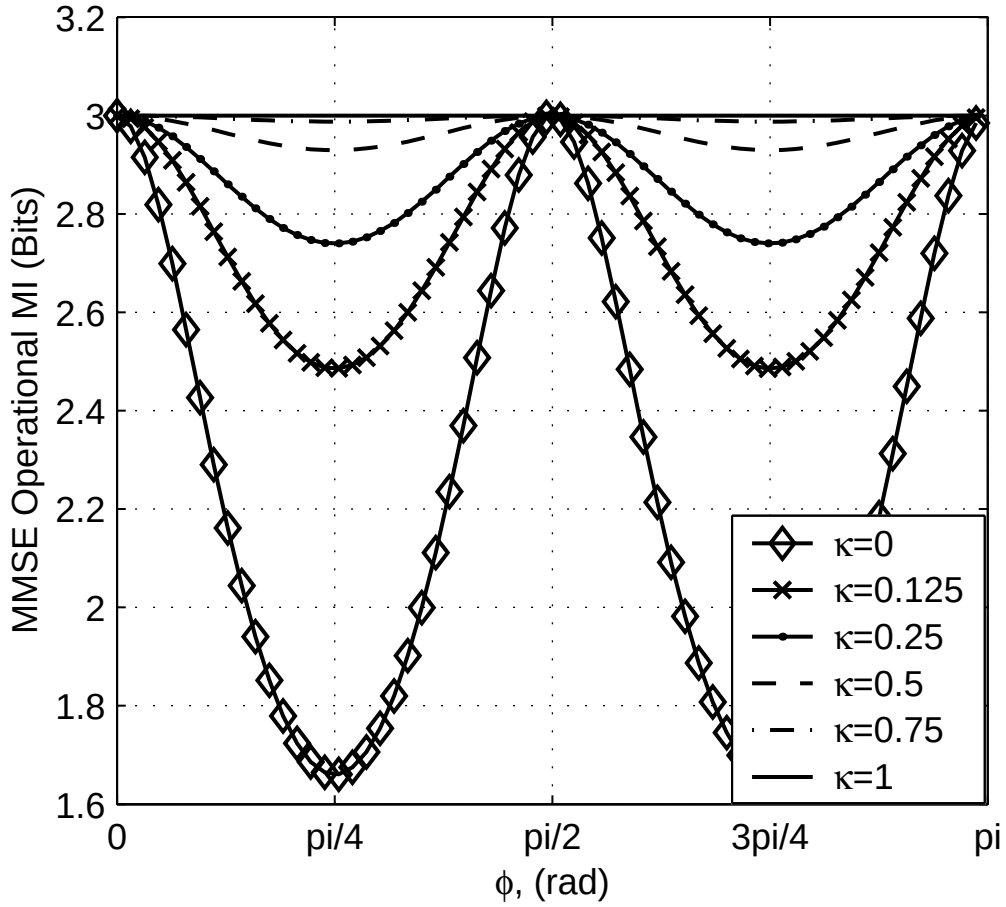


Figure 4.6: MMSE operational mutual information versus channel matrix parameter ϕ for a family of eigenskews κ.

Figure 4.6 shows the operational capacity of the MMSE suppression detector as a function of the angle ϕ for different eigenskews κ. The curves are plotted at $J_{\text{Gauss}} = 3$ bits, where J_{Gauss} is the Gaussian input capacity defined in (4.35). This figure shows that the operational capacity of the MMSE detector can be as low as half the actual Shannon capacity for some particular unfavorable rotation in the channel matrix parametrization. The worst case is $\kappa = 0$ and ϕ a multiple of $\pi/4$. This is exactly the same as the worst channel identified earlier under MAP detection.

Figure 4.7 shows the excess MI results for the rate-1/3, $\mathbf{a} = [1, 1]$ optimized LDPC code using the MAP, MMSE-SIC, and MMSE suppression detectors. For each eigenvalue skew, there are two points for each detector corresponding to the angles $\phi = 0, \phi = \pi/4$, which define the range between "best" and "worst" channels. Note that this range is largest for $\kappa = 0$ and MMSE suppression detector. For all detectors this range diminishes with increasing skew values. The performance loss on the worst channels is attributed to the gap between the operational capacity with the different detectors and the true Shannon capacity, rather than the performance of the LDPC code itself.

4.5 Simulation Results in Rayleigh Fading

In this section we examine the performance of LDPC coded MIMO systems in Rayleigh fading using the soft detectors introduced earlier. In our study, we assume that the number of receive antennas is the same as the number of transmit antennas.

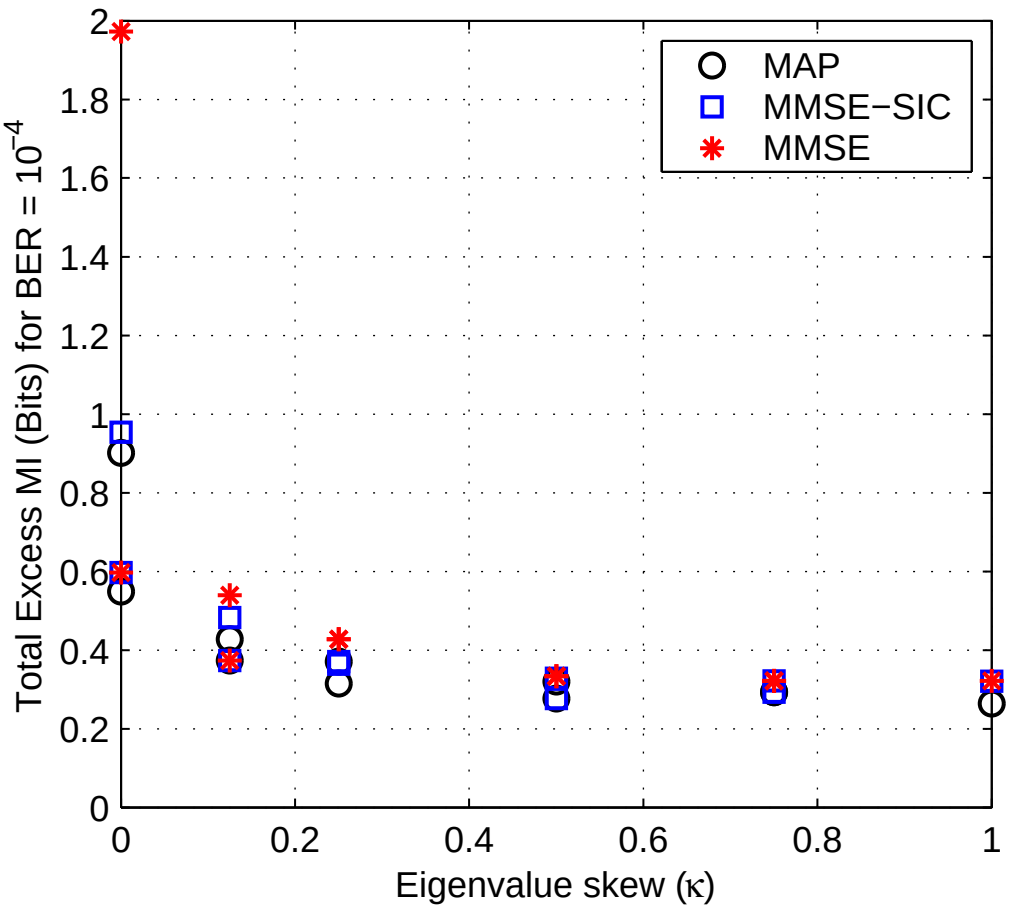


Figure 4.7: Excess MI performance of length 15,000, rate-1/3, $\mathbf{a} = [1, 1]$ optimized LDPC code at BER $= 10^{-4}$ as measured from the Gaussian input capacity across eigenskew and two distinct values of ϕ. The MAP, MMSE-SIC, and MMSE suppression detectors shown for comparison.

We assume a fast Rayleigh fading scenario, where the channel matrix is realized independently from one transmission time to the next. We compare our bit-error results with the theoretical channel capacity limit. Under the fast fading assumption, the theoretical capacity limit is the ergodic channel capacity given by [40]:

$$C = \mathsf{E}\left[\log_2 \det\left(\mathbf{I}_{n_r} + \frac{E_x}{N_0}\mathbf{H}\mathbf{H}^\dagger\right)\right]. \tag{4.45}$$

Figures 4.8 and 4.9 show the BER performance of the rate-1/3, $\mathbf{a} = [1, 0]$ optimized LDPC code versus average SNR per receiver antenna on 2×2, 4×4, and 8×8 fast fading Rayleigh channels. The first plot assumes a QPSK modulated system with resulting spectral efficiencies of $1.33, 2.66$ and 5.33 bits/s/Hz respectively. The second plot assumes 16-QAM modulation, resulting in spectral efficiencies of $2.66, 5.33$ and 10.66 bits/s/Hz respectively. On these plots we also show the channel capacity at the corresponding transmission bit rate for these systems. As expected, the MAP detector yields the best performance, which is 1.5 dB from the theoretical capacity at BER $= 10^{-4}$ on the 2×2 system modulating QPSK and around 2.0 dB on the 2×2 system modulating 16-QAM. The MMSE-SIC detector has essentially the same performance as the MAP detector on the 2×2 and 4×4 systems. For the 4×4 16-QAM and 8×8 QPSK and 16-QAM systems the computational complexity associated with the MAP detector is prohibitive and these result were not simulated.

Our simulation results show that the very low complexity MMSE suppression de-

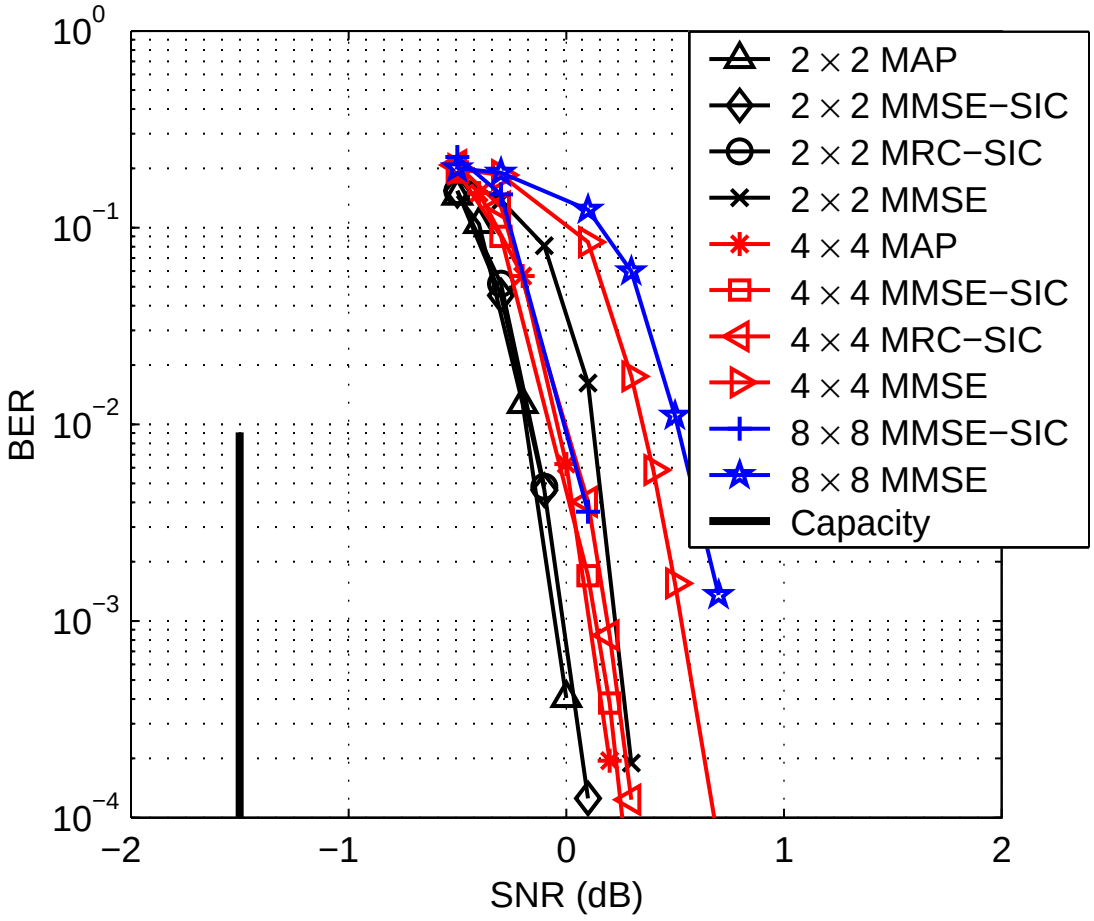


Figure 4.8: Performance of length 15,000, rate-1/3, $\mathbf{a} = [1, 0]$ optimized LDPC code on MIMO systems with MAP, MMSE-SIC, MRC-SIC and MMSE suppression detectors and QPSK modulation.

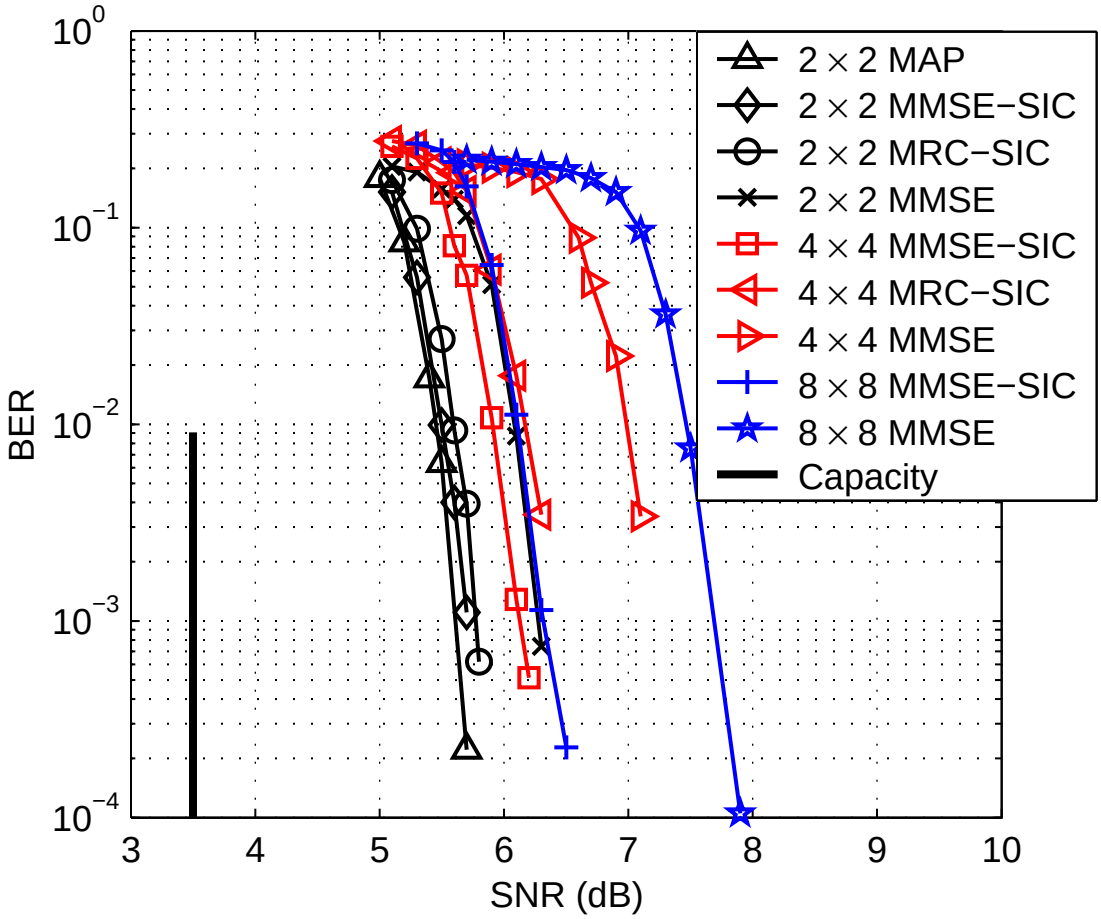


Figure 4.9: Performance of length 15,000, rate-1/3, $\mathbf{a} = [1, 0]$ optimized LDPC code on MIMO systems with MMSE-SIC, MRC-SIC and MMSE suppression detectors and 16-QAM modulation.

tector has a performance loss (with respect to the MMSE-SIC detector) of only 0.5 dB or less under QPSK modulation, but loses around 1.5 dB under 16-QAM modulation and 8×8 channel. The MRC-SIC detector closely follows the performance of the MMSE-SIC detector, with a loss of at most 0.5 dB.

The next set of results are presented in Figs. 4.10, 4.11, and 4.12, where we show the BER performance versus average SNR per receiver antenna on 2×2, 4×4, and 8×8 channels when the $\mathbf{a} = [1, 1]$ optimized LDPC code is used under a QPSK mapping. Again, these simulation results show that the MMSE-SIC detector has essentially the same performance as the MAP detector, and that the very low complexity MMSE suppression detector has a performance loss of only 0.5 dB or less. This loss is approximately cut in half if the MRC-HIC detector with an optimized cancellation threshold is used instead. In fact, in the 2×2 system, the MRC-HIC detector performance is very close to the MAP and MMSE-SIC detectors.

Though the majority of this section has focused on a comparison of the performance of the different detectors in terms of SNR in Rayleigh fast fading, we again emphasize the robustness of the codes as measured in terms of excess mutual information via Fig. 4.13. In this figure we show the excess MI per antenna required by the $\mathbf{a} = [1, 1]$ optimized code modulating QPSK to achieve BER $= 10^{-4}$ with different detectors and different antenna configurations.

The cases plotted in this figure are contrasted from the exhaustively parameterized 2×2 case where we were able to determine best and worst case channels. Here, instead,

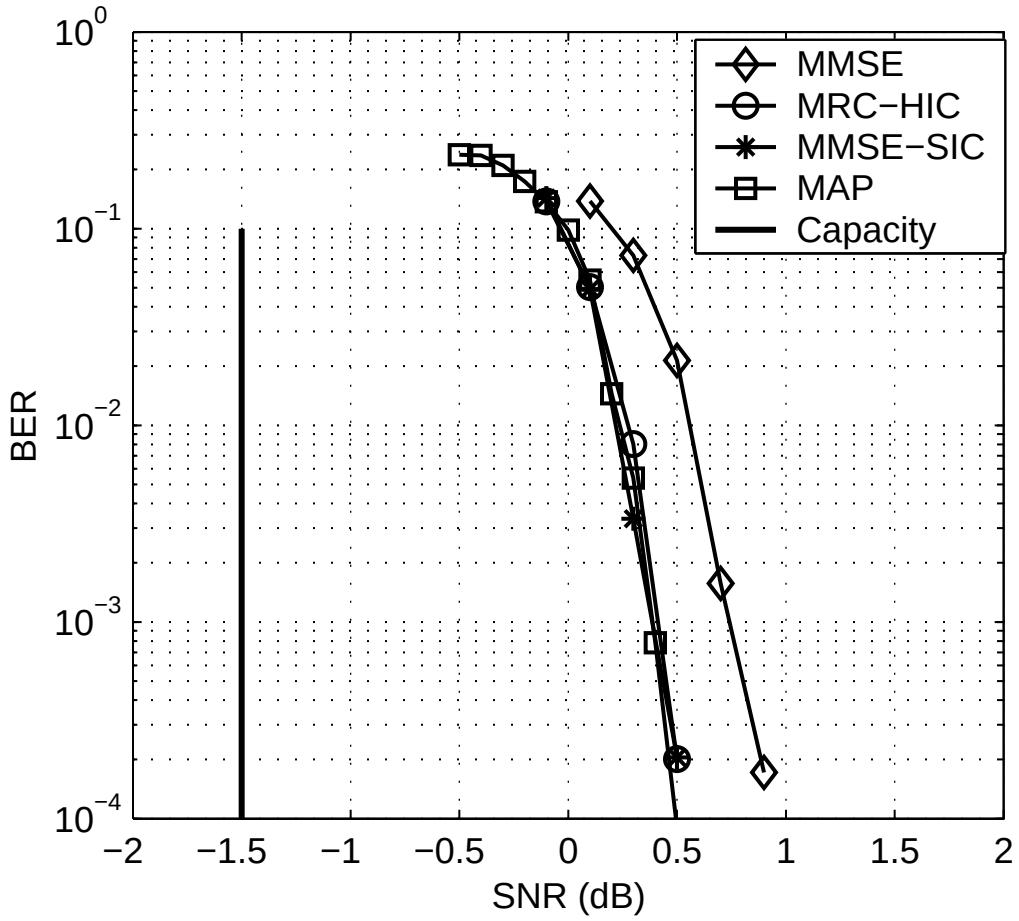


Figure 4.10: Performance of length 15,000, rate-1/3, $\mathbf{a} = [1, 1]$ optimized LDPC code on 2×2 system with MAP, MMSE-SIC, MRC-HIC and MMSE suppression detectors.

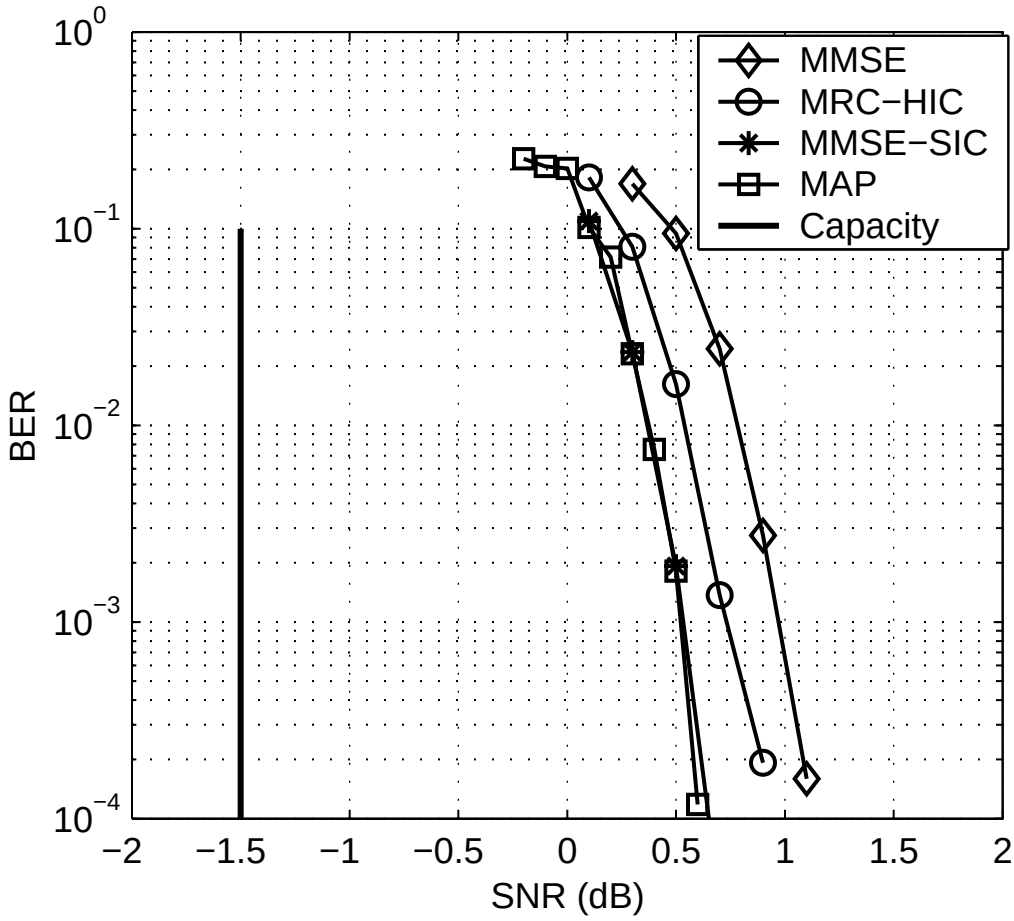


Figure 4.11: Performance of length 15,000, rate-1/3, $\mathbf{a} = [1, 1]$ optimized LDPC code on 4×4 system with MAP, MMSE-SIC, MRC-HIC and MMSE suppression detectors.

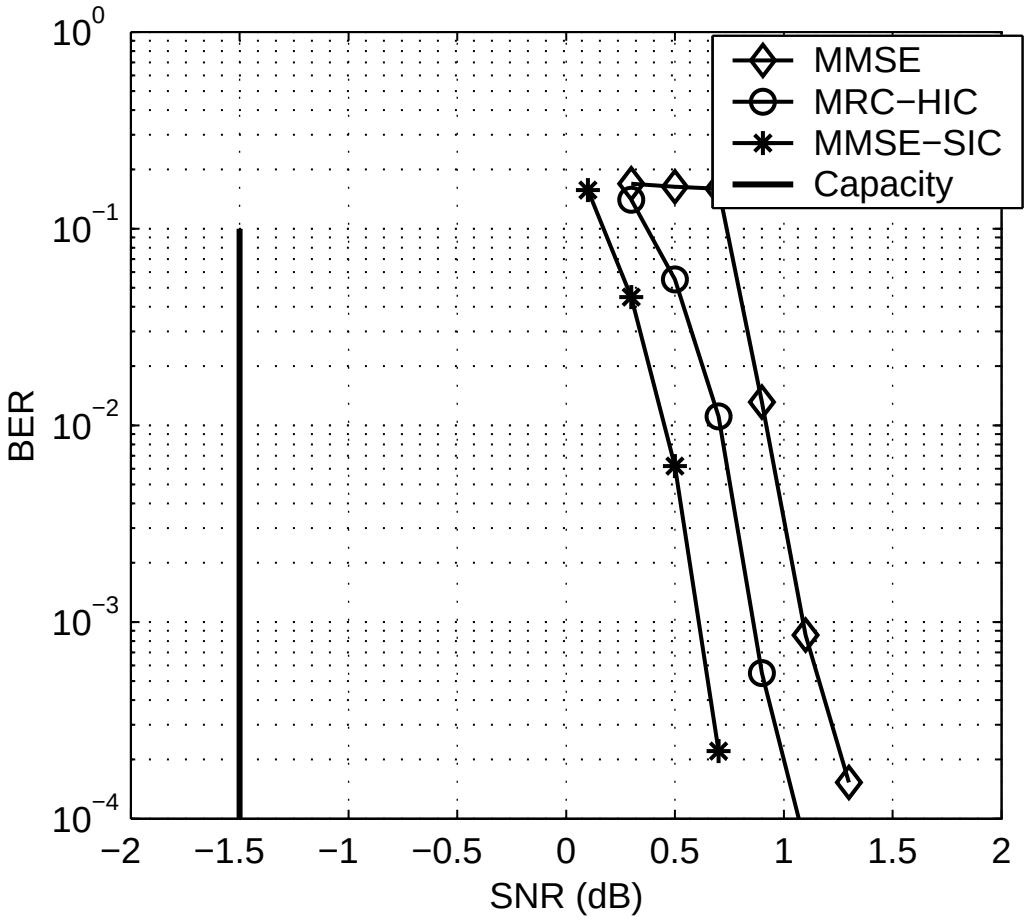


Figure 4.12: Performance of length 15,000, rate-1/3, $\mathbf{a} = [1, 1]$ optimized LDPC code on 8×8 system with MMSE-SIC, MRC-HIC and MMSE suppression detectors.

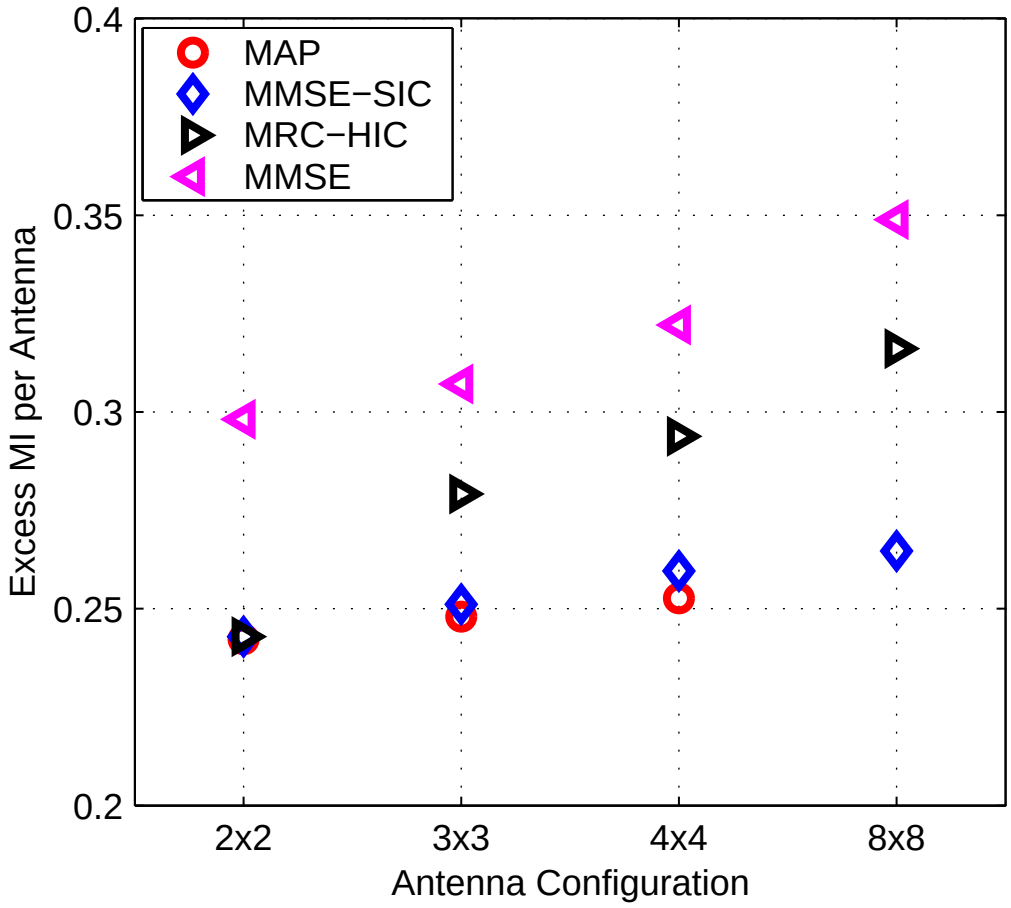


Figure 4.13: Performance of different detectors across increasing antenna multiplicities in terms of excess mutual information per transmit antenna in Rayleigh fading. Each excess MI is measured from QPSK constellation constrained ergodic Rayleigh capacity for the given channel.

all channel results are average together in conjunction with the Rayleigh distribution. Note that under MAP and MMSE-SIC detection, the excess MI per antenna remains essentially constant, at 0.25 bits, as the number of transmit/receive antennas increases from 2 to 4.

4.6 Conclusion

In this chapter we have introduced several soft reduced-complexity detection schemes suitable for systems that achieve high spectral efficiencies. The soft MAP detector is optimal, however the MMSE-SIC detector exhibits similar performance and has a lower overall computational complexity. The MMSE suppression, MRC-SIC, and MRC-HIC detectors offer lower complexities for relatively small performance penalties. As an example, for the 4×4 system, considering the MAP detector as the norm, the total complexity (with 30 iterations) of the MMSE-SIC detector is 7 times lower, that of the MRC-SIC detector is 19 time lower, and that of the MMSE suppression detector is 56 times lower. A performance loss of around 0.5 dB is observed with the MMSE suppression detector and QPSK, and a larger loss (1.5 dB) with 16-QAM.

We have also shown that the excess mutual information required to achieve a bit-error rate of 10^{-5} on the 2×2 channel with MAP detection and our best LDPC code (the rate-1/3, $\mathbf{a} = [1, 0]$ optimized code) varies by approximately 0.5 bits between the best and worst case channel when excess MI is measured against net QPSK constella-

tion constrained capacity. However, excess MI is essentially constant when measured against parallel independent constellation constrained capacity. Moreover, the excess MI measure with respect to the Gaussian input capacity of our $\mathbf{a} = [1, 1]$ optimized LDPC code with the simple MMSE suppression detector varies more significantly on channels where the eigenskew approaches zero due to the loss in operational mutual information of the MMSE detector. In other words, we observe that code performance closely tracks available operating mutual information. Therefore, LDPC codes themselves are robust to channel variation. The information loss incurred during the detection process is the root cause of excess mutual information variation with eigenskew.

Chapter 5

Conclusions

In this dissertation we presented several coding techniques for transmission over wireless multiple-input multiple-output channels.

In Chapter 2 we introduced a simple coding technique for D-BLAST that uses a single trellis code with finite-traceback Viterbi decoding. We examined the performance of universal trellis codes designed to have a distance structure that is matched to the periodic signal-to-noise ratio variation of the channel created by D-BLAST. We showed that a universal 64-state trellis code on a 2×2 D-BLAST system with long enough blocklengths displays universal behavior working on almost every 2×2 channel with at least the mutual information required by a standard 64-state AWGN trellis code. The only 2×2 channel where more mutual information is required is a certain rotation of the zero eigenvalue channel.

We introduced *full-overhead* and *reduced-overhead* versions of the trellis coded D-

BLAST system together with layering methods for both an optimal "just-in-time" and a "not just-in-time" finite traceback Viterbi decoding. We presented performance results for several trellis coded D-BLAST systems and showed that their frame-error rates are within 1.8 - 3.6 dB of the quasistatic Rayleigh fading outage capacity. An additional loss of 0.6 - 1.0 dB is incurred due to the overhead penalty associated with the diagonal layering.

In Chapter 3 we proved that D-BLAST under the MMSE criterion is optimal in the sense that it achieves the Shannon capacity for MIMO channels. We also showed that under a constrained input scenario and a Gaussian approximation on the MMSE filter output, the parallel channels created by D-BLAST have an aggregate capacity that is approximately equal to the MIMO constellation constrained capacity. The D-BLAST technique is useful in reducing the computational complexity for the constellation constrained capacity from exponential to linear in the number of antennas.

In Chapter 4 we introduced a scheme that uses a single LDPC code spatially multiplexed on multiple antennas and an iterative detection and decoding receiver. We presented several variants of LDPC coded MIMO systems with reduced-complexity detectors that perform roughly 2 - 4 dB from the theoretical capacity of the Rayleigh MIMO channel. The MMSE-SIC detector exhibits similar performance as the optimal MAP detector and has a lower overall computational complexity. The MMSE suppression, MRC-SIC, and MRC-HIC detectors offer yet lower complexities, for relatively small performance penalties.

We showed that the excess mutual information required for reliable communication varies more significantly with the matrix channel parameter ϕ as the eigenskew approaches zero. This variation is most severe for the MMSE suppression detector. We attribute the loss in performance on the zero eigenskew channels with particular unfavorable rotations to the gap between the operational capacity with the different detectors and the true Shannon capacity, rather than the performance of the LDPC code itself. We conclude that properly designed LDPC codes behave very much like universal codes in the sense that their performance lies in close proximity to the compound channel capacity for the linear Gaussian vector channels for all but a few channels.

Bibliography

[1] V. Tarokh A., Naguib, N. Seshadri, and A. R. Calderbank. Combined array processing and space-time coding. *IEEE Trans. Inform. Theory*, 45(4):1121–1128, May 1999.

[2] A. Paulraj, R. Nabar, and D. Gore. *Introduction to Space-Time Wireless Communications*. Cambridge Univ. Press. To be published.

[3] S.M. Alamouti. A simple transmit diversity technique for wireless communications. *IEEE Journal on Selected Areas in Communications*, 16:1451–1458, 1998.

[4] P. Almers, F. Tufvesson, and A. F. Molisch. Measurement of keyhole effect in a wireless multiple-input multiple-output (MIMO) channel. *IEEE Commun. Lett.*, 7(8):373–375, Aug. 2003.

[5] A. A. AlRustamani, A. Stefanov, and B. R. Vojcic. Turbo-greedy coding for multiple antenna systems. In *Proceedings of ICC*, volume 6, pages 1684 –1689, June 2001.

[6] J. B. Anderson and K. Balachandran. Decision depths of convolutional codes. *IEEE Trans. Inform. Theory*, 35:455–459, Mar. 1989.

[7] S. L. Ariyavisitakul. Turbo space-time processing to improve wireless channel capacity. *IEEE Trans. Commun.*, 48(8):1347–1359, Aug. 2000.

[8] C. Berrou, A. Glavieux, and P. Thitimajshima. Near Shannon limit error-correcting coding and decoding: Turbo codes. in *Proc. IEEE Int. Conf. Commun.*, Geneva, Switzerland, pages 1064–1070, May 1993.

[9] G. Caire and G. Colavolpe. On space-time coding for quasi-static multiple-antenna channels. In *Proc. Globecom'01*, Nov. 2001.

[10] G. Caire and G. Colavolpe. On low-complexity space-time coding for quasi-static channels. *IEEE Trans. Inform. Theory,* 49(6):1400–1416, June 2003.

[11] G. Caire, G. Taricco, and E. Biglieri. Bit-interleaved coded modulation. *IEEE Trans. on Inform. Theory*, 44:927–946, May 1998.

[12] G. J. Foschini. Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas. *Bell Labs Technical Journal*, 1(2):41–59, Autumn 1996.

[13] G. J. Foschini and M. J. Gans. On limits of wireless communications in a fading environment when using multiple antennas. *Wireless Personal Commun.*, 6(3):311–335, Mar. 1998.

[14] C. Fragouli, C. Komninakis, and R. D. Wesel. Minimality for punctured convolutional codes. In *Proc. IEEE ICC'01*, June 2001.

[15] G. H. Golub and C. F. van Loan. *Matrix Computations*. Baltimore, MD: Johns Hopkins Univ. Press, third edition, 1996.

[16] R. G. Gallager. Low-density parity-check codes. *IRE Trans. Inform. Theory*, IT-8:21–28, Jan. 1962.

[17] H. El Gamal and A. R. Hammons. A new approach to layered space-time coding and signal processing. *IEEE Trans. Inform. Theory*, 47(6):2321–2334, Sep. 2001.

[18] G. Ginis and J. M. Cioffi. On the relation between V-BLAST and the GDFE. *IEEE Commun. Lett.*, 5(9):364–366, Sep. 2001.

[19] H. Harashima and H. Miyakawa. Matched-transmission technique for channels with intersymbol interference. *IEEE Trans. Commun.*, 20(4):774–780, Aug. 1972.

[20] C. Jones, A. Matache, T. Tian, J. Villasenor, and R. Wesel. The universality of LDPC codes on wireless channels. In *Proceedings of MilCom*, Oct 2003.

[21] C. Jones, R. Tian, A. Matache, R. Wesel, and J. Villasenor. Robustness of LDPC codes on periodic fading channels. In *Proceedings of GlobeCom*, Nov 2002.

[22] C. Jones, R. Tian, R. Wesel, and J. Villasenor. The Universal Operation of LDPC Codes in Scalar Fading Channels. *In Submission to the IEEE Transactions on Communications*, Oct 2003.

[23] R. Knopp and P. Humblet. On coding for block fading channels. *IEEE Trans. Inform. Theory*, 46(1):189–205, Jan. 2000.

[24] C. Köse and R. D. Wesel. Universal space-time trellis codes. *IEEE Trans. Inform. Theory Special Issue on MIMO Systems*, 49(10):2717–2727, Oct. 2003.

[25] L. Lampe, R. Schober, and R. Fischer. Multilevel coding for multiple-antenna transmission. In *Proceedings of 2002 International Symposium on Information Theory*, volume 1, July 2002.

[26] Y. Liu, M. P. Fitz, and O. Y. Takeshita. Full rate space-time turbo codes. *IEEE JSAC*, 19(5):969–980, May 2001.

[27] B. Lu, G. Yue, and X. Wang. Performance analysis and design optimization of LDPC coded MIMO OFDM systems. *To appear in EURASIP JASP*, 51(11).

[28] B. Lu, G. Yue, and X. Wang. Performance analysis and design optimization of LDPC coded MIMO OFDM systems. *IEEE Trans. Signal Proc.*, 52(2):348–361, Feb. 2004.

[29] M. Luby, M. Mitzenmacher, A. Shokrollahi, and D. Spielman. Improved low-density parity-check codes using irregular graphs. *IEEE Trans. Inform. Theory*, 47:585–598, Feb. 2001.

[30] D. J. C. MacKay. Good error-correcting codes based on very sparse matrices. *IEEE Trans. Inform. Theory*, 45:399–431, Mar. 1999.

[31] U. Madhow and M. L. Honig. MMSE interference suppression for direct-sequence spread-spectrum CDMA. *IEEE Trans. Commun.*, 42(12):3178–3188, Dec. 1994.

[32] E. Malkamaki and H. Leib. Evaluating the performance of convolutional codes over block fading channels. *IEEE Trans. Inform. Theory*, 45(5):1643–1646, July 1999.

[33] A. Matache, R. D. Wesel, and J. Shi. Trellis coding for diagonally layered space-time systems. In *35th Asilomar Conference on Signals, Systems and Computers*, Nov. 2001.

[34] T. Richardson, M. Shokrollahi, and R. Urbanke. Design of capacity-approaching irregular low-density parity-check codes. *IEEE Trans. Inform. Theory*, 47:638–656, Feb. 2001.

[35] T. Richardson and R. Urbanke. The capacity of low-density parity check codes under message-passing decoding. *IEEE Trans. on Inform. Theory*, 47:599–618, Feb 2001.

[36] W. L. Root and P. P. Varaiya. Capacity of classes of Gaussian channels. *SIAM J. of Applied Math*, 16(6):1350–1393, Nov. 1968.

[37] M. Sellathurai and S. Haykin. TURBO-BLAST for high speed wireless communications. In *Proceedings of WCNC*, volume 1, pages 315–320, Sept 2000.

[38] M. Sellathurai and S. Haykin. Turbo-BLAST for wireless communications: theory and experiments. *IEEE Trans. Signal Proc.*, 50(10):2538–2546, Oct 2002.

[39] V. Tarokh, N. Seshadri, and A. R. Calderbank. Space-time codes for high data rate wireless communication: Performance criterion and code construction. *IEEE Trans. Inform. Theory*, 44(2):744–765, Mar. 1998.

[40] I. E. Telatar. Capacity of multi-antenna Gaussian channels. *European Trans. Telecommun.*, 10(6):585–595, Nov.-Dec. 1999. Available: `http://mars.bell-labs.com/cm/ms/what/mars/papers/proof/`.

[41] S. Ten Brink and B. Hochwald. Achieving near-capacity on a multiple-antenna channel. *IEEE Trans. on Comm.*, 51(3):389–399, Mar. 2003.

[42] M. Tomlinson. New automatic equalizer employing modulo arithmetic. *Electronic Letters*, 7:138–139, March 1971.

[43] A. van Zelst, R. van Nee, and G. A. Awater. Turbo-BLAST and its performance. In *Proceedings of VTC*, volume 2, pages 1282–1286, May 2001.

[44] U. Wachsmann, R. Fischer, and J. Huber. Multi-Level Codes: Theoretical Concepts and Practical Design Rules. *IEEE Trans. on Inform. Theory*, 45:1361–1391, May 1999.

[45] X. Wang and H. V. Poor. Iterative (Turbo) soft interference cancellation and decoding for coded CDMA . *IEEE Trans. on Comm.*, 47:1046–1061, July 1999.

[46] R. D. Wesel. Convolutional codes. *Wiley Encyclopedia of Telecommunications*, Dec. 2002.

[47] R. D. Wesel and J. M. Cioffi. Trellis code design for periodic interleavers. *IEEE Commun. Lett.*, 3(4):103–105, Apr. 1999.

[48] R. D. Wesel, X. Liu, and W. Shi. Trellis codes for periodic erasures. *IEEE Trans. Commun.*, 48(6):938–947, June 2000.