

UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Link adaptation schemes for MIMO wireless systems

Permalink

<https://escholarship.org/uc/item/0163g66w>

Author

Zhang, Yan

Publication Date

2005

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Link Adaptation Schemes for MIMO Wireless Systems

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Electrical Engineering
(Communication Theory and Systems)

by

Yan Zhang

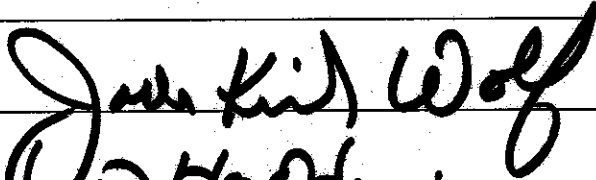
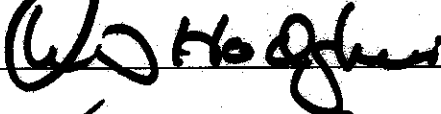
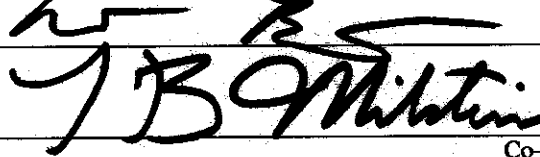
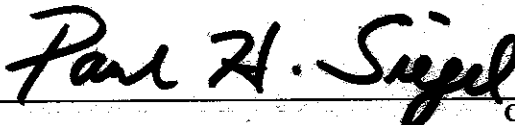
Committee in charge:

Professor Paul H. Siegel, Co-Chair
Professor Laurence B. Milstein, Co-Chair
Professor Robert Bitmead
Professor William S. Hodgkiss
Professor Jack K. Wolf

2005

Copyright
Yan Zhang, 2005
All rights reserved.

The dissertation of Yan Zhang is approved, and it is acceptable in quality and form for publication on microfilm:




Co-Chair

Co-Chair

University of California, San Diego

2005

To my parents

CONTENTS

	Signature Page	iii
	Dedication	iv
	Contents	v
	List of Figures	viii
	List of Tables	ix
	Acknowledgements	x
	Vita and Publications	xi
	Abstract of the Dissertation	xii
Chapter 1	Introduction	1
	1.1 Background	1
	1.2 Outline of the Dissertation	2
Chapter 2	The Tradeoff between Diversity Gain and Interference Suppression in a MIMO MC-CDMA System	5
	2.1 Introduction	5
	2.2 System Model	8
	2.2.1 Transmitter	8
	2.2.2 Channel Model	9
	2.3 Performance Analysis	10
	2.3.1 Output of the m th Correlator	10
	2.3.2 Output of the Adaptive Beamformer	12
	2.3.3 Maximal Ratio Transmission and Adaptive Beamforming Reception	14
	2.4 Numerical Results and Discussions	19
	2.5 Conclusion	25
	2.6 Acknowledgements	25
	Appendix 2.A The Computational Complexity Analysis for the Optimum and Suboptimum Algorithms	26
Chapter 3	Approaching V-BLAST Capacity with Adaptive Modulation and LDPC En- coding	27
	3.1 Introduction	27
	3.2 System Model	30
	3.2.1 Per-antenna-coded V-BLAST	30

3.2.2	Channel Model	30
3.2.3	Optimum Successive Detection	31
3.2.4	Adaptive Modulation	32
3.3	LDPC Code Design	33
3.3.1	Identically Independent Distributed (i.i.d) Channel Adapters	33
3.3.2	Density Evolution	33
3.4	PIC-MMSE Demodulation	37
3.4.1	Parallel Soft Interference Cancellation	38
3.4.2	MMSE Linear Filtering	38
3.5	Rate-compatible Puncturing of LDPC Codes	41
3.6	Outage Capacity of V-BLAST	44
3.6.1	Outage Capacity	44
3.6.2	LDPC Code Design for Non-ergodic Fading Channels	46
3.7	Conclusion	46
3.8	Acknowledgement	48
Chapter 4	Flexible-length Rate-compatible Punctured Irregular Repeat-accumulate Code Design	49
4.1	Introduction	49
4.2	Motivation	50
4.2.1	Limitations of Conventional Punctured LDPC Codes	51
4.2.2	Motivation for Using IRA Codes	52
4.3	Background on IRA Codes	53
4.3.1	Definition of IRA Codes	53
4.3.2	Density Evolution	55
4.4	Flexible-length Rate-compatible Punctured IRA Code Design	57
4.4.1	Analysis of Shortening IRA Codes	57
4.4.1.1	Grouping Check Nodes and Parity Nodes	57
4.4.1.2	Probability Density Functions of Messages from Information/Parity Nodes to Check Nodes	59
4.4.1.3	Derivations of the Recursive Equations for the Updated Mean Values	62
4.4.1.4	Optimization of Shortening Proportions $\left\{ \delta_l^{(0)} \right\}_{l=1}^{2\alpha}$	63
4.4.2	Analysis of Puncturing IRA Codes	64
4.4.2.1	Probability Density Functions of Messages from Information/Parity Nodes to Check Nodes	65
4.4.2.2	Derivations of the Recursive Equations for the Updated Mean Values	68
4.4.2.3	Optimization of Puncturing Proportions $\left\{ \pi_{i,j}^{(0)} \right\}_{j=2}^{d_i}$ and $\left\{ \pi_{p,l}^{(0)} \right\}_{l=1}^{2\alpha}$	70
4.5	Numerical Results	72
4.6	Conclusion	75

4.7 Acknowledgement	76
Appendix 4.A Derivation of ρ_j, ζ_j, ν_j and α_j	77
Appendix 4.B Derivation of $\psi_{p,l}^{(k-1)}$ and Proof of $\hat{\epsilon}_{p,j}^{(k)} = \epsilon_{p,j}^{(k)}$	79
Appendix 4.C Notation List	81
Chapter 5 Conclusion	83
Bibliography	86

LIST OF FIGURES

2.1	Transceiver with adaptive beamforming in a MC-CDMA system	6
2.2	Performance comparison of optimum algorithm and suboptimum algorithm for a $M = 2$, $L_T = 2$ and $L_R = 4$ system with varying number of interfering users	17
2.3	Performance comparison of optimum algorithm and suboptimum algorithm for a $M = 2$, $L_T = 1$ and $L_R = 8$ system with varying number of interfering users	18
2.4	Performance comparison of optimum algorithm and suboptimum algorithm for a $M = 2$, $L_T = 2$ and $L_R = 4$ with different correlations among antennas	18
2.5	BER versus E_b/η_0 for $K = 30$	20
2.6	BER versus E_b/η_0 for $K = 50$	20
2.7	BER versus E_b/η_0 for $K = 30$ and $L_T = 2$, $L_R = 4$ with varying correlations between receive antennas	22
2.8	BER versus E_b/η_0 for $K = 30$ and $L_T = 1$, $L_R = 8$ with varying correlations between receive antennas	23
2.9	BER versus E_b/η_0 for $K = 30$ and $N_s = 32$ per subcarrier with varying number of receive antennas	24
2.10	BER versus E_b/η_0 for $K = 50$ and $N_s = 32$ per subcarrier with varying number of receive antennas	24
3.1	Adaptively modulated LDPC encoded V-BLAST type system	30
3.2	MMSE decision feedback detector	31
3.3	Comparisons of the actual pdf to the approximated Gaussian pdf	35
3.4	Parallel interference cancellation and MMSE filtering demapper	37
3.5	Comparisons of the decoding convergence rate of Tx1 using modified PIC algorithm to that using the original PIC algorithm	39
3.6	Comparisons of the decoding convergence rate of Tx2 using modified PIC algorithm to that using the original PIC algorithm	40
3.7	Comparisons of the ergodic capacity to the maximal achievable rates of a 2×2 V-BLAST system with optimally designed LDPC codes and to that with rate-compatible punctured LDPC codes	43
3.8	BER performance of a 2×2 V-BLAST system with optimally designed LDPC codes and that with rate-compatible punctured LDPC codes	47
4.1	Adaptively modulated per-antenna-coded V-BLAST type system	51
4.2	Tanner graph for IRA codes with parameters $(f_1, \dots, f_{d_l}; a)$	54
4.3	Message flows between information (parity) and check nodes	58
4.4	Sum rate of an adaptively modulated 2×2 V-BLAST system with rate-compatible punctured IRA codes	74
4.5	BER performance of an adaptively modulated 2×2 V-BLAST system with rate-compatible punctured IRA codes	75
4.6	Grouping check nodes and parity nodes	79

LIST OF TABLES

3.1	Degree distribution of a rate = 0.25 LDPC code	42
3.2	Puncturing proportions of rate-compatible punctured LDPC codes	42
3.3	Degree distribution of a rate = 0.20 LDPC code	42
3.4	Puncturing proportions of rate-compatible punctured LDPC codes	42
3.5	Ergodic capacities and maximal achievable rates of a 2×2 V-BLAST system using optimally designed LDPC codes	43
3.6	Ergodic capacities and maximal achievable rates of a 2×2 V-BLAST system using punctured LDPC codes	43
3.7	Outage capacities and maximal achievable rates of a 2×2 V-BLAST system with quasi-static fading	46
4.1	Distribution of expunged information bits for codes of rate 0.25 and 0.35 using rate 0.2 mother code	72
4.2	Shortening proportion of parity bits for codes of rate 0.25 and 0.35 using rate 0.2 mother code	72
4.3	Puncturing proportion of information bits for codes of rate 0.25 and 0.35 using rate 0.2 mother code	73
4.4	Puncturing proportion of parity bits for codes of rate 0.25 and 0.35 using rate 0.2 mother code	73
4.5	Capacities and maximal achievable rates of a 2×2 V-BLAST system using rate-compatible punctured IRA codes	73

ACKNOWLEDGEMENTS

I would like to take this opportunity to thank all those people who have influenced my life and have been kind enough to share their precious resources with me.

First and foremost, I am deeply indebted to my advisors: Professor Paul Siegel and Professor Larry Milstein for guiding me throughout this dissertation work during the past few years. I have immensely benefitted from their keen insight, rigorous approach, unending patience, and constant support. I would also thank them for giving me the freedom to pursue my research topics that I have the most interest in.

I would sincerely like to thank Professor Jack Wolf for exposing me to his enthusiasm for this field which has kept me motivated all throughout. I would also like to thank my other committee members, Professors William Hodgkiss and Robert Bitmead, for reviewing my dissertation and providing valuable feedback.

I would like to thank my labmates at the STARlab. I am grateful to Joseph Soriaga for many fruitful discussions on channel codes design. Thanks to Jilei Hou for sharing his expertise on LDPC codes. Also, thanks to Michael Cheng, Henry Pfister, Li Zhu, Brian Kurkoski, Kai Tang, Hugo Tullberg, Junsheng Han, Patrick Amihood, Prasad Shamain, Qinghua Zhao, Mohammad Taghavi, Sharon Aviran, Zheng Wu, Zeinab Taghavi, Amir Hadi Djahanshahi and Seyhan Karakulak.

I would like to dedicate this dissertation to my family. I owe my deepest gratitude to my parents for everything they have done for me. My brother has always supported and encouraged me in every aspect of my life.

This work was supported by the Center for Wireless Communications at UCSD, by the UC Discovery Grant Program and by Intersil/Conexant.

This dissertation, is in part a reprint of the material in the following papers: “The tradeoff between diversity gain and interference suppression in a MIMO MC-CDMA system,” Yan Zhang, Laurence B. Milstein, and Paul H. Siegel, *IEEE Trans. Commun.*, vol. 53, no. 4, pp 623-631, Apr. 2005, “Approaching V-BLAST capacity with adaptive modulation and LDPC encoding,” Yan Zhang, Paul H. Siegel, and Laurence B. Milstein, submitted to *IEEE Transactions on Communications*. The dissertation author was the primary author of these papers.

VITA

October	Born, Nanjing, China
1996	B.S.E.E., Southeast University, Nanjing, China
1999	M.S.E.E, Southeast University, Nanjing, China
2006	Ph.D., Electrical Engineering (Communications Theory and Systems), University of California, San Diego

PUBLICATIONS

Yan Zhang, Laurence B. Milstein, and Paul H. Siegel, "The tradeoff between diversity gain and interference suppression in a MIMO MC-CDMA system," in *Proceedings of IEEE 2003 Military Communications Conference*, Boston, MD, USA, vol 2., pp. 1132-1137, Oct. 2003.

Yan Zhang, L. B. Milstein, and P. H. Siegel, "The tradeoff between diversity gain and interference suppression in a MIMO MC-CDMA system," *IEEE Trans. Commun.*, vol. 53, no. 4, pp 623-631, Apr., 2005

Yan Zhang, Paul H. Siegel, and Laurence B. Milstein, "Approaching V-BLAST capacity with adaptive modulation and LDPC encoding," Submitted to *IEEE Transactions on Communications*.

ABSTRACT OF THE DISSERTATION

Link Adaptation Schemes for MIMO Wireless Systems

by

Yan Zhang

Doctor of Philosophy in Electrical Engineering
(Communication Theory and Systems)

University of California San Diego, 2005

Professor Paul H. Siegel, Co-Chair
Professor Laurence B. Milstein, Co-Chair

Multiple input multiple output (MIMO) systems can significantly increase channel capacity, especially in power, bandwidth, or complexity limited systems. Link adaptation techniques, where signal transmission parameters such as the modulation and coding rate are dynamically adapted to the changing channel states, also are powerful tools for increasing the system spectral efficiency. Hence, adaptive MIMO systems are emerging as one of the key techniques for easing the bottleneck of traffic capacity in future Internet-intensive wireless networks.

We first examine the traditional role of multiple antennas in a CDMA uplink, combating fading and reducing multiple access interference. Specifically, we assume the fades of the antennas in the receive array are correlated, which reduces the diversity gain against fading without affecting the array's capability for interference suppression. Assuming perfect channel knowledge available at the transmitter, Maximal Ratio Transmission (MRT) is employed to weight the transmitted signal optimally in terms of combating signal fading. At the receiver, adaptive beamforming reception is adopted to both suppress MAI and combat the fading. We evaluate the antenna array performance with joint fading reduction and MAI suppression.

Among numerous space-time coding techniques, the Vertical Bell-Laboratories Layered Space-Time (V-BLAST) scheme has been adopted for 4G systems since it exhibits the best tradeoff between performance and complexity. Here we present a practical implementation of a V-BLAST type system, in which the MIMO open-loop capacity can be closely approached

by using adaptive modulation with appropriate channel codes and optimum successive detection (OSD). First, the constellation is selected based on the instantaneous capacity of each channel realization. Then the density evolution technique is employed to determine the maximal achievable rate of an LDPC code for each transmit antenna for each channel realization, at a given SNR. If the fading process is non-ergodic, the outage capacity corresponding to a given outage probability is used to measure the channel performance. As an example, we design the LDPC codes for an adaptively modulated 2×2 V-BLAST system to approach its outage capacity for a given outage probability.

Since the constellation size and channel code rate have to match each channel realization, codes of different rates and block lengths are required for different transmissions. Flexible-length rate-compatible punctured irregular repeat-accumulate (IRA) codes are introduced to accomplish this goal. We propose a two-step shortening and puncturing process to obtain codes of different rates and block lengths from one underlying IRA mother code while satisfying constraint imposed by the chosen modulation alphabet, fixed frame length (in symbols) and the target code rate. A key advantage of this approach is that the optimality of the degree distribution is maintained in the shortening process. Further, good performance of codes with different rates is guaranteed by optimizing the shortening and puncturing distributions. The shortening step has to preserve the code rate and information node degree distribution while reducing the mother code to the target block length. Higher rate codes can be obtained by puncturing the shortened mother code according to the optimal puncturing distributions of the information bits and the parity bits.

Chapter 1

Introduction

1.1 Background

The continued increase in demand for all types of wireless services, including voice, data and multimedia is fueling the need for higher capacity. This goal is particularly challenging for systems that are power, bandwidth, and complexity limited. However, wireless communication using multi-input multi-output (MIMO) has recently emerged as one of the most significant breakthroughs to improve system spectral efficiency [1, 2].

Pioneering work by Winters [3], Foschini [4], and Telatar [5] predicted remarkable spectral efficiencies for wireless MIMO systems when a rich scattering environment is present and channel state information can be accurately tracked. Many MIMO techniques have been developed to capitalize on the theoretical capacity gains [6] predicted by Shannon theory. Practical design of MIMO systems involves the development of finite-complexity transmission/reception signal processing algorithms such as space-time processing [7, 8, 9], space-time coding [10, 11, 12] and spatial multiplexing [13, 14, 15].

Smart antenna systems have been employed to get array gain and/or reduce cochannel interference for several decades [16]. In the conventional smart antenna system [17], only the transmitter or the receiver is actually equipped with multiple antenna elements, typically the base station [18]. The signals are processed at the transmitter and/or the receiver, resulting in an increase in average receiver SNR due to a coherent combining effect [19]. Transmit/receive array gain requires channel knowledge at the transmitter and receiver, respectively, and depends on the number of transmit and receive antennas.

Cochannel interference is inevitable in multi-user wireless communication systems. If the channel knowledge of the desired signal, and possibly that of the interference signals, is available, the received signal can be optimally combined such that the average SNIR is maximized. When multiple antennas are used, the differentiation between the spatial signatures of the desired signal and cochannel signals can be exploited to reduce interference. This idea is represented in transmitter/receiver beamforming [20]. Essentially, beamforming is used to increase the average SNR (SNIR) through focusing energy in a desired direction, in either the transmitter or the receiver [21].

Multiple antennas have also been used to achieve spatial diversity [22, 23], thus combating channel fading. Each pair of transmit and receive antennas provides a signal path from the transmitter to the receiver. By sending signals that carry the same information through different paths, a maximal diversity gain of $N_t N_r$ can be achieved if the paths between the N_t transmit antennas and N_r receive antennas are modeled as i.i.d flat Rayleigh fading. Space-time coding [24, 25] is one of the most prominent techniques to achieve the diversity gain in the absence of channel knowledge at the transmitter.

However, an alternative approach suggests that fading can be beneficial in a MIMO system [5]. Essentially, if the paths between individual transmit-receive antenna pairs are i.i.d Rayleigh fading channels, significant spatial multiplexing gain can be obtained by transmitting independent information on those parallel spatial channels [26, 27, 28]. Foschini has shown that in the high SNR regime, the capacity of a system with N_t transmit antennas, N_r receive antennas and i.i.d Rayleigh fading path between each antenna pair grows linearly with the minimum of N_t and N_r [29].

1.2 Outline of the Dissertation

In Chapter 2, the uplink of an asynchronous Multi-Carrier Direct Sequence Code Division Multiple Access (MC-DS-CDMA) [30] system with multiple antennas at both the transmitter and the receiver is considered. We analyze the system performance over a spatially correlated Rayleigh fading channel [31] with multiple access interference (MAI), and evaluate the antenna array performance with joint fading reduction and MAI suppression. Assuming perfect channel knowledge available at the transmitter, Maximal Ratio Transmission (MRT) [32] is employed to weight the transmitted signal optimally in terms of combating signal fading. At the receiver,

adaptive beamforming reception is adopted to both suppress MAI and combat fading. Note that while the correlation among the fades of the antennas in the receive array reduce the diversity gain against fading, the array still has the capability for interference suppression. We examine the effect of varying the number of transmit and receive antennas on both the diversity gain and the interference suppression.

The Vertical Bell-Laboratories Layered Space-Time (V-BLAST) [33] system has been actively investigated in recent years since it appears to be a robust, low-complexity and cost-effective solution for future wireless networks [13]. In Chapter 3, we present a practical implementation of the V-BLAST type system, in which the MIMO open-loop capacity [34] can be approached with conventional scalar coding, using adaptive modulation with appropriate channel codes [35], e.g., LDPC codes, and optimum successive detection (OSD) [36]. Depending upon the relation between the channel coherence time and the codeword length, the latter being determined by system transmission delay constraint, we use either ergodic capacity or outage capacity [37] as the performance measure. A block fading model is assumed for each channel realization to calculate the instantaneous capacity, which is used to select the signal constellation. Then, the density evolution technique [38] is employed to determine the maximal achievable rate of an LDPC code for each transmit antenna for a given channel realization, at a given SNR. Numerical results show that the average sum rate of our adaptively modulated LDPC encoded system is quite close to the V-BLAST capacity with both rate and power adaptations [39]. Considering the performance degradation caused by error propagation due to the imperfect feedback and relatively long decoding delay in the OSD detection, we use parallel soft interference cancellation (PIC) followed by MMSE filtering [40] in the BER performance simulation. If the system transmission delay is shorter than the channel coherence time, the outage capacity [41] corresponding to a given outage probability is used to measure the channel performance. As an example, we design the LDPC codes for an adaptively modulated 2×2 V-BLAST system to approach its outage capacity for a given outage probability.

Since the constellation size and channel code rate have to match each channel realization, codes of distinct rates and block lengths are required for different transmissions. Although it is feasible to implement rate-compatible punctured LDPC codes having different block lengths from one underlying mother code, an intensive computer search is required to design the parity check matrix for the mother code [42]. We choose irregular repeat-accumulate (IRA) code [43] for the flexible-length rate-compatible punctured code design due to its unique code struc-

ture. A two-step shortening and puncturing process is proposed to obtain codes of different rates and block lengths from one underlying IRA mother code, while satisfying constraint imposed by the chosen modulation alphabet, fixed frame length (in symbols) and the target code rate in Chapter 4. An key advantage of this approach is that the optimality of the degree distribution is maintained in the shortening process. Further, good performances of codes with different rates is guaranteed by optimizing the shortening and puncturing distributions. The shortening step preserves the code rate and the information node degree distribution, while reducing the mother code to a shorter block length. To maintain the code rate, the number of information bits and the number of parity bits are expunged proportionally. Meanwhile, the number of expunged information bits in different degree groups are also proportional to their respective cardinalities to preserve the information node degree distribution. Higher rate codes can be obtained by puncturing the shortened mother code. A standard Gaussian approximation (GA) [44] is used to optimize the shortening distribution of the parity bits and the puncturing distributions of the information and parity bits so that the performance of the punctured codes can approach the performance of the optimal codes very closely.

Chapter 2

The Tradeoff between Diversity Gain and Interference Suppression in a MIMO MC-CDMA System

2.1 Introduction

For systems with power, bandwidth, and delay limits, the use of multiple antennas has become an effective technique to improve the system capacity. However, when omni-directional antennas are used at the base station, the transmission/reception of each user's signal becomes a source of interference to other users located in the same cell, making the overall system interference limited. Beamforming is an effective way to reduce the interference, thus improving the system capacity. With this technology, each user's signal is transmitted and received by the base station only in the direction of that particular user, which can significantly reduce the overall interference in the system. Essentially, an array of antennas at the base station can direct different transmission/reception beams toward each user in the system, which is realized by signal processing at the baseband.

In beamforming, each user's signal is multiplied with complex weights that adjust the magnitude and phase of the signal to and from each antenna. This causes the output from the array of antennas to form a transmit/receive beam in the desired direction and minimize the output in other directions. Through adaptive beamforming, the base station can form narrow

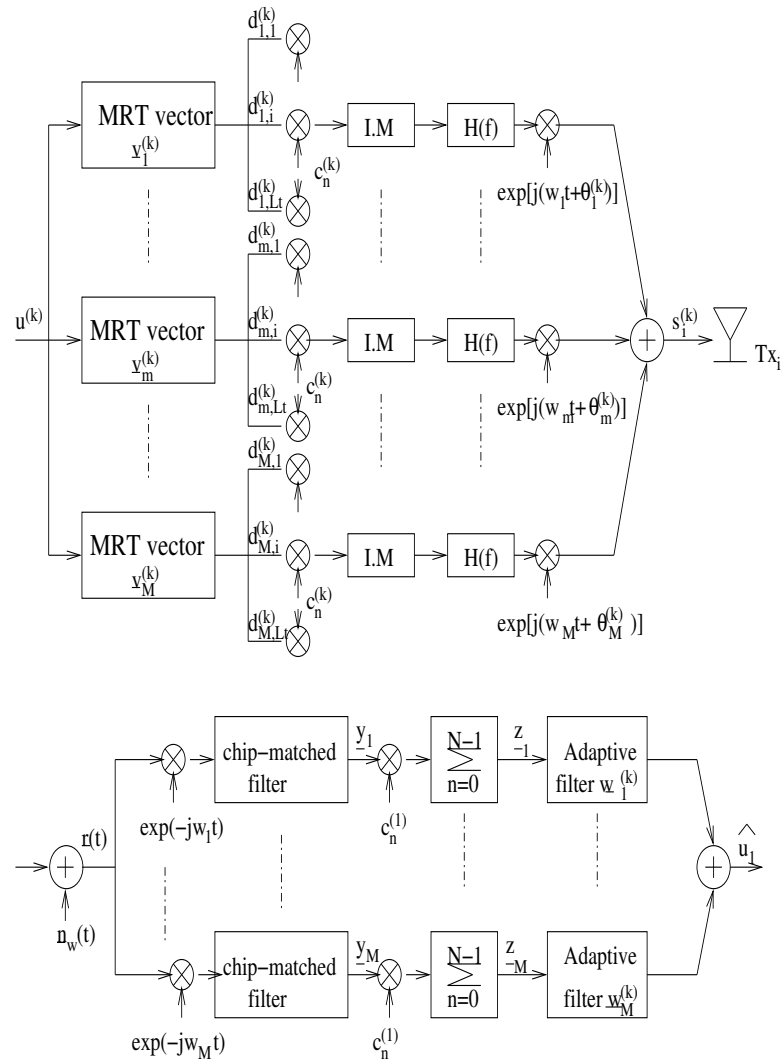


Figure 2.1: Transceiver with adaptive beamforming in a MC-CDMA system

beams toward the desired user and nulls towards interfering users, considerably improving the signal-to-interference-plus-noise ratio (SINR). Thus, beamforming can be used to improve the performance of wireless systems when there is a dominant direction-of-arrival (DOA) for the signal-of-interest [21].

Compared with spatial multiplexing [45, 13], beamforming is preferred in terms of complexity. On the other hand, beamforming, in general, has a much lower data rate compared to spatial-multiplexing in a single user multiple antenna system. However, in a multiple access channel, where K users, each with L_T transmit antennas, try to communicate with a common receiver with L_R receive antennas, beamforming is not only sufficient but also necessary for achieving the so-called sum capacity of multiple access channels if the number of users is much larger than the number of receive antennas [46]. This latter condition generally holds in a CDMA system.

It is natural to use multiple antennas at the receiver to achieve the diversity gain through maximal ratio combining (MRC). When multiple antennas are available at the transmitter, transmit diversity can be realized by sending symbols and their linear combinations from different antennas, i.e., space-time coding. With two transmit antennas, the Alamouti scheme [7] is the simplest, and yet one of the most elegant space time codes, and achieves full diversity gain without rate loss. Due to the orthogonal design, the two different symbols sent from different antennas can be separated after simple linear combinations of the received signal, which makes Alamouti scheme attractive enough to have been adopted in 3G cellular standards. Alamouti scheme can be extended to more than two antennas, and is also called space-time block coding (STBC) [23]. However, STBC cannot provide any coding gain that can be achieved by space-time trellis code (STTC) [47, 24] at the expense of much more detection complexity. Both STBC and STTC are appropriate for an open-loop MIMO system.

Compared with the gain achieved by the receive antenna array using MRC, the Alamouti scheme has a 3dB array gain loss resulting from allocating equal transmit power to each transmit antenna. If channel state information is known at the transmitter, the 3dB loss can be recovered by weighting the transmit symbol to match the channel state, and this is also known as maximal ratio transmission (MRT) [32].

In this chapter, an asynchronous Multi-Carrier Direct Sequence (MC-DS) CDMA system with multiple antennas at both the transmitter and the receiver is considered. We analyze the system performance over a spatially correlated Rayleigh fading channel with multiple access

interference (MAI), and evaluate the antenna array performance with joint fading reduction and MAI suppression. The outline of the remaining part is as follows. The system model and fading channel model used in the study are described in Section 2.2. Section 2.3 presents the analysis of system performance, and is followed by some numerical results and discussions in Section 2.4. Finally, conclusions are drawn in Section 2.5.

2.2 System Model

We describe a system model exploiting multiple antennas in a single cell MC-CDMA system. Assume that both the mobiles and the base station use an antenna array to transmit and receive signals, where each mobile has an antenna array of size L_T , and the base station has an antenna array of size L_R , as shown in Fig. 2.1. Assuming perfect channel knowledge available at the transmitter, MRT is employed to weight the transmitted signal optimally in terms of combating signal fading. Adaptive beamforming reception is adopted to suppress MAI and combat the fading.

2.2.1 Transmitter

For the block diagram shown in Fig. 2.1, the transmitted signal vector, of dimension $(L_T \times 1)$, in the m -th subband for user k , is given by

$$\begin{aligned} \underline{s}_m^{(k)}(t) &= \sqrt{P_k} c_k(t) \underline{d}_m^{(k)} \exp \left[j(\omega_m t + \theta_m^{(k)}) \right] \\ &= \sqrt{P_k} \sum_{n=-\infty}^{\infty} c_n^{(k)} h(t - nT_c) u_{\lfloor n/N_s \rfloor}^{(k)} \\ &\quad \cdot \underline{v}_m^{(k)} \exp \left[j(\omega_m t + \theta_m^{(k)}) \right], \end{aligned} \quad (2.1)$$

where $u_i^{(k)}$ is the i -th data symbol of user k , $\underline{v}_m^{(k)}$ is a transmission weight vector for user k in the m -th subband, ω_m is the subcarrier frequency, $\theta_m^{(k)}$ is a random carrier phase associated with user k in the m -th subcarrier band and is uniformly distributed over $[0, 2\pi)$, the spreading sequence of the interfering users, $c_n^{(k)}$, $k = 2, \dots, K$, are assumed to be i.i.d. random variables taking values ± 1 with equal probability, while that of the desired user, $c_n^{(1)}$, is taken to be deterministic, $h(t)$ is the impulse response of the baseband chip wave-shaping filter, and $1/T_c$ is the chip rate of a band-limited MC-DS-CDMA system. We assume the chip wave-shaping

filter $H(f)$ is bandlimited so that the spectra in each subband do not overlap. We also define $x(t) = F^{-1} |H(f)|^2$ and assume that $x(t)$ satisfies the Nyquist criterion, i.e., $x(nT_c) = \delta(n)$. The processing gain is defined as $N_s = T_s/T_c$, and is taken to be much smaller than the period of the spreading sequence, where T_s is the symbol duration. Then we can write the transmitted signal vector from the k th user as

$$\begin{aligned} \underline{s}^{(k)}(t) &= \sqrt{P_k} \sum_{n=-\infty}^{\infty} u_{\lfloor n/N_s \rfloor}^{(k)} c_n^{(k)} h(t - nT_c) \\ &\quad \cdot \sum_{m=1}^M v_m^{(k)} \exp \left[j(\omega_m t + \theta_m^{(k)}) \right]. \end{aligned} \quad (2.2)$$

2.2.2 Channel Model

The channel model is taken to be a slowly varying Rayleigh fading channel for each subcarrier, with transfer function $\xi_{m,r,l}^{(k)} = \alpha_{m,r,l}^{(k)} \exp(j\beta_{m,r,l}^{(k)})$, for $r = 1, \dots, L_R$ and $l = 1, \dots, L_T$, where l is the index for the transmit antennas and r is the index for the receive antennas. We assume that $\{\alpha_{m,r,l}^{(k)}\}$ and $\{\beta_{m,r,l}^{(k)}\}$ are statistically independent for different users, and that $\{\alpha_{m,r,l}^{(k)}\}$ and $\{\beta_{m,r,l}^{(k)}\}$ are, respectively, i.i.d Rayleigh random variables with a unit second moment, and uniform random variables over $[0, 2\pi)$ for different transmit antennas. However, the array gain and the phase of the different elements in the receive antenna array are correlated, where the correlation is determined by parameters such as direction of arrival $\phi_l^{(k)}$, angular spread $\Delta_l^{(k)}$, spacing between neighboring receive antennas D_r , and the wavelength of the carrier signal λ .

Specifically, using the model in [31], the composite channel gains for all the antennas in the array are represented as

$$\begin{aligned} \zeta_{m,i+1,l}^{(k)} &= \zeta_{m,1,l}^{(k)} \exp(-j2\pi i \sin \frac{\phi_l^{(k)}}{\lambda}) \\ &= \alpha_{m,l}^{(k)} \exp \left[j \left(\beta_{m,1,l}^{(k)} \right)' - j2\pi i \sin \frac{\phi_l^{(k)}}{\lambda} \right], \end{aligned} \quad (2.3)$$

where $\left(\beta_{m,r,l}^{(k)} \right)' = \beta_{m,r,l}^{(k)} + \theta_m^{(k)}$ is uniformly distributed over $[0, 2\pi)$. If we make the additional, physically-reasonable, assumption that the angles of arrival, $\phi_{l,n}^{(k)}$'s, are uniformly distributed over $\left[\phi_l^{(k)} - \Delta_l^{(k)}, \phi_l^{(k)} + \Delta_l^{(k)} \right]$, a closed-form spatial correlation formula can be obtained [31].

That is,

$$\begin{aligned} E \left[\zeta_{m,i,l}^{(k)} \left(\zeta_{m,j,l}^{(k)} \right)^* \right] &= Rs(\Delta_l^{(k)}, \phi_l^{(k)}, D_r, \lambda) \\ &= Rs_l^I(i, j) + jRs_l^Q(i, j), \end{aligned} \quad (2.4)$$

where $Rs_l^I(i, j)$ and $Rs_l^Q(i, j)$ are given by

$$\begin{aligned} Rs_l^I(i, j) &= J_0 \left(\frac{2\pi D_r |i - j|}{\lambda} \right) + 2 \sum_{n=1}^{\infty} J_{2n} \left(\frac{2\pi D_r |i - j|}{\lambda} \right) \\ &\quad \cdot \cos(2n\phi_l^{(k)}) \text{sinc}(2n\Delta_l^{(k)}) \end{aligned} \quad (2.5)$$

and

$$\begin{aligned} Rs_l^Q(i, j) &= 2 \sum_{n=0}^{\infty} J_{2n+1} \left(\frac{2\pi D_r |i - j|}{\lambda} \right) \\ &\quad \cdot \sin \left((2n + 1)\phi_l^{(k)} \right) \text{sinc} \left((2n + 1)\Delta_l^{(k)} \right) \end{aligned} \quad (2.6)$$

respectively, for $l = 1, \dots, L_T$, and where the J_n 's are Bessel functions of integer order. When this correlation is high, the signals at the antennas tend to fade at the same time, and the diversity benefit of antenna arrays against fading is significantly reduced. On the other hand, because independent fading is not required for interference suppression, antenna arrays can suppress interference even with complete correlation. Thus, we need to evaluate the antenna array performance with joint fading reduction and interference suppression.

We define a channel matrix $H_m^{(k)}$ by putting the channel gain of each transmit and receive antenna pair in the m -th subband into a matrix of size $L_R \times L_T$. That is to say, the (i, j) th entry in $H_m^{(k)}$ is $\xi_{m,i,j}^{(k)}$.

2.3 Performance Analysis

2.3.1 Output of the m th Correlator

The received signal vector in the antenna array is obtained as

$$\underline{r}(t) = \sum_{k=1}^K \sum_{m=1}^M H_m^{(k)} \underline{s}_m^{(k)}(t - \tau_k) + \underline{n}_w(t) \quad (2.7)$$

where τ_k is an arbitrary time delay uniformly distributed over $[0, T_s]$, and $\underline{n}_w(t)$ is the AWGN vector added to the receive antenna array and each of its elements is a zero-mean complex Gaussian random process with two-sided spectral density η_0 . An asynchronous MC-DS-CDMA is assumed, but the receiver is synchronized to the desired transmission, say that of user 1; thus, we assume that the power and delay of the desired signal are, respectively, $P_1 = 1$ and $\tau_1 = 0$, without loss of generality.

We evaluate the performance of the first user with the assumption of perfect carrier, code, and bit synchronization. After down-converting to baseband, we can write the complex baseband received signal vector at the antenna array in the m -th subband as

$$\begin{aligned} \underline{y}_m(t) &= \sum_{k=1}^K \sqrt{P_k} \sum_{n=-\infty}^{\infty} c_n^{(k)} h(t - nT_c - \tau_k) \\ &\quad \cdot u_{\lfloor n/N_s \rfloor}^{(k)} H_m^{(k)} \underline{v}_m^{(k)} \exp(j\theta_m^{(k)}) + \underline{n}_m(t) \\ &= \sum_{n=-\infty}^{\infty} u_{\lfloor n/N_s \rfloor}^{(1)} c_n^{(1)} h(t - nT_c) G_m^{(1)} \underline{v}_m^{(1)} + \underline{i}_m^{(1)}(t) + \underline{n}_m(t), \end{aligned} \quad (2.8)$$

where

$$\underline{i}_m^{(1)}(t) = \sum_{k=2}^K \sqrt{P_k} \sum_{n=-\infty}^{\infty} u_{\lfloor n/N_s \rfloor}^{(k)} c_n^{(k)} \cdot h(t - nT_c - \tau_k) G_m^{(k)} \underline{v}_m^{(k)} \quad (2.9)$$

is the composite of multiple access interference (MAI), and $(G_m^{(k)}) = H_m^{(k)} \exp(j\theta_m^{(k)})$, for $m = 1, \dots, M$ and $k = 1, \dots, K$. The noise is given by

$$\underline{n}_m(t) = \underline{n}_w(t) \exp(-j\omega_m t), \quad (2.10)$$

and is a complex AWGN process.

The output of the correlator during the i -th symbol interval, obtained by summing the corresponding despread N_s chip-matched filter output samples in the m -th branch, $\underline{z}_m^{(1)}(i)$, is given by

$$\begin{aligned} \underline{z}_m^{(1)}(i) &= \frac{1}{N_s} \sum_{n'=(i-1)N_s}^{iN_s-1} c_{n'}^{(1)} \left[\underline{y}_m(t) \star h(-t) \right]_{t=n'T_c} \\ &= \frac{1}{N_s} G_m^{(1)} \underline{v}_m^{(1)} \sum_{n'=(i-1)N_s}^{iN_s-1} c_{n'}^{(1)} \sum_{n=-\infty}^{\infty} c_n^{(1)} \\ &\quad \cdot u_{\lfloor n/N_s \rfloor}^{(1)} \int_{-\infty}^{\infty} h((n' - n)T_c + \tau) h(\tau) d\tau + \underline{I}_m^{(1)}(i) + \underline{N}_m^{(1)}(i) \\ &= u_i^{(1)} G_m^{(1)} \underline{v}_m^{(1)} + \underline{I}_m^{(1)}(i) + \underline{N}_m^{(1)}(i), \end{aligned} \quad (2.11)$$

where \star represents convolution. $u_i^{(1)} G_m^{(1)} \underline{v}_m^{(1)}$ is the signal component for the desired user,

$$\underline{N}_m^{(1)}(i) = \frac{1}{N_s} \sum_{n=(i-1)N_s}^{iN_s-1} c_n^{(1)} \{ \underline{v}_m(t) \star h(-t) \}_{|t=nT_c} \quad (2.12)$$

is the component due to thermal noise, and

$$\begin{aligned} \underline{I}_m^{(1)}(i) &= \sum_{k=2}^K \frac{\sqrt{P_k}}{N_s} \sum_{n'=(i-1)N_s}^{iN_s-1} c_{n'}^{(1)} \sum_{n=-\infty}^{\infty} u_{\lfloor n/N_s \rfloor}^{(k)} c_n^{(k)} x((n' - n)T_c - \tau_k) G_m^{(k)} \underline{v}_m^{(k)} \\ &= \sum_{k=2}^K \frac{\sqrt{P_k}}{N_s} \sum_{n'=(i-1)N_s}^{iN_s-1} c_{n'}^{(1)} \sum_{n=-\infty}^{\infty} \mu_n^{(k)} x((n' - n)T_c - \tau_k) G_m^{(k)} \underline{v}_m^{(k)} \\ &= \sum_{k=2}^K \frac{\sqrt{P_k}}{N_s} R_{k,1}(i) G_m^{(k)} \underline{v}_m^{(k)} \end{aligned} \quad (2.13)$$

is the multiple access interference. In (2.13), $\mu_n^{(k)} = u_{\lfloor n/N_s \rfloor}^{(k)} c_n^{(k)}$ and

$$R_{k,1}(i) = \sum_{n'=(i-1)N_s}^{iN_s-1} c_{n'}^{(1)} \sum_{n=-\infty}^{\infty} \mu_n^{(k)} x((n' - n)T_c - \tau_k) \quad (2.14)$$

is the cross-correlation function of the spreading signal between user k and user 1 during the i -th symbol interval. Here we absorb $u_{\lfloor n/N_s \rfloor}^{(k)}$ into $c_n^{(k)}$, since both are random variables taking values of ± 1 with equal probability. By the Liapounoff version of the central limit theorem, $\underline{I}_m^{(1)}(i)$ can be modeled as an asymptotically complex Gaussian vector as long as the following condition is satisfied [30]:

$$\sum_{n=-\infty}^{\infty} |x(nT_c - \tau)| < \infty$$

for all τ , where $0 \leq \tau < T_c$.

2.3.2 Output of the Adaptive Beamformer

The correlator outputs from each receive antenna in each subband are combined with the beamforming vector $\underline{w}_1 = [(\underline{w}_1^{(1)})^T, \dots, (\underline{w}_M^{(1)})^T]^T$ to produce an estimate of the transmitted symbol of the desired user, where $\underline{w}_m^{(1)}$ is the beamforming vector for the m -th subband, $m = 1, \dots, M$. Define the correlator output vector $\underline{z}_1(i) = [(\underline{z}_1^{(1)}(i))^T, \dots, (\underline{z}_M^{(1)}(i))^T]^T$. Then

the estimated data symbol can be represented as

$$\begin{aligned}\hat{u}_{1,i} &= \underline{w}_1^\dagger \underline{z}_1(i) \\ &= S_{1,i} + I_{1,i} + N_{1,i},\end{aligned}\quad (2.15)$$

where \dagger denotes complex conjugate,

$$\begin{aligned}S_{1,i} &= \sum_{m=1}^M (\underline{w}_m^{(1)})^\dagger \underline{S}_m^{(1)}(i) \\ &= u_i^{(1)} \sum_{m=1}^M (\underline{w}_m^{(1)})^\dagger G_m^{(1)} \underline{v}_m^{(1)} \\ &= u_i^{(1)} \underline{w}_1^\dagger G_1 \underline{v}_1,\end{aligned}\quad (2.16)$$

$$\begin{aligned}I_{1,i} &= \sum_{m=1}^M (\underline{w}_m^{(1)})^\dagger \underline{I}_m^{(1)}(i) \\ &= \sum_{k=2}^K \frac{\sqrt{P_k}}{N_s} R_{k,1}(i) \sum_{m=1}^M (\underline{w}_m^{(1)})^\dagger G_m^{(k)} \underline{v}_m^{(k)} \\ &= \sum_{k=2}^K \frac{\sqrt{P_k}}{N_s} R_{k,1}(i) \underline{w}_1^\dagger G_k \underline{v}_k,\end{aligned}\quad (2.17)$$

$G_k = \text{diag} \left[G_1^{(k)} \quad \dots \quad G_M^{(k)} \right]$ for $k = 1, 2, \dots, K$, $\underline{v}_k = \left[(\underline{v}_1^{(k)})^T, \dots, (\underline{v}_M^{(k)})^T \right]^T$, and

$$N_{1,i} = \sum_{m=1}^M (\underline{w}_m^{(1)})^\dagger \underline{N}_m^{(1)}(i), \quad (2.18)$$

$$\underline{N}_1(i) = \left[(\underline{N}_1^{(1)}(i))^T, \dots, (\underline{N}_M^{(1)}(i))^T \right]^T.$$

Now we proceed to determine the optimum transmit and receive weight vectors \underline{v}_1 and \underline{w}_1 , respectively, for the desired user. Since the MAI $I_1(i)$ can be modeled as an asymptotically zero-mean complex Gaussian vector, and is independent of the AWGN vector $\underline{N}_1(i)$, the conditional SINR γ_i of the estimated data $\hat{u}_{1,i}$, conditioned on G_1 , is given by

$$\begin{aligned}
\gamma_i &= \frac{|S_{1,i}|^2}{\text{Var}(I_{1,i}) + \text{Var}(N_{1,i})} \\
&= \frac{\underline{w}_1^\dagger G_1 \underline{v}_1 \underline{v}_1^\dagger G_1^\dagger \underline{w}_1}{\underline{w}_1^\dagger E \left\{ \frac{1}{2} \left[\underline{I}_1(i) \underline{I}_1^\dagger(i) + \underline{N}_1(i) \underline{N}_1^\dagger(i) \right] \right\} \underline{w}_1} \\
&\simeq \frac{\left| \underline{w}_1^\dagger G_1 \underline{v}_1 \right|^2}{\underline{w}_1^\dagger \left[\sum_{k=2}^K \frac{P_k}{2N_s} R_I(0) \Gamma_k^\dagger R_{MI}^{(k)} \Gamma_k + \frac{\eta_0}{N_s} I_{MLR} \right] \underline{w}_1}, \tag{2.19}
\end{aligned}$$

where

$$\begin{aligned}
\text{Var}(I_{1,i}) &= \frac{1}{2} E \left[\underline{I}_1(i) \underline{I}_1^\dagger(i) \right] \\
&= \sum_{k=2}^K \frac{P_k}{2N_s} \left[R_I^{(k)}(0) + 2 \sum_{m=1}^{N_s-1} R_I^{(k)}(mT_c) \cdot \sum_{n=m}^{N_s-1} c_n^{(1)} c_{n-m}^{(1)} \right] \Gamma_k^\dagger R_{MI}^{(k)} \Gamma_k \\
&\simeq \sum_{k=2}^K \frac{P_k}{2N_s} R_I(0) \Gamma_k^\dagger R_{MI}^{(k)} \Gamma_k. \tag{2.20}
\end{aligned}$$

$R_I(\tau)$ is the autocorrelation of $R_{k,1}(i)$ (see Eq. (2.14)), Γ_k is a matrix given by

$$\begin{aligned}
\Gamma_k &= \begin{bmatrix} V_1^{(k)} & \cdots & 0 \\ \vdots & V_m^{(k)} & 0 \\ 0 & \cdots & V_M^{(k)} \end{bmatrix} \\
V_m^{(k)} &= \underbrace{\begin{bmatrix} \underline{v}_m^{(k)} & \cdots & 0 \\ \vdots & \underline{v}_m^{(k)} & 0 \\ 0 & \cdots & \underline{v}_m^{(k)} \end{bmatrix}}_{L_R L_T \times L_R}
\end{aligned}$$

and $R_{MI}^{(k)}$ is a matrix whose elements are the cross-correlations of the channel gains of user k .

2.3.3 Maximal Ratio Transmission and Adaptive Beamforming Reception

In statistically optimum beamforming, the weights are chosen based on the statistics of the data received at the array. Loosely speaking, the goal is to “optimize” the beamformer response so that the output contains minimal contributions due to noise and signals arriving from directions other than the desired signal direction. There are several different criteria for

choosing statistically optimum beamformer weights, with perhaps the most obvious one being the maximization of SNR.

By using MRT, we set $\underline{v}_k = c_k G_k^\dagger \underline{w}_k$, where c_k is a constant used for normalization. Subject to the transmit power constraint $\|\underline{v}_k\|^2 = 1$, we have $|c_k| = \frac{1}{\|G_k^\dagger \underline{w}_k\|}$. Then the transmit weight vector is given by

$$\underline{v}_k = \frac{G_k^\dagger \underline{w}_k}{\|G_k^\dagger \underline{w}_k\|}. \quad (2.21)$$

After using (2.21) in (2.16), and considering $k = 1$ as the desired user, we obtain

$$\begin{aligned} |S_{1,i}|^2 &= \left| \underline{w}_1^\dagger G_1 \underline{v}_1 \right|^2 \\ &= \underline{w}_1^\dagger G_1 G_1^\dagger \underline{w}_1. \end{aligned} \quad (2.22)$$

Now the goal is to choose beamforming weight vector \underline{w}_1 which maximizes the SNR γ_i of (2.19). Subject to the normalization constraint, we have

$$\hat{\underline{w}}_1 = \arg \max_{\|\underline{w}_1\|^2=1} \frac{\underline{w}_1^\dagger G_1 G_1^\dagger \underline{w}_1}{\underline{w}_1^\dagger \left(\sum_{j=2}^K \frac{P_k}{2N_s} R_I(0) \Gamma_j^\dagger R_{MI}^{(j)} \Gamma_j + \frac{\eta_0}{N_s} I_{MLR} \right) \underline{w}_1}. \quad (2.23)$$

Then the optimum weight vector $\hat{\underline{w}}_1$ is the principle eigenvector of $(\sum_{j=2}^K \frac{P_k}{2N_s} R_I(0) \Gamma_j^\dagger R_{MI}^{(j)} \Gamma_j + \frac{\eta_0}{N_s} I_{MLR})^{-1} G_1 G_1^\dagger$, and γ_i is the corresponding eigenvalue [48], i.e., the maximum eigenvalue of $(\sum_{j=2}^K \frac{P_k}{2N_s} R_I(0) \Gamma_j^\dagger R_{MI}^{(j)} \Gamma_j + \frac{\eta_0}{N_s} I_{MLR})^{-1} G_1 G_1^\dagger$.

To compute $\hat{\underline{w}}_1$, we need matrix Γ_k of user k , consisting of the transmit weight vector \underline{v}_k . However, this is not available, since it depends on receive weight vector \underline{w}_k (see Eq.(2.21)), which in turn cannot be computed without the knowledge of \underline{v}_j for $j \neq k$ (see Eq. (2.23)). So we cannot apply Eq.(2.23) directly to get the optimum weight vector $\hat{\underline{w}}_1$. As a consequence, one alternative is to use an iterative algorithm to solve the problem. Initially, we assume that \underline{v}_k is an equal weight vector, i.e., we weight each branch equally. Now it is possible to compute the beamforming weight vector $\hat{\underline{w}}_k$ for each user using Eq.(2.23). In turn, we can compute the corresponding transmit weight vector \underline{v}_k for each user using Eq.(2.21). By using these updated \underline{v}_k 's, we further update the $\hat{\underline{w}}_1$ iteratively until no improvement of SNR can be observed. This algorithm is quite complicated, in that the receiver has to recalculate the receive weight vector and feed it back to the corresponding transmitter. Note that this has to be done for all users anytime there is a noticeable change of state for any one of them. As just one example, this is has to be done whenever the number of active users changes in the system.

Considering the complexity of adjusting the receive and transmit weight vectors based upon the corresponding channel state information for all the active users, and the computational complexity that this involves, as an alternative, we can replace the optimum criterion which maximizes SINR with an ad hoc criterion which only maximizes the received power for the desired user. Following the steps described above, we obtain the optimum receive weight vector $\underline{w}_m^{(1)}$ for each subband as

$$\hat{\underline{w}}_m^{(1)} = \arg \max_{\|\underline{w}_1\|^2=1} \left\{ \left| (\underline{w}_m^{(1)})^\dagger G_m^{(1)} (G_m^{(1)})^\dagger \underline{w}_m^{(1)} \right|^2 \right\}. \quad (2.24)$$

Therefore, the receive weight vector $\hat{\underline{w}}_m^{(1)}$ is the scaled principal eigenvector of $G_m^{(1)}(G_m^{(1)})^\dagger$, and the received power in each subband $P_{r,m}^{(1)}$ is the corresponding eigenvalue, i.e., the maximum eigenvalue $\hat{\lambda}_m$ of $G_m^{(1)}(G_m^{(1)})^\dagger$. Although it is difficult to find the pdf $p_{\hat{\lambda}_m}(\lambda)$ of $\hat{\lambda}_m$ for the ensemble of matrices $G_m^{(1)}(G_m^{(1)})^\dagger$, bounds on the $\hat{\lambda}_m$ can be easily found. The fact that $G_m^{(1)}(G_m^{(1)})^\dagger$ is a Hermitian and positive semi-definite matrix guarantees its eigenvalues to be nonnegative. Hence, the $\hat{\lambda}_m$ are bounded by

$$\frac{\sum_{l=1}^{L_R} \lambda_l}{L_R} \leq \frac{\sum_{l=1}^{L_R} \lambda_l}{\text{rank} \left(G_m^{(1)} (G_m^{(1)})^\dagger \right)} \leq \hat{\lambda}_m \leq \sum_{l=1}^{L_R} \lambda_l = \text{Tr} \left(G_m^{(1)} (G_m^{(1)})^\dagger \right),$$

where $\text{rank} \left(G_m^{(1)} (G_m^{(1)})^\dagger \right) \leq \min(L_T, L_R)$, and $\text{Tr}(\cdot)$ is the trace of the matrix. Therefore

$$\begin{aligned} \frac{\text{Tr} \left(G_m^{(1)} (G_m^{(1)})^\dagger \right)}{L_R} &\leq P_{r,m}^{(1)} \\ &\leq \text{Tr} \left(G_m^{(1)} (G_m^{(1)})^\dagger \right) \\ &= \sum_{l=1}^{L_T} \sum_{r=1}^{L_R} \left| \zeta_{m,r,l}^{(1)} \right|^2 \\ &= L_R \sum_{l=1}^{L_T} \left| \alpha_{m,l}^{(1)} \right|^2 \end{aligned}$$

The last equality holds, since $\zeta_{m,r+1,l}^{(1)} = \alpha_{m,l}^{(1)} \exp \left(j(\beta_{m,1,l}^{(1)})' - j2\pi r \sin \frac{\phi_l^{(1)}}{\lambda} \right)$ is assumed. Thus,

$$P_r^{(1)} = \sum_{m=1}^M P_{r,m}^{(1)} \leq L_R \sum_{m=1}^M \sum_{l=1}^{L_T} \left| \alpha_{m,l}^{(1)} \right|^2, \quad (2.25)$$

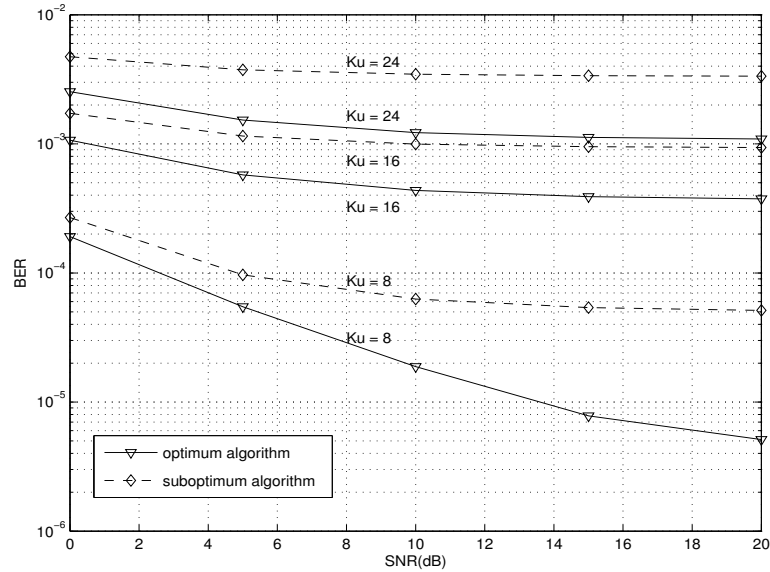


Figure 2.2: Performance comparison of optimum algorithm and suboptimum algorithm for a $M = 2$, $L_T = 2$ and $L_R = 4$ system with varying number of interfering users

and we see that this scheme achieves a diversity of the order ML_T , since the $\alpha_{m,l}^{(1)}$'s are assumed uncorrelated.

This scheme overcomes the disadvantages of the iterative algorithm, although it may suffer performance degradation. To quantify the performance loss, we resorted to simulation. We compared the performance of using the optimum algorithm and the suboptimum one, and the results are shown in Fig. 2.2, Fig. 2.3 and Fig. 2.4. Note that, in the low SNR region, the additive noise is typically larger than the MAI (especially when the number of interfering users is small), and $\sum_{j=2}^K \frac{P_k}{2N_s} R_I(0) \Gamma_j^\dagger R_{MI}^{(j)} \Gamma_j + \frac{\eta_0}{N_s} I_{ML_R}$ is dominated by the covariance matrix of the noise, which is a scaled identity matrix. Thus, optimizing the numerator term P_r is equivalent to optimizing the SINR, γ_i . As expected, the results in these figures show that the gain achieved by using the iterative algorithm is not significant in the low SNR region. As the SNR increases, the additive noise is no longer the dominant element. In the medium-to-high SNR region, the gain becomes more obvious by using the iterative algorithm.

We further observe that the improvement is smaller when there are more interfering users in the system. It seems that when the number of interfering users is large, the covariance matrix of the MAI, $\sum_{k=2}^K \frac{P_k}{2N_s} R_I(0) \Gamma_k^\dagger R_{MI}^{(k)} \Gamma_k$, is close to a scaled identity matrix. Although

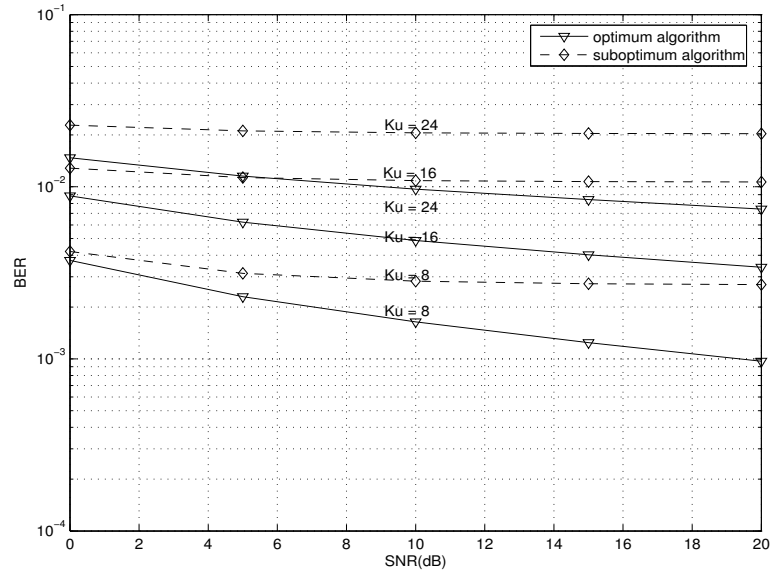


Figure 2.3: Performance comparison of optimum algorithm and suboptimum algorithm for a $M = 2$, $L_T = 1$ and $L_R = 8$ system with varying number of interfering users

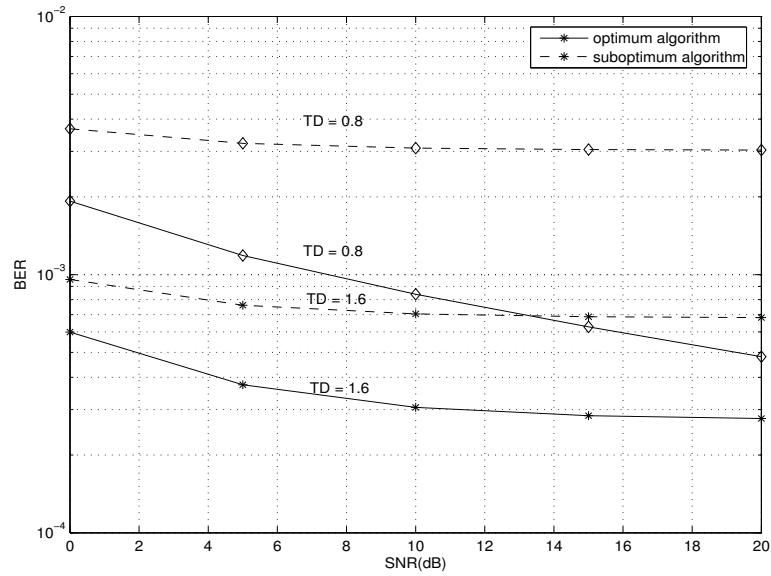


Figure 2.4: Performance comparison of optimum algorithm and suboptimum algorithm for a $M = 2$, $L_T = 2$ and $L_R = 4$ with different correlations among antennas

we cannot give a rigorous mathematical proof, an intuitive explanation based on the numerical results is as follows: The elements along the main diagonal of each $R_{MI}^{(k)}$ are unity, while the off-diagonal elements are complex numbers whose norms are smaller than unity (this condition holds as long as the antennas in the array are not fully correlated and the signals of different users arrive from directions uniformly distributed over $[-\frac{\pi}{2}, \frac{\pi}{2}]$). Also, it can be shown that all the diagonal elements in the matrix $\frac{P_k}{2N_s} R_I(0) \Gamma_k^\dagger R_{MI}^{(k)} \Gamma_k$ are positive real numbers. The off-diagonal elements are still complex numbers, whose real and imaginary parts could be either positive or negative. So the more the terms are in the summation, the more likely the polarities of those off-diagonal elements are averaged out, and the more accurate is the approximation of the covariance matrix of the MAI, $\sum_{k=2}^K \frac{P_k}{2N_s} R_I(0) \Gamma_k^\dagger R_{MI}^{(k)} \Gamma_k$, looking like a scaled identity matrix. Thus, as the number of interfering users increases, the improvement by using the optimum criterion with the iterative algorithm diminishes.

When the correlations among the antennas in the receive array become smaller, so does the improvement from using the iterative algorithm, as observed in Fig. 2.4. As we know, when the antennas become less correlated, the off-diagonal elements in $R_{MI}^{(k)}$ are much smaller than unity, while the diagonal elements are unity. Thus, the approximation of $R_{MI}^{(k)}$ by an identity matrix is more appropriate, and there is less gain to be achieved by using the optimum algorithm.

2.4 Numerical Results and Discussions

Given a fixed information rate and total bandwidth allocation, the product $MN_s = N^{sc}$ must be held constant, where N^{sc} is the processing gain of a single carrier CDMA system and N_s is the corresponding value for each subcarrier in the MC-CDMA system. We assume that the fading seen by each of the transmit antennas is independent. At the receiver, L_R receive antennas are deployed for adaptive beamforming reception, where L_R can be a large enough number so that the fading experienced by each receive antenna might be correlated. M independent subcarriers can provide M -th order frequency diversity gain, while L_T independent transmit antennas and L_R independent receive antennas result in an extra $L_T L_R$ order of spatial diversity gain. So fixing the value of $M L_T L_R$ fixes the maximal diversity gain achievable by the system. When the fading is, in fact, correlated, the diversity gain from the receive antenna array is reduced. However, independent fading is not required for interference suppression, so correlated receive antennas can still be used for MAI suppression. If we fix the product of $L_T L_R$, just

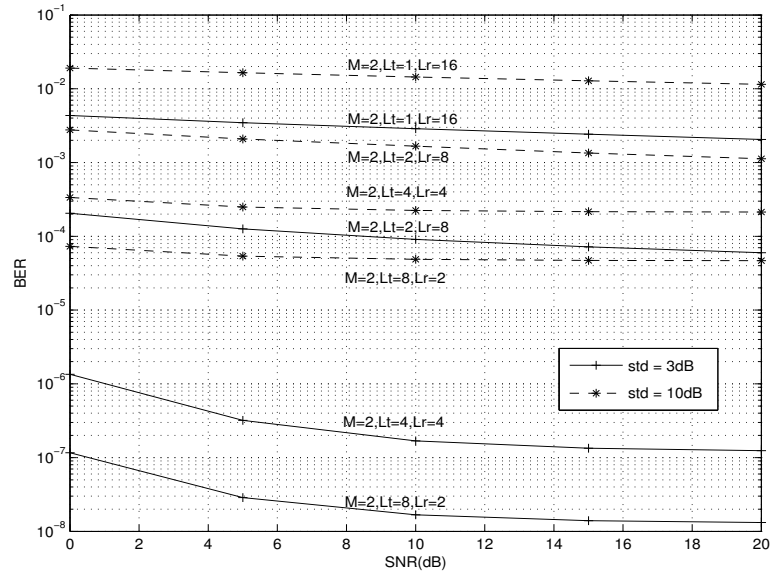


Figure 2.5: BER versus E_b/η_0 for $K = 30$

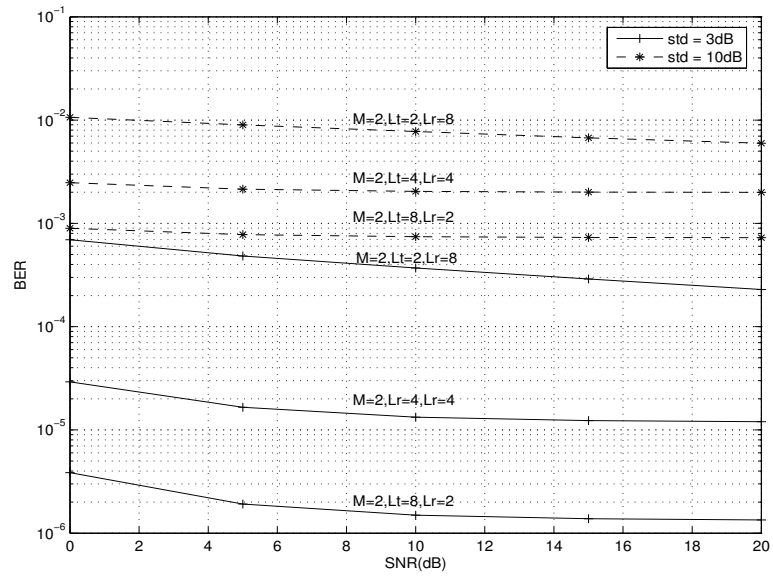


Figure 2.6: BER versus E_b/η_0 for $K = 50$

for the sake of having a frame of reference for the performance tradeoff, then increasing L_T will increase the diversity gain against fading while sacrificing the receive antenna array's capability of MAI suppression.

We assume the use of a raised-cosine filter characteristic, with rolloff factor $\alpha = 0.5$, for pulse shaping. We further assume the processing gain for a single carrier system to be fixed at $N^{sc} = 32$ unless it is explicitly stated. Since it is difficult to analytically derive the pdf of the instantaneous SNR, $f_\gamma(\gamma)$, we cannot obtain a closed-form expression for the BER. To circumvent this problem, a Monte-Carlo simulation is carried out. After one million trials, the SINR distribution of the combined outputs at the receiver is accumulated and $f_\gamma(\gamma)$ is numerically determined. The SINR value γ for each combined output is applied to the conditional bit error probability for a BPSK system, $\phi(\sqrt{2\gamma})$, and the average BER is calculated by integrating $Pe = \int_0^\infty \phi(\sqrt{2\gamma})f_\gamma(\gamma)d\gamma$.

We consider an MC-DS-CDMA system with 30 users, where the interference power is log-normally distributed with either a $3dB$ or a $10dB$ standard deviation. The average BER versus E_b/η_0 , for different sets of parameters, is shown in the figure 2.5. With the frequency diversity order fixed at $M = 2$, and $L_T L_R$ fixed to be 16, we find that the system employing 8 transmit antennas and 2 receive antennas is much better than one employing 4 transmit antennas and 4 receive antennas. This is primarily due to the eight-fold diversity gain from the eight transmit antennas with independent fading. Note that since the total length of the receive array is fixed at a value such that the multiple receive antennas experience correlated fading, the resulting effective diversity order achieved by the 4-antenna array is less than twice that achieved by the 2-antenna array, although the MAI suppression capability is enhanced with more receive antennas. We also compare in Fig. 2.5 the performance of other systems with the value of $M L_T L_R$ held constant. The worst case is $M = 2$, $L_T = 1$, and $L_R = 16$, since there is no transmit diversity gain and most of the receive diversity gain is lost due to the high correlations among the antennas in the receive array. We further evaluate the system performance with a more severe near-far problem, i.e, interference power is log-normally distributed with a $10dB$ standard deviation. Compared to the system with better power control, the BER performance of all of the above systems degrades by at least one order of magnitude. It is further observed that the degradations are more significant for the systems with $L_R = 2$ receive antennas than they are for the systems with more receive antennas. This phenomenon can be explained as follows. The system's ability to suppress MAI is augmented by using more receive antennas, while sacrificing some diversity

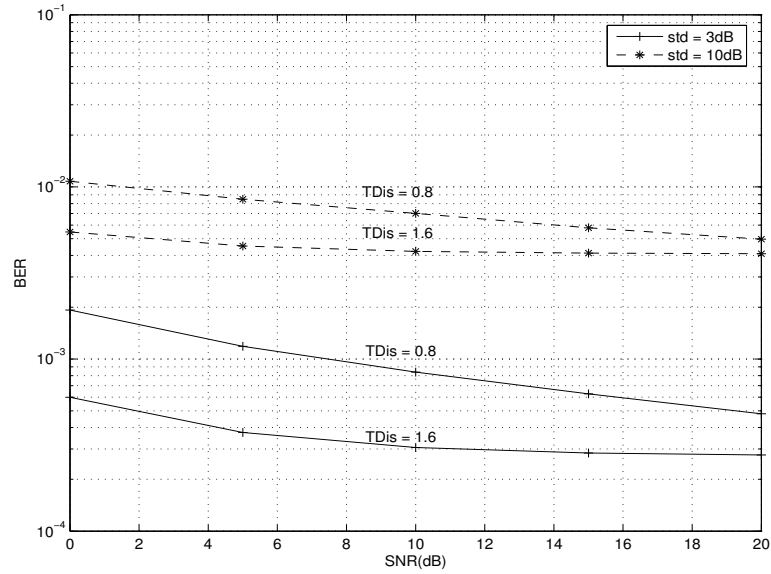


Figure 2.7: BER versus E_b/η_0 for $K = 30$ and $L_T = 2$, $L_R = 4$ with varying correlations between receive antennas

gain. Since MAI becomes more dominant in a system with a severe near-far problem, we find that the performance gap between the $L_R = 2$ and the $L_R = 4$, $L_R = 8$ or $L_R = 16$ systems decreases dramatically.

In Fig. 2.6, we plot BER performance curves for some of the systems in Fig. 2.5 when K is increased to 50. Compared with the curves plotted in Fig. 2.5 for systems with $K = 30$, there is smaller degradation when $L_R = 16$ receive antennas are employed. However, the degradation is much more conspicuous when $L_R = 2$ rather than $L_R = 16$ receive antennas are used. These observations indicate that systems with a larger number of receive antennas are more robust to various changes in the wireless environment, say, when the number of active users is constantly varying and/or the power control cannot be accurately implemented. Thus, it is beneficial to deploy more receive antennas in a dynamic wireless system to keep relatively stable service quality.

In Fig. 2.7 and Fig. 2.8, the BER performance when the correlations among receive antennas are varied by changing the spacing between neighboring antennas is shown. The fades become more correlated as we narrow the spacing. As we know, correlation results in loss of diversity gain against fading. However, the beamforming gain for MAI suppression is enhanced.

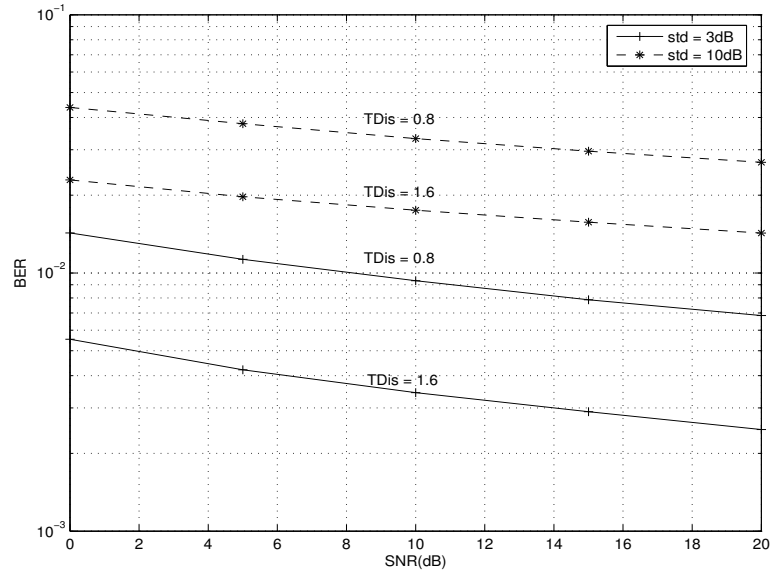


Figure 2.8: BER versus E_b/η_0 for $K = 30$ and $L_T = 1$, $L_R = 8$ with varying correlations between receive antennas

This fact can be seen from the curves plotted in those two figures. When MAI is dominant, e.g., interference power distributed with $10dB$ standard deviation, the performance degradation is much less than that in a system with better power control, where fading is the more dominant source of degradation. It is seen from the figures that the relative performance gap of the system with $L_T = 1$, $L_R = 8$ is larger than that of the system with $L_T = 2$, $L_R = 4$ when the correlations among the array increase due to a decrease in the spacing of neighboring antennas.

Note that while most of our results correspond to the product of ML_TL_R being held constant, in Fig. 2.9 and Fig. 2.10 we show the effect of doubling and tripling the number of receive antennas while keeping both M and L_T constant, for both the optimum and the suboptimum algorithms. The resulting performance improvement is not as significant as might be expected. The reason is that we cannot double or triple the order of the diversity by doubling or tripling the number of receive antennas, since once again fades on the antennas become more correlated due to the decreasing distance between antenna elements. It is obvious from Figs. 2.9 and 2.10 that the performance improvement using the optimum algorithm is more significant than using the suboptimum algorithm when the number of receive antennas is increased. However, this performance gain is obtained at the expense of additional complexity, especially when

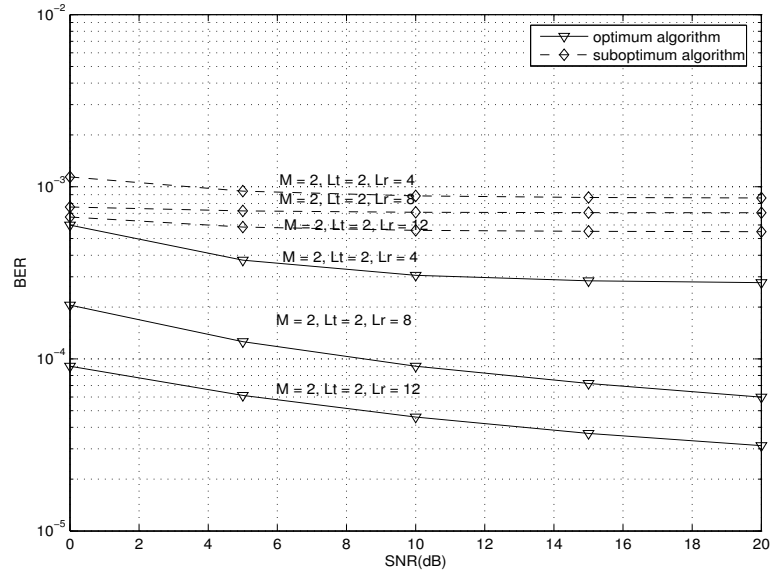


Figure 2.9: BER versus E_b/η_0 for $K = 30$ and $N_s = 32$ per subcarrier with varying number of receive antennas

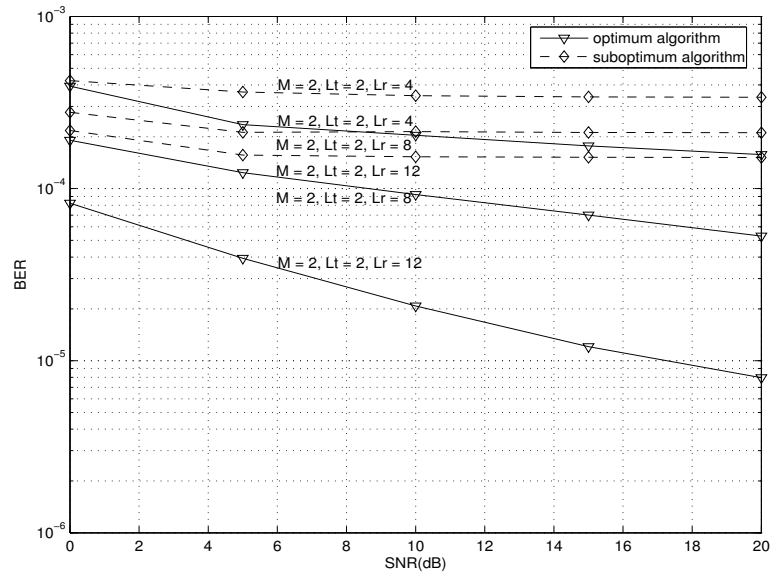


Figure 2.10: BER versus E_b/η_0 for $K = 50$ and $N_s = 32$ per subcarrier with varying number of receive antennas

the number of receive antennas is large. Furthermore, the performance gap between the optimum algorithm and the suboptimum algorithm will decrease if we consider using noisy channel estimates instead of the perfect channel state information. This is because the optimum algorithm needs the channel state information of all the users to calculate the weight vector for each user, whereas each receiver using the suboptimum algorithm only needs its own channel state estimate. Lastly, from Appendix A, it is obvious that the number of calculations involved in the optimum algorithm is greater than that involved in the suboptimum algorithm.

2.5 Conclusion

We proposed an MC-DS-CDMA system employing multiple antennas at both the mobile and the base station. Maximal ratio transmission and adaptive beamforming reception are used to achieve the maximum received SNR for the desired user in a multiple access channel with correlated Rayleigh fading. The conditional SNR is analytically derived and the average BER is investigated via simulation. By varying the number of transmit antennas, receive antennas and subcarriers, we find a tradeoff between obtaining diversity gain against fading and MAI suppression. In a spatially correlated Rayleigh fading channel, as long as the interferers arrive from directions uniformly distributed over $[-\frac{\pi}{2}, \frac{\pi}{2}]$, using more receive antennas is preferred in a dynamic wireless system, since the effect of wireless environment changes (e.g., when the number of active users is varying and/or the accuracy of power control is varying) on the performance is smaller with more rather than less receive antennas. The benefit of using only a single transmit antenna is easier implementation in a small mobile unit. However, when the number of active users is stable and/or accurate power control is always maintained, using two independent transmit antennas with a smaller number of receive antennas is preferred.

2.6 Acknowledgements

Parts of the material in this chapter have appeared in Yan Zhang, Laurence B. Milstein, and Paul H. Siegel, : “The tradeoff between diversity gain and interference suppression in a MIMO MC-CDMA system,” *IEEE Trans. Commun.*, vol. 53, no. 4, pp 623-631, Apr. 2005. The dissertation author was the primary author of these papers.

Appendix 2.A The Computational Complexity Analysis for the Optimum and Suboptimum Algorithms

1. The computation complexity of suboptimum algorithm

(a) the number of multiplications: $K (ML_R^2L_T + M^2L_RL_T)$;

(b) the number of divisions: $K (ML_R + ML_T)$;

(c) the number of other operations: KM eigenvalue decomposition of a square matrix of size L_R .

2. The computation complexity of optimum algorithm depends on the number of iterations for convergence. In each iteration,

(a) the number of multiplications: $K (M^3L_R^3L_T + M^3L_R^3L_T^2 + 2M^3L_R^2L_T + M^2L_RL_T)$;

(b) the number of divisions: $K (ML_R + ML_T)$;

(c) the number of other operations: K eigenvalue decomposition and inversion of a square matrix of ML_R .

Outside the loop, number of multiplications: $M^3L_R^2L_T$.

According to the simulation results, empirically, the number of iterations ranges from 3 to 11, which depends on the specific value for each parameter, K , M , L_T and L_R . Generally, the number of iterations increases as K or L_R increases.

Chapter 3

Approaching V-BLAST Capacity with Adaptive Modulation and LDPC Encoding

3.1 Introduction

In order to attain the MIMO capacity, it is necessary to signal through the channel's eigenmodes [5] with optimal power and rate allocations across those modes. Thus, a specialized transmit structure is required to perform the eigenmode signaling. It is very challenging to incorporate the MIMO capacity-achieving transceiver structures into existing systems. Instead, a V-BLAST type system is more likely to be adopted in the next generation of wireless systems, since it uses a much simpler transceiver structure with conventional scalar coding.

In a V-BLAST [33] system, a single data stream is encoded and demultiplexed into multiple substreams determined by the number of transmit antennas. Each substream is then modulated into symbols and fed to its respective antenna. The N_t transmit antennas operate at the same symbol rate with synchronized symbol timing. At the receiver side, each of the N_r receive antennas receives the signals radiated from all N_t transmit antennas. A successive detection technique is used to separate the signal from each transmit antenna.

Since it allocates equal power and equal rate to every transmit antenna, the open-loop V-BLAST can only attain a fraction of the MIMO capacity [29]. To overcome this problem, rate

and power adaptations at each antenna are introduced in [39, 49, 50]. It has been theoretically proven in [39] that V-BLAST with Per-Antenna Rate Control (PARC) achieves the performance of an open-loop scheme with space-time trellis codes [47]. In [39], the authors proposed several schemes with rate and/or power adaptations, and obtained the corresponding capacity results based on the assumption that capacity-achieving channel codes are employed. However, for an uncoded system, there is a coding penalty Γ equal to 3.333 at a BER of 10^{-3} . The problem left open is how to design channel codes to compensate the coding penalty in a practical system. A general channel coding and decoding strategy is presented in [51] to achieve the capacity of fading channel when the channel information is available at both the transmitter and the receiver. Here, we explore the use of adaptive coded-modulation techniques, specifically using LDPC codes mapped to Gray-labelled modulation constellations, to approach the MIMO capacity.

In [52], the density evolution technique is used to analyze the performance of message-passing decoders on a binary-input symmetric AWGN channel, enabling the accurate determination of the noise thresholds of LDPC code ensembles. The application of the concentration theorem and the density evolution technique to determine the noise threshold of LDPC code ensembles is simplified by the symmetry of the channel and the decoding algorithm. Specifically, under appropriate symmetry conditions, it suffices to consider the performance of the all-zeroes codeword [52].

Our objective is to develop a similar algorithmic approach for the analysis of LDPC component codes for this V-BLAST type system with adaptive modulation. However, we cannot apply the density evolution technique and the concentration theorem to this scheme directly because, in general, the equivalent binary-input component channels are not symmetric. However, as shown in [53], by introducing i.i.d channel adapters, we can force the symmetry of the equivalent binary-input component channels. Thus, the analysis and design of binary LDPC codes are greatly simplified.

An optimum successive detection (OSD) [36], parametrized by feedforward and feedback equalization vectors, is employed to detect the transmitted vector on a per-substream basis in our V-BLAST type system. In the OSD algorithm, the feedforward vectors and feedback vectors are obtained under the assumption that the detected symbols are error-free. However, this condition is not realistic, and error propagation causes system performance degradation. Further, if channel codes are used, the decoding delay incurred by the successive detection could be intolerable, even in a system with a small number of antennas and channel codes with moderate

block lengths. To overcome these problems, we choose a more practical demodulation method based on parallel soft interference cancellation (PIC) followed by MMSE linear filtering [40].

For the non-ergodic fading channel, the outage capacity, which reflects the tradeoff between outage probability and supportable rate, is used as an information theoretic measure. Outage capacity is appropriate if the system delay constraint is shorter than the channel coherence time, meaning that one cannot average over the fades. This condition is true for most of the practical communication systems. In this chapter, we calculate the outage capacity within the prescribed constraints on outage and power, and design optimal LDPC codes with adaptive modulation to approach the outage capacity for a 2×2 V-BLAST type system.

The outline of this chapter is as follows. In Section 3.2, we illustrate the successive detection used in this V-BLAST type system and show the ergodic capacity using OSD. In Section 3.3, we present the channel code design for this adaptive modulation system, i.e., the LDPC codes designed with the density evolution technique, which generates the maximal achievable rate code and the corresponding degree distribution pair for a target BER of 10^{-5} . We introduce parallel interference cancellation (PIC) followed by MMSE filtering to reduce the decoding delay and the effect of error propagation in Section 3.4. We adopt rate-compatible punctured LDPC codes to simplify the code design, and compare numerical and simulation performance results of the system using optimally designed codes with those of the system using punctured codes in Section 3.5. In Section 3.6, we extend the discussion to a non-ergodic fading process, and design the optimal LDPC codes with the adaptive modulation to approach the outage capacity of the system with a specific outage constraint. Finally, we draw the conclusions in Section 3.7.

Notation: underscore denotes column vectors, $(\cdot)^T$ and $(\cdot)^\dagger$ denote transpose and conjugate transpose, respectively, $|\cdot|^2$ denotes vector norm while $|\cdot|$ denotes cardinality and I_N denotes an $N \times N$ identity matrix. The subscripts and superscripts of vectors are to be interpreted as follows: $\underline{x}_m^n = (x_m, \dots, x_n)$ for $m \leq n$. Throughout the chapter, we use symbols Tx and Rx for transmit antenna and receive antenna, respectively.

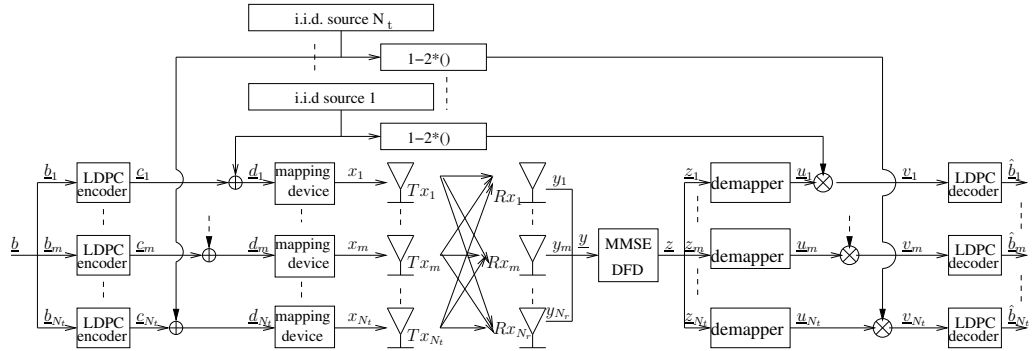


Figure 3.1: Adaptively modulated LDPC encoded V-BLAST type system

3.2 System Model

3.2.1 Per-antenna-coded V-BLAST

To facilitate rate adaptation for each transmit antenna, we extend the original V-BLAST structure [33] with per-antenna-coded, i.e., we use a distinct encoder for each transmit antenna. As shown in Fig. 3.1, each substream consists of independent coded symbols x_m , $m = 1, 2, \dots, N_t$, each with transmit power $p_m = E|x_m|^2$, under a total power constraint $\sum_{m=1}^{N_t} p_m \leq P_T$. Here, we allocate equal power to each transmit antenna, since power adaptation only improves the capacity slightly, at the expense of an exhaustive computer search if rate adaptation is already employed [39]. Furthermore, we assume that each transmitted symbol is drawn from an M -QAM constellation and that different constellations can be used for different substreams.

3.2.2 Channel Model

We let H be a random channel matrix corresponding to a block fading channel. Furthermore, flat fading is assumed, thus the channel gain matrix is written as

$$H = \begin{bmatrix} h_{1,1} & \cdots & h_{1,N_t} \\ \vdots & \ddots & \vdots \\ h_{N_r,1} & \cdots & h_{N_r,N_t} \end{bmatrix}$$

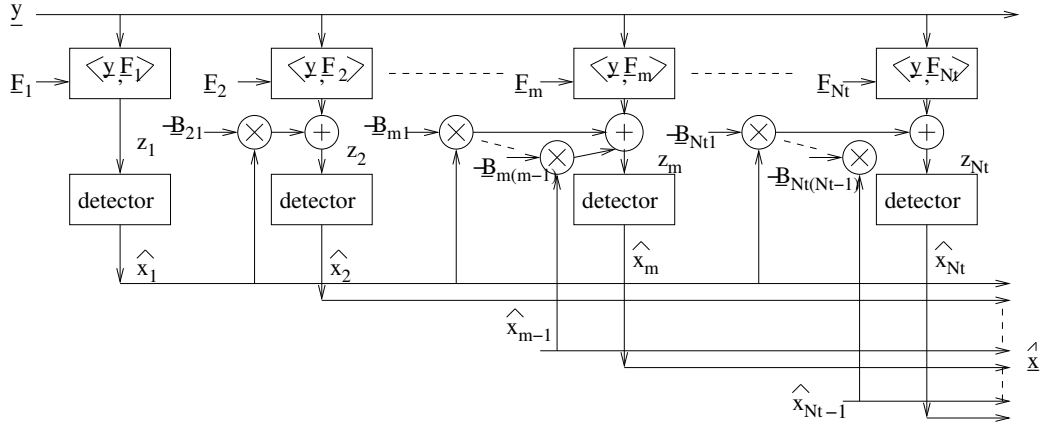


Figure 3.2: MMSE decision feedback detector

where $h_{n,m}$ is the channel gain from transmit antenna m to receive antenna n , $n \in [1, N_r]$ and $m \in [1, N_t]$. For a rich scattering environment, the entries of H are i.i.d. zero-mean complex Gaussian random variables, each with unit variance.

We use a discrete-time baseband representation of the detection process for a single transmitted symbol vector, assuming symbol-synchronous receiver sampling and ideal timing. Letting $\underline{x} = (x_1, \dots, x_{N_t})^T$ denote the vector of transmit symbols, we get the corresponding received vector

$$\underline{y} = \sum_{m=1}^{N_t} \underline{h}_m x_m + \underline{n}, \quad (3.1)$$

where \underline{n} is a noise vector with i.i.d. components each having zero-mean and variance η_0 , and $\underline{h}_m = [h_{1,m}, \dots, h_{N_r,m}]^T$ is the channel gain vector associated with transmit antenna m .

3.2.3 Optimum Successive Detection

A successive detector, parametrized by feedforward and feedback equalization vectors, detects the transmitted vector on a per-substream basis in an arbitrary but fixed order [36]. The feedforward equalizers are denoted by the set of length- N_r vectors $\{\underline{F}_m\}_{m=1}^{N_t}$, whereas the feedback equalizers are denoted by the sets of length- N_r vectors $\{\underline{B}_{m1}, \dots, \underline{B}_{m,m-1}\}_{m=2}^{N_t}$, as shown in Fig. 3.2.

In particular, the first symbol x_1 is detected based on $z_1 = \underline{F}_1^\dagger \underline{y}$, while the m -th symbol

x_m , $m > 1$, is detected based on $z_m = \underline{F}_m^\dagger \left(\underline{y} - \sum_{j=1}^{m-1} \underline{B}_{m,j} \hat{x}_j \right)$, where \hat{x}_j denotes the estimate of x_j . If we assume all the detected symbols are error-free, then z_m can be represented as

$$z_m = \underline{F}_m^\dagger \underline{h}_m x_m + \sum_{j=m+1}^{N_t} \underline{F}_m^\dagger \underline{h}_j x_j + \sum_{j=1}^{m-1} \underline{F}_m^\dagger (\underline{h}_j - \underline{B}_{m,j}) x_j + \underline{F}_m^\dagger \underline{n}. \quad (3.2)$$

Optimum successive detection (OSD) [36] maximizes the conditional mutual information between x_m and z_m , conditioned on the detected symbols x_1, \dots, x_{m-1} , $I(x_m, z_m | \underline{x}_1^{m-1})$. Define $H(k) \triangleq [\underline{h}_k, \dots, \underline{h}_{N_t}]$ and $P(k) \triangleq \text{diag}[p_k, \dots, p_{N_t}]$ as the partial matrix of H and $P = \text{diag}[p_1, \dots, p_{N_t}]$, respectively. For the m th substream, the optimal feedback vectors are given by

$$\underline{B}_{m,j}^{\text{opt}} = \underline{h}_j \quad (3.3)$$

for $1 \leq j \leq m-1$ and the optimal feedforward vector is obtained as

$$\underline{F}_m^{\text{opt}} = \left(H(m+1)P(m+1)H^H(m+1) + \eta_0 I_{N_r} \right)^{-1} \underline{h}_m \quad (3.4)$$

Using OSD, the instantaneous capacity of the m -th antenna is given by [36]

$$C_m^{\text{osd}} = \log_2 \left(1 + p_m \underline{h}_m^\dagger \left(H(m+1)P(m+1)H^\dagger(m+1) + \eta_0 I_{N_r} \right)^{-1} \underline{h}_m \right) \quad (3.5)$$

and the corresponding ergodic capacity can be obtained by taking the expectation over H .

3.2.4 Adaptive Modulation

Note that the idealized capacity results assume an infinite-length code to achieve infinitesimally small bit error rates, whereas in an actual system, only finite-length coding with non-zero error rates are feasible. Further, the idealized results assume a continuous rate set, but in an actual system, only rates from a discrete rate set are feasible.

For a practical coding scheme with a nonzero target BER, the rate of the m -th antenna, R_m , is determined by the specific coding method, modulation constellation and the target BER. LDPC codes are used in our scheme, in conjunction with M -QAM constellations. Specifically, once we obtain C_m^{osd} from Eq. 3.3, we use a $2^{\lceil C_m^{\text{osd}} \rceil + 1}$ -QAM constellation for the m -th substream, where $\lceil x \rceil$ is the smallest integer greater than or equal to x . Then, for a fixed SNR level, the density evolution technique is used to obtain the LDPC code with the maximal achievable rate, $r_{m,c}$, such that the BER of the system is below the target BER. Therefore, the rate for

the m -th substream is $R_m = (\lceil C_m^{osd} \rceil + 1) r_{m,c}$. The ergodic capacity for the m -th substream can be obtained by averaging R_m over a large number of channel realizations. The results for different SNR levels are listed in Table 3.5.

3.3 LDPC Code Design

3.3.1 Identically Independent Distributed (i.i.d) Channel Adapters

We now describe the way in which we extend the density evolution technique to this adaptively modulated LDPC-encoded V-BLAST type system. We cannot apply the density evolution technique to this scheme directly because, for the specific Gray mapping we are using, it is easy to see that the equivalent channels are not symmetric. Therefore, the decoding analysis of the all-zeroes codeword alone may not suffice to predict the average decoder behavior.

Here we force the symmetry of the equivalent binary-input component channels by using i.i.d channel adapters. As shown in Fig. 3.1, each i.i.d channel adapter has three modules. The first one is a source which generates the binary symbol vector \underline{t}_m , $m \in [1, N_t]$, according to an i.i.d equiprobable distribution. The second one is a $mod - 2$ adder: $\underline{d}_m = \underline{c}_m \oplus \underline{t}_m$, where \underline{c}_m is the LDPC coded bits vector. The last module is a sign adjuster: $\underline{v}_m = \underline{u}_m \cdot (1 - 2\underline{t}_m)$, where \underline{u}_m is the a posteriori probability (APP) module output and \underline{v}_m is the LDPC decoder input. It is obvious that the last module undoes the effect of the second module. Therefore, each equivalent binary-input channel m , $m \in [1, N_t]$, is transformed into a new binary-input symmetric channel with corresponding input \underline{c}_m and output \underline{v}_m .

3.3.2 Density Evolution

The conditional pdf of z_m , conditioned on the channel gain matrix H , and the transmitted symbols q_j , $q_j \in Q_j$ with $|Q_j| = 2^{s_j}$, for $m \leq j \leq N_t$, is given by

$$p\left(Z_m = z_m \mid X_m = q_m, \underline{X}_{j=m+1}^{N_t} = [q_{m+1}, \dots, q_{N_t}], H\right) = \frac{1}{2\pi\sigma_m^2} \exp\left(-\frac{|z_m - \mu_m|^2}{2\sigma_m^2}\right), \quad (3.6)$$

where $z_m = \underline{F}_m^\dagger \underline{h}_m q_m + \sum_{j=m+1}^{N_t} \underline{F}_m^\dagger \underline{h}_j q_j + \underline{F}_m^\dagger \underline{n}$, assuming the previously detected symbols are error-free. Here, $\mu_m = \underline{F}_m^\dagger \underline{h}_m q_m + \sum_{j=m+1}^{N_t} \underline{F}_m^\dagger \underline{h}_j q_j$, and σ_m^2 , due to the Gaussian noise,

is obtained as

$$\begin{aligned}
\sigma_m^2 &= \frac{1}{2} \text{Var} \left(\underline{F}_m^\dagger \underline{n} \right) \\
&= \frac{1}{2} \underline{F}_m^\dagger E \left\{ \underline{n} \underline{n}^\dagger \right\} \underline{F}_m \\
&= \eta_0 \underline{F}_m^\dagger \underline{F}_m.
\end{aligned} \tag{3.7}$$

By averaging over $\underline{X}_{j=m+1}^{N_t} = [q_{m+1}, \dots, q_{N_t}]$, where $[q_{m+1}, \dots, q_{N_t}] \in Q_{m+1} \times \dots \times Q_{N_t}$, we get the conditional pdf

$$\begin{aligned}
p(Z_m = z_m | T) &\stackrel{(a)}{=} \sum_{\underline{X}_{j=m+1}^{N_t} \in Q_{m+1}^{N_t}} p(Z_m = z_m, \underline{X}_{j=m+1}^{N_t} | T) \\
&\stackrel{(b)}{=} \sum_{\underline{X}_{j=m+1}^{N_t} \in Q_{m+1}^{N_t}} p(Z_m = z_m | S) p(\underline{X}_{j=m+1}^{N_t} | T) \\
&\stackrel{(c)}{=} \sum_{\underline{X}_{j=m+1}^{N_t} \in Q_{m+1}^{N_t}} \prod_{j=m+1}^{N_t} p(X_j | H) p(Z_m = z_m | S) \\
&\stackrel{(d)}{=} \frac{1}{2^{\sum_{j=m+1}^{N_t} s_j}} \sum_{\underline{X}_{j=m+1}^{N_t} \in Q_{m+1}^{N_t}} p(Z_m = z_m | S)
\end{aligned} \tag{3.8}$$

where T and S denote the events $\{X_m = q_m, H\}$ and $\{X_m = q_m, \underline{X}_{j=m+1}^{N_t}, H\}$, respectively. $Q_{m+1}^{N_t}$ denotes the subspace $Q_{m+1} \times \dots \times Q_{N_t}$. To arrive at (a) and (b), the theorem of total probability and the definition of conditional probability are applied; respectively. Equation (c) holds because, conditioned on the channel gain matrix H , the transmitted symbols from distinct antennas are independent of each other. Therefore the joint probability $p(X_{j=m+1}^{N_t} | X_m = q_m, H)$ is independent of X_m , and $p(X_{j=m+1}^{N_t} | H)$ can be further factored into the product of the individual probabilities $\prod_{j=m+1}^{N_t} p(X_j | H)$. The last equation (d) is obtained by the fact that $p(X_j | H) = \frac{1}{2^{s_j}}$, since we assume that the symbols in the chosen constellation are equiprobably transmitted.

We have observed from our numerical calculations that $p(Z_m = z_m | X_m = q_m, H)$ can be approximated as a Gaussian probability density function (pdf) as long as the constellation size is no greater than 16-QAM. Fig. 3.3 shows the actual pdf and the approximated pdf, which validates our approximation. We can explain the phenomenon as follows. In our scheme, M turns out to be less than or equal to 16 for most of the channel realizations, if the SNR is no greater than 10dB. Since the feedforward vector \underline{F}_m is designed to suppress the interference, $\underline{F}_m^\dagger \underline{h}_j x_j$, $j \in [m+1, N_t]$, is much smaller than $\underline{F}_m^\dagger \underline{h}_m x_m$. Thus, the mean values for z_m , for all $m = 1, \dots, N_t$, conditioned on $\{X_m = q_m, \underline{X}_{j=m+1}^{N_t} \in Q_{m+1} \times \dots \times Q_{N_t}, H\}$, are all

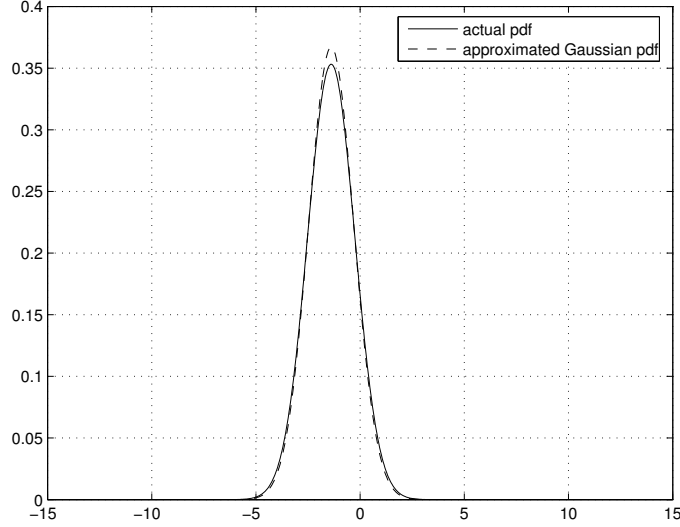


Figure 3.3: Comparisons of the actual pdf to the approximated Gaussian pdf

close to $\underline{F}_m^\dagger \underline{h}_m x_m$, so the mixture of those Gaussian probability density functions can be approximated as another Gaussian probability density function with mean value $\mu'_m = \underline{F}_m^\dagger \underline{h}_m x_m$ and variance slightly larger than σ_m^2 . For the N_t -th substream, $p(Z_{N_t} = z_{N_t} | X_{N_t} = q_{N_t}, H)$ is precisely a Gaussian probability density function, since no interference exists.

Note that $\underline{d}_m = [d_{m,0}, \dots, d_{m,s_m-1}]^T$, as shown in Fig. 3.1, is the input to the mapping device. The l -th component $u_{m,l}$ of \underline{u}_m , for $0 \leq l < s_m$, is the log-likelihood aposteriori probability ratio (LAPPR) of the corresponding input of the m -th mapper, $d_{m,l}$. From the relation between z_m and \underline{u}_m , conditioned on z_m and H , we can calculate the conditional pdf $p(U_{m,l} = u_{m,l} | X_m = q_m, H)$. Using the mapping relationship between q_m and \underline{d}_m , $q_m = \Phi_m(\underline{d}_m)$, where $\Phi_m(\cdot)$ is the mapping function for the m -th antenna, we have

$$\begin{aligned}
 u_{m,l} &= \log \frac{p(d_{m,l} = 0 | Z_m = z_m, H)}{p(d_{m,l} = 1 | Z_m = z_m, H)} \\
 &= \log \frac{p(Z_m = z_m | d_{m,l} = 0, H)}{p(Z_m = z_m | d_{m,l} = 1, H)} \\
 &= \log \frac{E_{q_m \in \Omega_{m,l}^0} p(Z_m = z_m | X_m = q_m, H)}{E_{q_m \in \Omega_{m,l}^1} p(Z_m = z_m | X_m = q_m, H)}
 \end{aligned} \tag{3.9}$$

where $\Omega_{m,l}^i$ denotes the set of symbols of $\Phi_m(\underline{d}_m)$, for which $d_{m,l} = i$, for $i = 0, 1$.

To apply the density evolution technique, we need \bar{v}_m , the LAPPR value of \underline{c}_m , to be

used as the decoder input of the m -th component code. Denote $f_{m,l}^0(v_{m,l})$ as the initial density of $v_{m,l}$, and define $f_m^0(\bar{v}_m) \triangleq \left(\sum_l f_{m,l}^0\right)/s_m$ as the initial density of \bar{v}_m . We also define the following events: $E_1^{(i)} \triangleq \{d_{m,l} = i, H\}$ and $E_2^{(i)} \triangleq \{d_{m,l} = i, (d_{m,k})_{k=0, k \neq l}^{s_m-1}, H\}$. Noting that $v_{m,l} = u_{m,l}(1 - 2d_{m,l})$ (see Fig. 3.1), we obtain

$$\begin{aligned}
f_{m,l}^0(v_{m,l}) &\stackrel{(a)}{=} p(V_{m,l} = v_{m,l} | H) & (3.10) \\
&\stackrel{(b)}{=} \frac{1}{2} \left[p(U_{m,l} = v_{m,l} | E_1^{(0)}) + p(U_{m,l} = -v_{m,l} | E_1^{(1)}) \right] \\
&\stackrel{(c)}{=} \frac{1}{2} \sum_{D_{m(l)}} \left[p(U_{m,l} = v_{m,l}, (d_{m,k})_{k=0, k \neq l}^{s_m-1} | E_1^{(0)}) \right. \\
&\quad \left. + p(U_{m,l} = -v_{m,l}, (d_{m,k})_{k=0, k \neq l}^{s_m-1} | E_1^{(1)}) \right] \\
&\stackrel{(d)}{=} \frac{1}{2} \sum_{D_{m(l)}} \left[p(U_{m,l} = v_{m,l} | E_2^{(0)}) \cdot Pr((d_{m,k})_{k=0, k \neq l}^{s_m-1} | E_1^{(0)}) \right. \\
&\quad \left. + p(U_{m,l} = -v_{m,l} | E_2^{(1)}) \cdot Pr((d_{m,k})_{k=0, k \neq l}^{s_m-1} | E_1^{(1)}) \right] \\
&\stackrel{(e)}{=} \frac{1}{2} \sum_{D_{m(l)}} \left[p(U_{m,l} = v_{m,l} | E_2^{(0)}) \cdot Pr((d_{m,k})_{k=0, k \neq l}^{s_m-1} | H) \right. \\
&\quad \left. + p(U_{m,l} = -v_{m,l} | E_2^{(1)}) \cdot Pr((d_{m,k})_{k=0, k \neq l}^{s_m-1} | H) \right] \\
&\stackrel{(f)}{=} \frac{1}{2} \prod_{k=0, k \neq l}^{s_m-1} Pr(d_{m,k}) \\
&\quad \cdot \sum_{D_{m(l)}} \left[p(U_{m,l} = v_{m,l} | E_2^{(0)}) + p(U_{m,l} = -v_{m,l} | E_2^{(1)}) \right] \\
&\stackrel{(g)}{=} \frac{1}{2^{s_m}} \sum_{D_{m(l)}} \left[p(U_{m,l} = v_{m,l} | E_2^{(0)}) + p(U_{m,l} = -v_{m,l} | E_2^{(1)}) \right]
\end{aligned}$$

where $\sum_{D_{m(l)}}$ denotes the multiple summations over binary variables $d_{m,0}, \dots, d_{m,l-1}, d_{m,l+1}, \dots, d_{m,s_m-1}$, $\sum_{d_{m,0}} \dots \sum_{d_{m,l-1}} \sum_{d_{m,l+1}} \dots \sum_{d_{m,s_m-1}}$. Equations (b) and (c) are derived from the definition of $E_1^{(i)}$ and the theorem of total probability. Equation (d) is result of the definition of $E_2^{(i)}$ and the definition of conditional probability. Equations (e) - (g) hold based on the assumptions that the coded bits in a symbol are independent and each symbol is equiprobably chosen from the constellation. Generally, unlike in the BPSK case, $V_{m,l}$ is no longer a Gaussian random variable. Note that the numerator and the denominator in Equation 3.7 are a mixture of a group of Gaussian pdfs. However, the mean value of z_m , conditioned on different $q_m \in \Omega_{m,l}^0$ or $q_m \in \Omega_{m,l}^1$, is no longer approximately independent of the value of q_m , and the

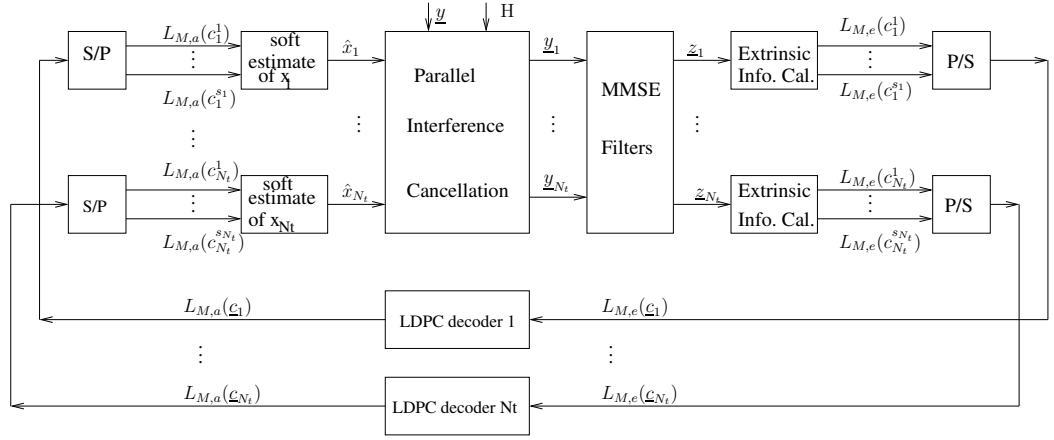


Figure 3.4: Parallel interference cancellation and MMSE filtering demapper

difference is determined by the specific mapping method and labelling. Simulation results show that $E_{q_m \in \Omega_{m,l}^0} p(Z_m = z_m | X_m = q_m, H)$ and $E_{q_m \in \Omega_{m,l}^1} p(Z_m = z_m | X_m = q_m, H)$, regarded as functions of the random variable Z_m , may have multiple maximal values, and therefore they are certainly non-Gaussian, so a Monte Carlo method has to be used. We first obtained the numerical distribution of $U_{m,l}$, conditioned on H , and used it as the initial density of the LAPPs of the augmented binary-input component channel. The density evolution technique was then employed to derive the maximal achievable rate $r_{m,c}$ and the corresponding degree distribution pair $(\lambda_m(x), \rho_m(x))$ for the m -th LDPC component code.

3.4 PIC-MMSE Demodulation

In a practical system, feedback errors cause significant system performance degradation, since $\{F_i\}_{i=1}^{N_t}$ and $\{[B_{j,i}]_{i=1}^{j-1}\}_{j=2}^{N_t}$ are no longer the optimal vectors for maximizing the conditional mutual information, $I(x_m, z_m | x_1^{m-1})$, when perfect feedback does not hold. Another prominent disadvantage of the OSD algorithm is the impractical decoding delay if channel codes are used, since the feedback symbol estimates are generated by re-encoding the decoded information bits and re-modulating the re-encoded bits. Even in a system with a small number of antennas and channel codes with moderate block lengths, this delay could be intolerable.

3.4.1 Parallel Soft Interference Cancellation

A more practical demodulation method based on parallel soft interference cancellation and MMSE linear filtering [40] is selected to mitigate those problems. The PIC-MMSE demapper is illustrated in Fig. 3.4. First, based on the *a priori* information of the coded bits, $L_{M,a}(\underline{c}_i)$, $i = 1, \dots, N_t$, we compute the soft estimates $\hat{\underline{x}} = [\hat{x}_1, \dots, \hat{x}_{N_t}]^T$ of the transmitted vector $\underline{x} = [x_1, \dots, x_{N_t}]^T$, i.e.,

$$\begin{aligned}\hat{x}_m &= E[x_m] = \sum_{m=0}^{2^{s_m}-1} q_m \Pr(x_m = q_m) \\ &= \sum_{m=0}^{2^{s_m}-1} q_m \prod_{n=0}^{s_m-1} \Pr(c_{m,n} = b_{m,n}),\end{aligned}\quad (3.11)$$

where $q_m = \Phi_m(\underline{c}_m)$. Since we assume the coded bits are independent, the symbol error probability can be factored into the product of the bit error probabilities, where $\Pr(c_{m,n} = b_{m,n}) = \exp((1 - b_{m,n}) \cdot L_{M,a}(c_{m,n})) / (1 + \exp(L_{M,a}(c_{m,n})))$, $b_{m,n} = 0, 1$ [53].

Define $\hat{\underline{x}}_m \triangleq [\hat{x}_1, \dots, \hat{x}_{m-1}, 0, \hat{x}_{m+1}, \dots, \hat{x}_{N_t}]$. To estimate transmitted symbol x_m , a soft interference cancellation is performed on the received vector \underline{y} , and we obtain

$$\underline{y}_m = \underline{y} - H\hat{\underline{x}}_m. \quad (3.12)$$

3.4.2 MMSE Linear Filtering

To further suppress the residual interference plus noise, we feed each \underline{y}_m to an MMSE filter and get

$$z_m = \underline{w}_m^\dagger \cdot \underline{y}_m. \quad (3.13)$$

Here, \underline{w}_m is chosen to minimize the mean square error between the symbol x_m and the filter output z_m , and, as in [40], we have

$$\begin{aligned}\underline{w}_m &= E[\underline{y}_m \underline{y}_m^\dagger]^{-1} E[\underline{y}_m z_m^*] \\ &= \left(\underline{h}_m \underline{h}_m^\dagger + \sum_{j=1, j \neq m}^{N_t} \frac{E|x_j|^2 - |\hat{x}_j|^2}{E|x_m|^2} \underline{h}_j \underline{h}_j^\dagger + \frac{2\sigma^2}{E|x_m|^2} I_{N_r} \right)^{-1} \underline{h}_m,\end{aligned}\quad (3.14)$$

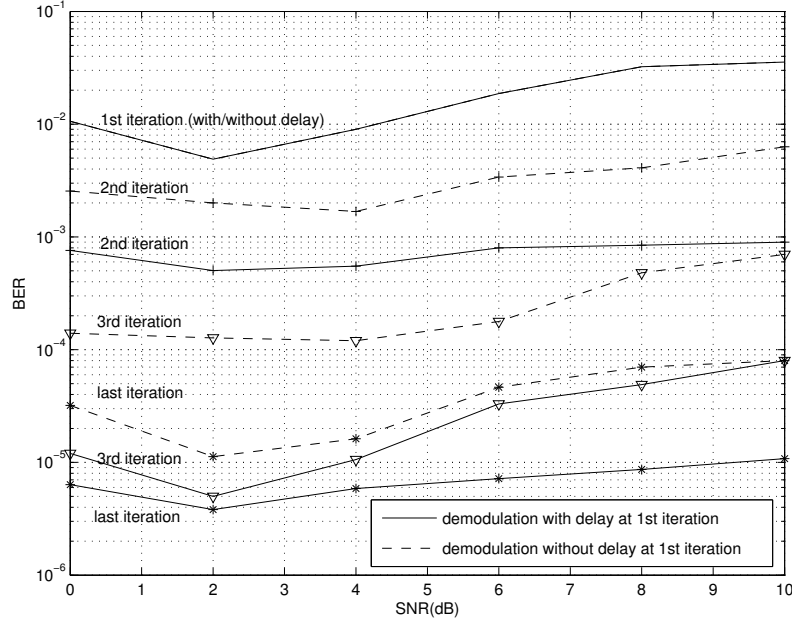


Figure 3.5: Comparisons of the decoding convergence rate of Tx1 using modified PIC algorithm to that using the original PIC algorithm

where $E |x_m|^2 = \sum_{k=0}^{2^{s_m}-1} |q_k|^2 Pr(x_m = q_k)$. Using an approximation introduced in [40], we regard z_m as the output of an equivalent AWGN channel with input x_m , i.e.,

$$z_m = \beta_m x_m + \xi_m, \quad (3.15)$$

where $\beta_m = E[z_m x_m^*] / E|x_m|^2 = w_m^\dagger h_m$, and ξ_m is a zero-mean Gaussian random variable with variance $\nu_m^2 = E|z_m - \beta_m x_m|^2 = E|x_m|^2 \cdot (\beta_m - |\beta_m|^2)$.

So the extrinsic information of each coded bit is given by [53]

$$L_{M,e}(c_{m,n}) = \ln \frac{\sum_{q_m \in \Omega_{m,n}^1} p(z_m | q_m) \cdot \exp\left(\sum_{j \in J_n(Q_m)} L_{M,a}(c_{m,j})\right)}{\sum_{q_m \in \Omega_{m,n}^0} p(z_m | q_m) \cdot \exp\left(\sum_{j \in J_n(Q_m)} L_{M,a}(c_{m,j})\right)} \quad (3.16)$$

where $\Omega_{m,n}^b$ is the set of symbols with $c_{m,n} = b$, and $J_n(Q_m)$ is the set of indices within symbol q_m such that $c_{m,j} = 0, j = 0, \dots, s_m - 1, j \neq n$. $L_{M,e}(c_m)$ is used by the m -th LDPC decoder as the *a priori* information $L_{D,a}(c_m)$. For the LDPC decoder, we use the standard sum-product algorithm to generate the extrinsic information $L_{D,e}(c_m)$. In this way, the soft symbol estimates can be improved by further iterations between the demodulator and the LDPC decoder.

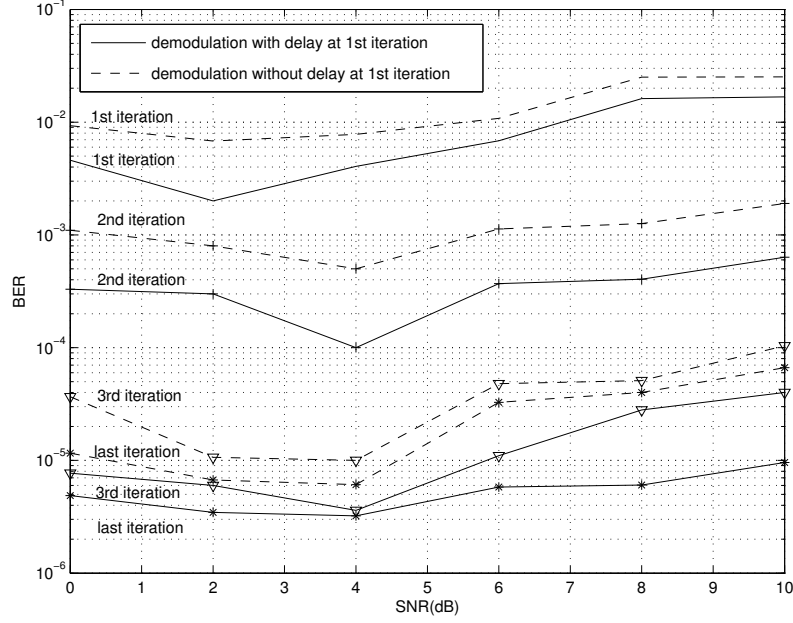


Figure 3.6: Comparisons of the decoding convergence rate of Tx2 using modified PIC algorithm to that using the original PIC algorithm

Finally, to speed up the convergence rate, we modified the algorithm above as follows. In the first iteration, since no symbol estimates are available if we demodulate the N_t branches simultaneously, the m -th demapper cannot get any *a priori* information, $L_{D,a}(c_m)$, from the LDPC decoders, so no cancellation is done. Hence, we delay the demodulation of the m -th antenna for $m - 1$ time slots, so that the symbol estimates, $\hat{x}_1, \dots, \hat{x}_{m-1}$, of the tentatively demodulated symbols x_1, \dots, x_{m-1} can be used as the feedback for interference cancellation. In the simulations, we find that this modification enhances the convergence significantly, as shown in Fig 3.5. We can explain this phenomenon by noting that the extrinsic information from the m -th demodulator, $L_{M,e}(c_m)$, is improved due to the interference cancellation, which in turn accelerates the convergence of the LDPC decoding. As a consequence, the extrinsic information from the LDPC decoder, $L_{D,e}(c_m)$, is better and thus helps the demodulation during the next iteration. We found that only 3-4 iterations between the demapper and the LDPC decoder suffice to achieve the target BER 10^{-5} , when we use the delayed demodulation at the first iteration, as shown in Fig. 3.5 and 3.6.

3.5 Rate-compatible Puncturing of LDPC Codes

To implement this adaptive scheme, we have to design for each antenna, distinct LDPC codes optimal for the corresponding channel realizations. Since these optimal LDPC codes may have different degree distributions, the transmitter (receiver) has to adjust the encoder (decoder) structure for each channel realization. This requirement increases the complexity of the transceiver. To reduce the transceiver complexity, we use the idea of rate-compatible punctured LDPC codes proposed in [54] so that only one code, with a given degree distribution $(\lambda(x), \rho(x))$ and rate $r(\lambda, \rho)$, is used for all the channel realizations.

The basic idea of puncturing is summarized as follows: For the puncturing problem, variable nodes having the same edge degree are put into a group denoted as G_j , $j \in [2, d_l]$. We randomly puncture a proportion $\pi_j^{(0)} \in [0, 1]$ of symbols in G_j , where $\pi_j^{(0)}$ is optimized according to the desired puncturing fraction. The $\pi_j^{(0)}$'s specify a puncturing distribution, $\pi^{(0)}(x) = \sum_{j=2}^{d_l} \pi_j^{(0)} x^{j-1}$. The puncturing fraction $p^{(0)}$ is defined as the ratio of the number of punctured variable nodes to the number of variable nodes, and is expressed as

$$p^{(0)} = \frac{\sum_{j=2}^{d_l} \pi_j^{(0)} n_j}{n}, \quad (3.17)$$

where $n_j = |G_j|$ is the number of variable nodes with degree j . Hence, the code rate of a punctured LDPC code is determined by a three-tuple distribution $(\lambda(x), \rho(x), \pi^{(0)}(x))$, i.e.,

$$r(\lambda, \rho, \pi^{(0)}) = \frac{r(\lambda, \rho)}{1 - p^{(0)}}. \quad (3.18)$$

Based on the convergence condition using the *Gaussian Approximation* (GA) [44], the puncturing coefficients $\pi_j^{(0)}$'s are optimized to achieve the target puncturing fraction $p^{(0)}$ at a fixed SNR threshold. We choose the code with the lowest non-zero rate, i.e., the code designed for the poorest channel realization, as the base code for each transmit antenna, and puncture the left degree distribution $\lambda(x)$ to get the higher rate punctured codes.

The maximal achievable rates with the punctured LDPC codes at different SNR levels for each antenna are listed in Table 3.6. The results of puncturing patterns to get the higher rates codes from a base code are also listed in Table 3.1 – Table 3.4. Comparing the numerical results in Fig. 4.5, we find that the maximal rates achieved by rate-compatible punctured LDPC codes are slightly worse than the maximal rates achieved by optimizing the LDPC codes for each SNR level.

Table 3.1: Degree distribution of a rate = 0.25 LDPC code

d_l	2	3	6	10	19	50
λ_j	0.2816	0.1869	0.1813	0.0719	0.1427	0.1356
d_r	5	6				
ρ_j	0.6415	0.3585				

Table 3.2: Puncturing proportions of rate-compatible punctured LDPC codes

rate \ $\pi_j^{(0)}$	2	3	6	10	19	50
0.30	0.2376	0.1206	0.0005	0.0358	0.0260	0.1332
0.35	0.2778	0.2881	0.3764	0.0667	0.1930	0.4702
0.40	0.3934	0.3131	0.4018	0.3103	0.4569	0.4903
0.45	0.4642	0.3966	0.4239	0.4809	0.4987	0.4981
0.50	0.4875	0.4825	0.5439	0.5717	0.5994	0.5998

Table 3.3: Degree distribution of a rate = 0.20 LDPC code

d_l	2	3	6	10	19	50
λ_j	0.3411	0.2027	0.1950	0.0589	0.1523	0.05
d_r	4	5				
ρ_j	0.5685	0.4315				

Table 3.4: Puncturing proportions of rate-compatible punctured LDPC codes

rate \ $\pi_j^{(0)}$	2	3	6	10	19	50
0.25	0.3070	0.2027	0.1950	0.0589	0.1523	0.0
0.30	0.3392	0.3008	0.3494	0.3008	0.4836	0.0
0.35	0.4427	0.3808	0.4422	0.4745	0.4953	0.0
0.40	0.5017	0.4826	0.5233	0.5398	0.5499	0.0
0.45	0.5521	0.5441	0.5967	0.5960	0.5988	0.0
0.50	0.6152	0.5486	0.6252	0.6454	0.6485	0.0

Table 3.5: Ergodic capacities and maximal achievable rates of a 2×2 V-BLAST system using optimally designed LDPC codes

SNR	$R1$	$R2$	R	C^{osd}
0	0.6451	0.8662	1.5113	1.6982
2	0.8390	1.1561	1.9951	2.2688
4	1.0893	1.4061	2.4954	2.9287
6	1.3335	1.8497	3.1832	3.7092
8	1.6388	2.3294	3.9682	4.5846
10	2.0819	2.8133	4.8952	5.5513

Table 3.6: Ergodic capacities and maximal achievable rates of a 2×2 V-BLAST system using punctured LDPC codes

SNR	$R1$	$R2$	R	C^{osd}
0	0.6006	0.8118	1.4124	1.6982
2	0.7911	1.0937	1.8848	2.2688
4	1.0308	1.3412	2.3720	2.9287
6	1.2747	1.7365	3.0112	3.7092
8	1.5726	2.2125	3.7851	4.5846
10	1.9970	2.6952	4.7292	5.5513

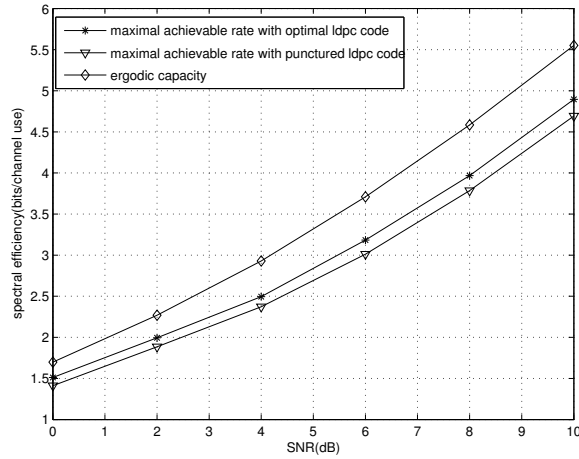


Figure 3.7: Comparisons of the ergodic capacity to the maximal achievable rates of a 2×2 V-BLAST system with optimally designed LDPC codes and to that with rate-compatible punctured LDPC codes

In the BER performance simulations, we fix the number of symbols transmitted from each antenna in each frame to be 2,500, since symbol synchronization is required to implement the parallel interference cancellation. Thus the block length of the LDPC code corresponding to each channel realization has to be adjusted due to the distinct constellation. For example, if *QPSK* is the chosen constellation, then the block length of the LDPC code is 5,000.

The same rule applies to the rate-compatible punctured LDPC codes. In the latter case, the block length of the base code has to be adjusted by the puncturing portion $p^{(0)}$ corresponding to each channel realization. Specifically, if *QPSK* is the chosen constellation and the desirable punctured code rate is 0.4, then the block length of the base code of rate 0.25, N_u , is calculated as $N_u = \frac{N_p \times r(\lambda, \rho, \pi^{(0)})}{r(\lambda, \rho)} = \frac{5000 \times 0.4}{0.25} = 8,000$, where N_p is the code block length after puncturing. To realize the symbol synchronization in the punctured code case, the block length of the base code has to vary from frame to frame to accommodate the chosen constellation and puncturing fraction for each frame. Details of designing codes with different rates and different block lengths from one underlying mother code will be discussed in next chapter.

The BER performance of the system with the optimal LDPC codes and rate-compatible punctured LDPC codes for each SNR level are shown in Fig. 3.8. Generally, the system with the optimal LDPC codes outperforms the system with the punctured LDPC codes. It is observed that the BER performances of both systems appear to degrade in the high SNR region. We offer the following as a possible explanation. The BER performance obtained by the DE technique reflects the average performance of the code ensemble. However, the performance of a bad code construction deviates from the average. As the SNR increases, the channel code rate tends to be high, which makes the performance more sensitive to the specific code construction since it is harder to design the loop-free graph structure.

3.6 Outage Capacity of V-BLAST

3.6.1 Outage Capacity

The outage capacity is defined as the maximum rate below which arbitrarily reliable transmission is possible for a given percentage of channel realizations. In the context of the V-BLAST setup, the transmission rate on each of the subchannels has to be no greater than the respective capacities of those subchannels, so that the assumption of perfect feedback can be

justified.

As in the ergodic fading channel case, the outage capacity of V-BLAST can be achieved by Gaussian distributed input symbols [29]. This means that the m -th subchannel capacity is a random variable of the form $\log [1 + \gamma_m]$, where γ_m is the received SNR of the m -th subchannel. If the subchannels are transmitting at rates $\{r_m\}_{m=1}^{N_t}$, the outage probability of the V-BLAST is defined as [55]

$$P_{out}^{VB}(r_1, \dots, r_{N_t}) \triangleq P \left[\log(1 + \gamma_1) \leq r_1 \bigcup \dots \bigcup \log(1 + \gamma_{N_t}) \leq r_{N_t} \right]. \quad (3.19)$$

The outage capacity C_ϵ^{VB} for $\epsilon \in [0, 1]$ is the maximum sum rate at which arbitrarily reliable transmissions are possible for all the subchannels for $(1 - \epsilon) \times 100\%$ of the channel realizations. It is given by [55]

$$C_\epsilon^{VB} = \max_{\{r_m\}_{m=1}^{N_t}} \left\{ \sum_{m=1}^{N_t} r_m : P_{out}^{VB}(r_1, \dots, r_{N_t}) \leq \epsilon \right\}. \quad (3.20)$$

To obtain the outage capacity within the prescribed constraints on outage and power, and to design the optimal LDPC codes with adaptive modulation to approach the outage capacity, we use a 2×2 MIMO system to illustrate the process. The achievable outage region for V-BLAST for a specific channel realization H_t is

$$C^{VB}(H_t) = \{(r_1, r_2) : r_1 \leq \log(1 + \gamma_1(H_t)), r_2 \leq \log(1 + \gamma_2(H_t))\}. \quad (3.21)$$

For the designated rate pair (r_1, r_2) , an outage event occurs if either $r_1 > \log(1 + \gamma_1(H_t))$ or $r_2 > \log(1 + \gamma_2(H_t))$. Define an indicator function $I(H_t)$ by

$$I(H_t) = \begin{cases} 1 & \text{an outage event occurs} \\ 0 & \text{otherwise} \end{cases}. \quad (3.22)$$

For the prescribed outage constraint $P_{out}^{VB}(r_1, r_2) \leq \epsilon$, we get the outage region C_{out}^{VB} by transmitting on N independent channel realizations. Specifically, we compute the achievable outage region for each of the N instances, and record the total number of outages $\sum_{t=1}^N I(H_t)$ for a given rate pair (r_1, r_2) . According to the definitions of outage capacity and outage probability, all the rate pairs (r_1, r_2) in the outage region satisfy the following condition:

$$\sum_{t=1}^N I(H_t) / N \leq \epsilon, \quad (3.23)$$

and the outage capacity is the rate pair with the largest sum rate.

Table 3.7: Outage capacities and maximal achievable rates of a 2×2 V-BLAST system with quasi-static fading

SNR	$R1$	$R2$	R	C^{osd}
5	0.2663	0.5920	0.8583	1.0622
10	0.6622	1.2813	1.9435	2.2159
15	0.9872	1.9685	2.9557	3.9542
20	1.3685	3.6198	4.9883	6.4482

3.6.2 LDPC Code Design for Non-ergodic Fading Channels

Once we have the supportable rate distribution based on N instantaneous capacity pairs, $(\log(1 + \gamma_1(H_t)), \log(1 + \gamma_2(H_t)))$, we can calculate the outage capacities for different outage constraints. Specifically, if $N = 100$ and $\epsilon = 10\%$, then 10 out of the N channel realizations with the lowest sum capacity could be experiencing an outage. The lowest sum capacity corresponding to the remaining channel realizations is regarded as the outage capacity for this specific outage constraint.

Here we use the same methodology as before to determine the modulation size and obtain the LDPC code with maximal achievable rate to approach the outage capacity. For each channel realization H_t , the constellation is chosen by the corresponding capacity, $C_m^{osd}(H_t)$, for each antenna. Then the density evolution technique is used to obtain the maximal achievable rate, $r_{m,c}(H_t)$, for a target BER of 10^{-5} , where $m = 1, 2$ is the index of transmit antenna. Once we get those rate pairs for all the channel realizations, we calculate the supportable sum rate $\sum_m R_{m,f}(H_t) = \sum_m (R_{m,M}(H_t) \times r_{m,c}(H_t))$, and sort them in an non-decreasing order, where $R_{m,M}(H_t)$ is the modulated symbol rate of the m -th antenna for the t -th channel realization. Then we approximate the $N\epsilon + 1$ smallest supportable rate as the maximal achievable sum rate for the 2×2 V-BLAST type system using OSD detection. The outage capacities and maximal achievable sum rates for a given outage probability at different SNR levels are listed in Table 3.7.

3.7 Conclusion

In this chapter, we propose a practical scheme for a V-BLAST type system, in which the MIMO open-loop capacity is approached closely by means of adaptive modulation using

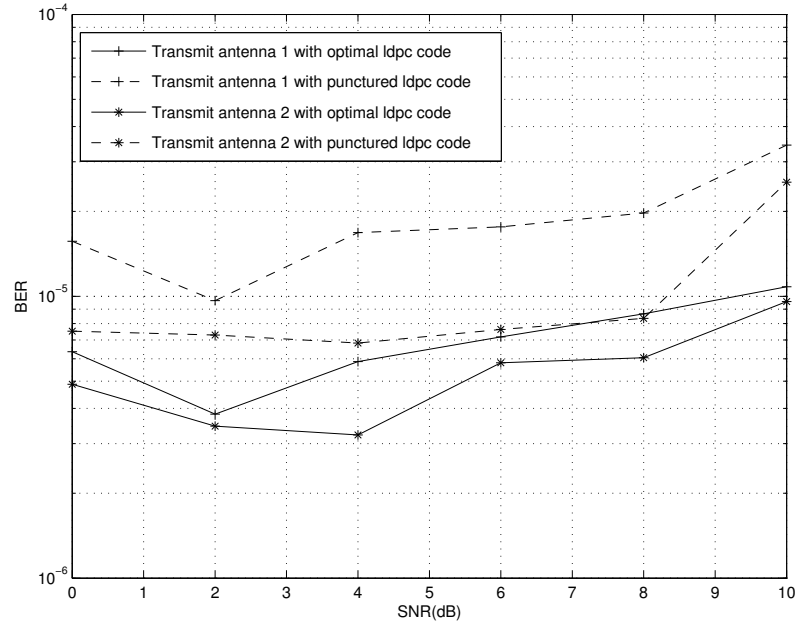


Figure 3.8: BER performance of a 2×2 V-BLAST system with optimally designed LDPC codes and that with rate-compatible punctured LDPC codes

LDPC codes. For each transmit antenna, the constellation is chosen based upon its instantaneous capacity using OSD, and the density evolution technique is employed to determine the maximal achievable rate of the LDPC code in each channel realization.

Numerical results show that the spectral efficiency of the designed system for achieving the target BER is quite close to the capacity of the V-BLAST system with per-antenna rate and power adaptations. To simplify the channel code design, we replace the optimal LDPC codes with rate-compatible punctured LDPC codes for each antenna, at the expense of slightly reduced spectral efficiency.

In performance simulations, we use PIC followed by MMSE filtering to overcome the problems caused by error propagation and the decoding delay in the OSD algorithm. Again, the results show that the system using the optimal codes slightly outperforms the system using the punctured codes. Further, we calculate outage capacities of this adaptively modulated V-BLAST type system with a given outage probability for the non-ergodic fading channels, which is more meaningful for most practical systems. Then to approach the outage capacities, we employ similar methodology to design the LDPC codes and choose the modulation alphabet for each

antenna.

3.8 Acknowledgement

Parts of the material in this chapter have appeared in Yan Zhang, Paul H. Siegel, and Laurence B. Milstein, “Approaching V-BLAST capacity with adaptive modulation and LDPC encoding,” Yan Zhang, Paul H. Siegel, and Laurence B. Milstein, submitted to *IEEE Transactions on Communications*. The dissertation author was the primary author of these papers.

Chapter 4

Flexible-length Rate-compatible Punctured Irregular Repeat-accumulate Code Design

4.1 Introduction

The Vector-LDPC architecture proposed in [56] allows for the implementation of multiple codes of different rates and block lengths in one device, which is very important for channel adaptation. In [42], Wesel et al. applied information nulling and row-combining to the parity check matrix of the mother LDPC code so that codes of different rates and block lengths can be realized from one fundamental code structure. This scheme has an advantage over the Vector-LDPC architecture in that the connections between variable nodes and check nodes are maintained while encoding/decoding codes with different rates and block lengths. However, to avoid a performance degradation due to cycles which may result from combining the rows of the parity check matrix, extra constraints and computations are required to generate the appropriate parity check matrix for the mother code. Moreover, codes of certain rates have worse performance if the rows in the parity check matrices of the higher rate codes are obtained by combining different numbers of rows from the original parity check matrices.

In this chapter, we explore an alternative way to simplify the design and implementation of codes of different rates and block lengths by using irregular repeat-accumulate (IRA)

codes. These codes were proposed by Jin, Khandekar and McEliece proposed IRA codes in [43] [57], who were inspired by the markedly better performance of irregular LDPC codes relative to that of regular LDPC codes.

On binary-input output-symmetric channels, IRA codes have performance competitive with that of optimized irregular LDPC codes, but with much simpler encoder structures. In fact, they can be encoded and decoded with linear complexity in the block length. Thus, IRA codes are an appealing choice for a practical system.

Instead of modifying parity check matrices to generate a family of codes with different rates and block lengths as in [42], we introduce a two-step design process that first shortens a mother IRA code, and then punctures the shortened code. We use the design procedure to generate a family of flexible-length, rate-compatible codes with degree distributions that ensure good performance.

An ensemble of irregular LDPC codes is represented by a Tanner graph consisting of variable nodes and check nodes, with connections between variable nodes and check nodes specified by the degree distribution pair $(\lambda(x), \rho(x))$. Similarly, a Tanner graph of an ensemble of IRA codes consists of variable nodes and check nodes, but with the variable nodes further subdivided into information nodes and parity nodes. In this chapter, we show how this characteristic makes it possible to use one encoder and decoder structures for multiple codes of different rates and block lengths. To demonstrate the practicality of our approach, we apply the proposed code design technique to an adaptively modulated V-BLAST type system that approaches the V-BLAST capacity with a relatively low implementation complexity.

The outline of this chapter is as follows. The motivation for using rate-compatible punctured IRA codes is presented in Section 4.2. We summarize the basics of IRA codes in Section 4.3. The design of flexible-length rate-compatible punctured IRA codes is discussed in Section 4.4. Numerical results are shown in Section 4.5. Conclusions are drawn in Section 4.6.

Notation: $\lceil x \rceil$ denotes the closest integer to a real number x , with the exception that the closest integer greater than x is chosen if there are two closest integers to x .

4.2 Motivation

We use adaptive modulation and capacity-approaching channel codes to increase the system sum rate for a per-antenna-coded V-BLAST type system, as shown in Fig. 4.1. Thus,

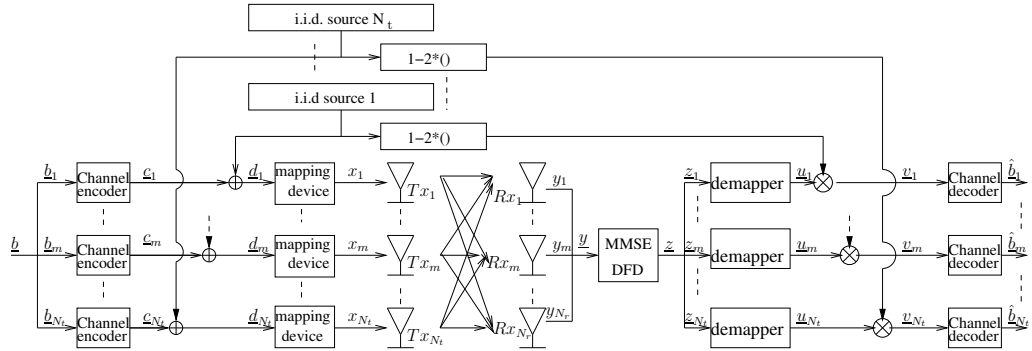


Figure 4.1: Adaptively modulated per-antenna-coded V-BLAST type system

codes with variable rates and variable block lengths should be designed to match different channel realizations. To reduce the implementation complexity and make the scheme appropriate for a practical system, a family of codes derived from one underlying mother code is more desirable than a collection of codes designed individually for different channel realizations.

4.2.1 Limitations of Conventional Punctured LDPC Codes

If the transmitted frame contains a fixed number of symbols, and if the modulation alphabet is adapted to the channel state, then different frames may represent different numbers of coded bits. The approach to designing rate-compatible punctured LDPC codes introduced in [54] optimizes the puncturing distribution based on the mother code degree distribution, and the punctured code length is determined by the mother code length and both code rates. However, in our scheme, the resulting punctured code length is predetermined by the chosen modulation alphabet and the number of symbols in the frame.

Consider the following example, in which we fix the frame length to be 2500 symbols. Now, suppose that for the first channel realization, the designed code rate is 0.25 with 8AMPM modulation. Then, the number of coded bits to be transmitted in this frame is $2500 \times 3 = 7500$. If QPSK is chosen for the second channel realization and the designed code rate is 0.35, the number of coded bits to be transmitted is $2500 \times 2 = 5000$. Due to the relationship between the code rates and code lengths of the mother code and punctured codes, the mother code length has to be adjusted according to the predetermined punctured code lengths.

Specifically, with a rate 0.2 mother code, the puncturing fractions for these two real-

izations are given by $p_1^{(0)} = 1 - \frac{r_b}{r_{d,1}} = 0.2$ and $p_2^{(0)} = 1 - \frac{r_b}{r_{d,2}} = \frac{3}{7}$, where r_b is the mother code rate, and $r_{d,1}$, $r_{d,2}$ are the target code rates, respectively. The puncturing fraction is also defined as the ratio of the number of punctured variable nodes to the total number of variable nodes. For the first channel realization, we see that the length n_b of the mother code must satisfy $(1 - p_1^{(0)}) n_b = 7500$ bits long, implying that $n_b = 9375$. On the other hand, for the second channel realization, the length n_b of the mother code must satisfy $(1 - p_1^{(0)}) n_b = 5000$, implying that $n_b = 8750$. Hence, it is required that the mother code length be adjusted for each specific channel realization according to the chosen modulation alphabet and the target code rate. This also means that the transmitter (resp., receiver) has to be equipped with different encoder (resp., decoder) structures for different channel realizations, since the connections between variable nodes and check nodes in the parity check matrices are different due to the variable code lengths.

4.2.2 Motivation for Using IRA Codes

It is highly desirable in an actual system to encode (resp., decode) multiple codes with different rates and block lengths using one fundamental encoder (resp., decoder). To accomplish this, we use IRA codes, which explicitly distinguish between information bits and parity bits, and a two-phase construction incorporating code shortening and puncturing. Specifically, we use a mother code with the largest block length anticipated, say 10000, and adapt it to different block lengths and code rates required for different channel realizations.

The approach can be illustrated by application to the scenario described above. Rather than directly puncturing the mother code to a higher rate code, we first shorten the code, reducing the number of information bits and the number of parity bits proportionally so that the resulting code length satisfies the constraint imposed by the chosen modulation size, the frame length (in symbols) and the target code rate, while preserving the mother code rate. Furthermore, the number of information bits in each node degree group is reduced in proportion to the respective cardinality so that the optimized degree distribution for the mother code is preserved (at least approximately). This is a key advantage of our approach.

Then we puncture this shortened mother code to a higher rate code. More specifically, first we shorten the rate 0.2 mother code from 10000 bits to 9375 bits, preserving the code rate 0.2. This means that 125 information bits and 500 parity bits are expunged from the original

codeword. We next puncture this shortened rate 0.2 code of length 9375 bits to a rate 0.25 code of length 7500 bits.

At the encoder, where 1875 data bits are to be transmitted, 125 zeros are padded, so that we can get a codeword of 10000 bits. At the decoder, we insert zeros bits (absolutely reliable estimates with log-likelihood ratio (LLR) messages set to ∞) at positions corresponding to the 125 expunged information bits, and insert erasures (LLR messages set to 0) at positions corresponding to the expunged parity bits and all the punctured bits. The received 7500 coded bits plus the 125 zeros and the 2375 erasures are fed to the decoder for the length 10000 mother code. Among the 2000 decoded information bits, we select the 1875 desired bits. Of course, the shortening and puncturing distributions of the information bits and parity bits must be known to both the transmitter and the receiver.

4.3 Background on IRA Codes

4.3.1 Definition of IRA Codes

Before going into the details of rate-compatible punctured IRA code design, we will summarize the basic properties of IRA codes. A Tanner graph describing an ensemble of IRA codes is shown in Fig. 4.2. The ensemble of IRA codes is specified by parameters $(f_2, \dots, f_{d_l}; a)$, where a fraction f_i of information bits is repeated i times and connected to i check nodes, for $i = 2, \dots, d_l$, and d_l is the maximal information node degree. The distribution $\{f_i \geq 0, i = 2, \dots, d_l : \sum_{i=2}^{d_l} i f_i = 1\}$ is referred to as the repetition profile, and is used as a degree of freedom in the optimization of the IRA ensemble [57]. For $(k+r, k)$ codes, there are k information nodes on the left, $r = (k \sum_i i f_i) / a$ check nodes in the middle, and r parity nodes on the right. Each check node is connected to a information nodes, and two parity nodes. The connections between check nodes and information nodes are determined by the interleaver and are random while the connections between check nodes and parity nodes are in a simple zigzag pattern. The code rate is easily shown to be $R_s = \frac{k}{k+r} = \frac{a}{a + \sum_i i f_i}$.

If the interleaver in the Tanner graph is fixed, a systematic IRA code can be obtained with k information bits $\underline{b} = (b_1, \dots, b_k)$ and r parity bits $\underline{x}_2 = (x_{2,1}, \dots, x_{2,r})$. The value of a parity bit is determined uniquely by the condition that the mod-2 sum of the values of the variable nodes connected to each check node is zero. If the values of the bits on the ra edges

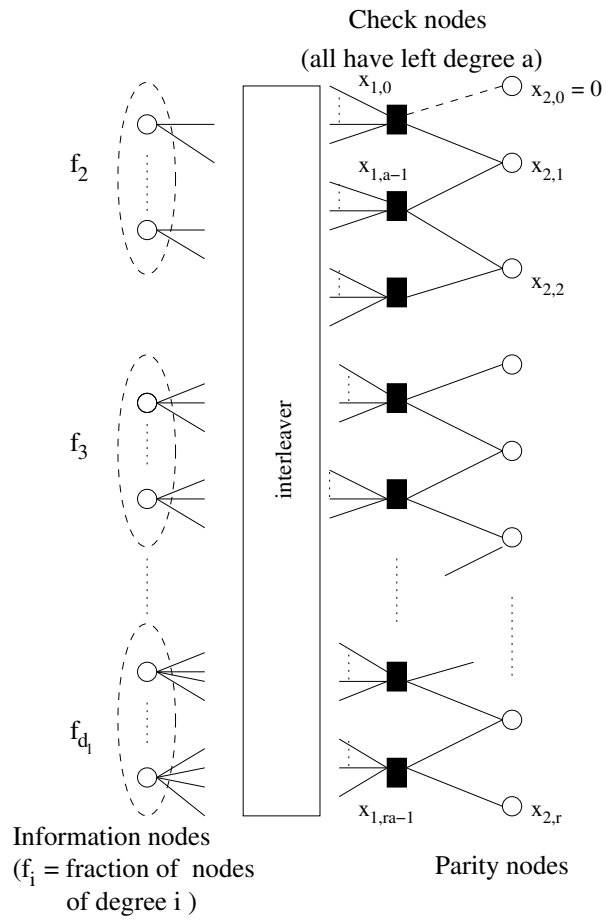


Figure 4.2: Tanner graph for IRA codes with parameters $(f_1, \dots, f_{d_1}; a)$

coming out of the interleaver are $(x_{1,0}, \dots, x_{1,ra-1})$, we have the recursion

$$x_{2,j+1} = x_{2,j} + \sum_{i=0}^{a-1} x_{1,aj+i}, \quad (4.1)$$

for $j = 0, \dots, r-1$, with the initial condition $x_{2,0} = 0$.

4.3.2 Density Evolution

Although density evolution (DE) for IRA codes is analogous to that for irregular LDPC codes [58], there are aspects of code optimization using density evolution that are unique to IRA codes. Specifically, one message-passing cycle in a standard LDPC code optimization only involves updating check-to-variable messages and variable-to-check messages, while for IRA codes, it involves updating information-to-check messages, check-to-parity messages, parity-to-check messages and check-to-information messages.

Using the density evolution technique, we describe the optimization of the IRA codes with respect to achieving the maximum rate while satisfying the target BER at a given SNR level. Let p_l and \tilde{p}_l denote the probability density functions(pdf's) of an information-to-check node message and a parity-to-check node message, respectively, at the l -th iteration. Let q_l and \tilde{q}_l denote the pdf's of a check-to-information node message and a check-to-parity node message, respectively, at the l -th iteration. Under the cycle-free condition, p_l , \tilde{p}_l , q_l and \tilde{q}_l satisfy the following recursion [58]:

$$p_l = f_u \otimes \lambda(q_l) \quad (4.2)$$

$$\tilde{p}_l = f_u \otimes \tilde{q}_l \quad (4.3)$$

$$q_l = \Gamma^{-1} \left(\Gamma(\tilde{p}_{l-1})^{\otimes 2} \otimes \Gamma(p_{l-1})^{\otimes a-1} \right) \quad (4.4)$$

$$\tilde{q}_l = \Gamma^{-1} \left(\Gamma(\tilde{p}_{l-1}) \otimes \Gamma(p_{l-1})^{\otimes a} \right) \quad (4.5)$$

for $l = 1, 2, \dots$, with initial condition $p_0 = \tilde{p}_0 = \Delta_0$ (Δ_x is a Dirac function with a shift of x), where f_u denotes the density of the channel output message, \otimes denotes convolution of densities, and $f^{\otimes m}$ denotes m -fold convolution. In (4.2),

$$\lambda(f) \triangleq \sum_{i=2}^{d_l} \lambda_i f^{\otimes(i-1)}, \quad (4.6)$$

In (4.4) and (4.5), $\Gamma(f_x)$ is the density of $y = \gamma(x)$ when the density of x is f_x , where the mapping $\gamma: \mathbb{R} \rightarrow \mathcal{F}_2 \times \mathbb{R}_+$ is defined by [38]

$$\gamma(x) = \left(\text{sign}(x), -\log \tanh \frac{|x|}{2} \right), \quad (4.7)$$

and

$$\text{sign}(x) = \begin{cases} 0 & \text{if } x > 0 \\ 0 & \text{with probability } 1/2 \text{ if } x = 0 \\ 1 & \text{with probability } 1/2 \text{ if } x = 0 \\ 1 & \text{if } x < 0. \end{cases} \quad (4.8)$$

Also, Γ^{-1} denotes the inverse mapping of Γ , i.e., $\Gamma^{-1}(g_y)$ is the density of $x = \gamma^{-1}(y)$ when the density of y is g_y .

It is obvious that $(\Delta_\infty, \Delta_\infty)$ is a fixed point of the above recursion. It is also shown in [58] that the fixed point $(\Delta_\infty, \Delta_\infty)$ for the density evolution is locally stable if and only if

$$\lambda_2 < \frac{\exp(s)(\exp(s) - 1)}{a + 1 + \exp(s)(a - 1)}, \quad (4.9)$$

where $s = -\log \left(\int \exp(-z/2) f_u(z) dz \right)$. Following an argument similar to that used in [38], it can be shown that if the stability condition holds, then there exists an $\xi > 0$ such that if for some $l \in \mathbb{N}$

$$Pe(R_s p_l(p_0, \tilde{p}_0) + (1 - R_s) \tilde{p}_l(p_0, \tilde{p}_0)) < \xi, \quad (4.10)$$

then $Pe(R_s p(p_0, \tilde{p}_0) + (1 - R_s) \tilde{p}(p_0, \tilde{p}_0))$ converges to zero as $l \rightarrow \infty$. Otherwise, there exists an $\xi > 0$ such that for all $l \in \mathbb{N}$

$$Pe(R_s p_l(p_0, \tilde{p}_0) + (1 - R_s) \tilde{p}_l(p_0, \tilde{p}_0)) > \xi. \quad (4.11)$$

For the binary-input output-symmetric channels, the IRA code ensemble design can be summarized as the solution of the following optimization problem [58]:

$$\begin{cases} \text{maximize} & a \sum_{i=2}^{d_l} \lambda_i / i \\ \text{subject to} & \sum_{i=2}^{d_l} \lambda_i = 1, \lambda_i \geq 0 \\ \text{and to} & \text{BER} < \text{target BER}. \end{cases} \quad (4.12)$$

Due to the intensive computation involved in the full density evolution, Gaussian approximation (GA) [44] is usually used for code ensemble optimization, with negligible loss of

accuracy. However, the result in [58] shows that the code ensemble optimized by GA may have a higher noise threshold than the actual channel parameter for a given capacity. Therefore, we use the full density evolution to optimize the mother code ensemble to make sure that BER of the designed code ensemble achieves the target BER. Nonetheless, we apply GA to get the shortening and puncturing distributions for the rate-compatible punctured IRA codes to make the analysis of the shortening/puncturing process more tractable.

4.4 Flexible-length Rate-compatible Punctured IRA Code Design

4.4.1 Analysis of Shortening IRA Codes

As we illustrated in Section 4.2, two steps are used to adapt the low rate mother code to a higher code rate. The first step maintains the code rate and information node degree distribution $f(x) = \sum_{j=2}^{d_l} f_j x^j$, while reducing the number of coded bits to the target codeword length before the puncturing step.

4.4.1.1 Grouping Check Nodes and Parity Nodes

Information bits having the same node degree are put into a group denoted by IG_j , $j \in [2, d_l]$, and the number of expunged information bits in each group is in proportion to its cardinality. Specifically, suppose the rate r_b mother code has codeword length n_b , consisting of k_j information bits in IG_j , $j \in [2, d_l]$, and $m_p = \frac{\sum_{j=2}^{d_l} j \cdot k_j}{a}$ parity bits. If the target codeword length before the puncturing step is n_{sb} , we randomly expunge l_i information bits, consisting of $l_{i,j} = \left\lceil \left(1 - \frac{n_{sb}}{n_b}\right) k_j \right\rceil$ bits in IG_j , $j \in [2, d_l]$, and $l_p = \left\lceil \left(1 - \frac{n_{sb}}{n_b}\right) m_p \right\rceil$ parity bits.

In the following, we develop a criterion for expunging l_p parity bits for a particular IRA code structure so that the system has guaranteed good performance. To facilitate the analysis, we consider the two check nodes connected to each parity bit as a check node pair cp_i , for $i = 1, \dots, m_p$. Each pair has $2a$ edges connecting to the information nodes on its left side. The check nodes can be grouped according to the number of incident edges connected to the information nodes which are associated with the expunged bits. The corresponding check node groups are denoted by CG_j , for $j \in [1, a]$. Two check node edge degree distributions, $\rho(x) = \sum_{j=1}^a \rho_j x^{j-1}$ and $\zeta(x) = \sum_{j=1}^a \zeta_j x^{j-1}$, are used to describe the connections to the information nodes and parity nodes, where ρ_j and ζ_j are the fractions of edges connecting the

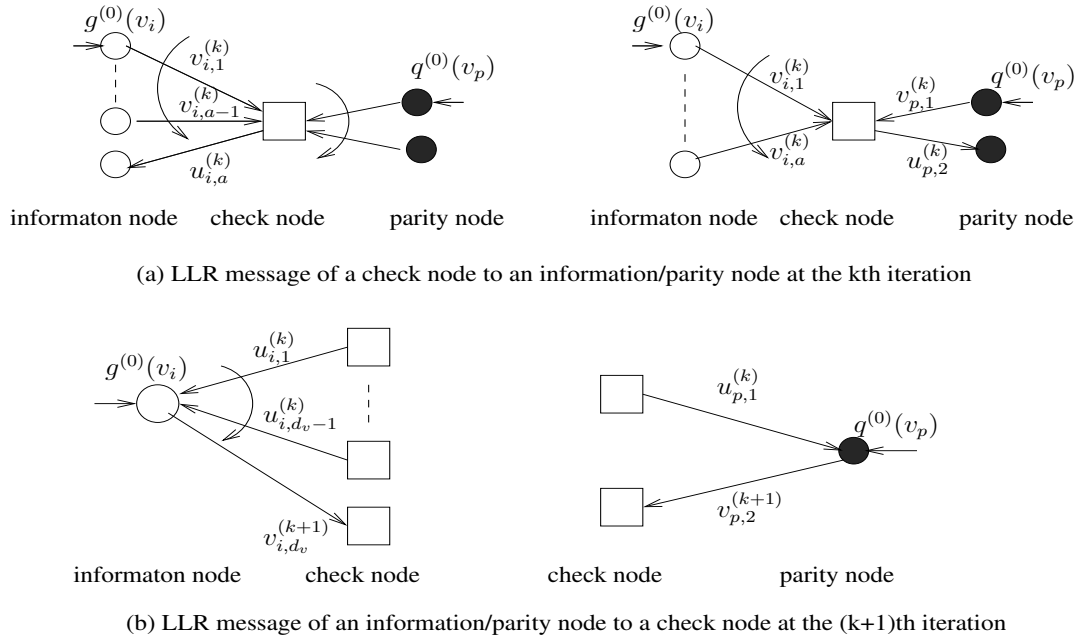


Figure 4.3: Message flows between information (parity) and check nodes

check nodes in CG_j to the information nodes and to the parity nodes, respectively. The derivations of ρ_j and ζ_j are shown in Appendix A.

Accordingly, we have $2a + 1$ groups of check node pairs CPG_j , $j = 0, \dots, 2a$, where CPG_j consists of check node pairs connecting to j information nodes associated with the transmitted information bits. Also, we have $2a + 1$ corresponding groups of parity nodes, PG_j , with edge degree distribution $\nu(x) = \sum_{j=1}^{2a} \nu_j x^{j-1}$, where ν_j is the fraction of edges connecting the parity nodes to the check node pairs in CPG_j . If the proportion of the number of parity nodes in PG_j is α_j , and the shortening proportion of each group is $\delta_j^{(0)}$, $j = 0, \dots, 2a$, then the total number of expunged parity nodes is $\sum_{j=0}^{2a} \delta_j^{(0)} \alpha_j m_p$. The $\delta_j^{(0)}$'s are optimized to satisfy the convergence condition discussed below. Note that the parity nodes connected to check node pairs in CPG_0 should be expunged since all the connected information nodes are associated with the expunged information bits, i.e., $\delta_0^{(0)} = 1$. The derivations of ν_j and α_j are in Appendix A.

4.4.1.2 Probability Density Functions of Messages from Information/Parity Nodes to Check Nodes

The message flow between a check node and information (resp., parity) nodes, and between an information (resp., a parity) node and check nodes are drawn in Fig. 4.3. In the sum-product decoding algorithm, if at least one of the information or parity nodes sends a zero message connected to a check node, the check node sends zero messages to the incident information nodes and parity nodes (excluding the information nodes associated with the expunged bits).

Let $\varepsilon_i^{(k)}$ and $e_p^{(k)}$ denote the probabilities of a check-to-information-node message and a parity-to-check-node message being zero, respectively. It is obvious that

$$\varepsilon_i^{(k)} = 1 - \left(1 - e_p^{(k)}\right)^2, \quad (4.13)$$

since no information nodes send zero messages.

The probability of a check-to-parity-node message being zero at the k -th iteration, $\varepsilon_p^{(k)}$, can be written as

$$\varepsilon_p^{(k)} = \sum_{j=1}^a \zeta_j \varepsilon_{p,j}^{(k)}, \quad (4.14)$$

where $\varepsilon_{p,j}^{(k)}$ is the probability of a message from a check node in CG_j to a parity node being zero. Since each check node connects to two parity nodes, and no information nodes send zero messages, the message from a check node will not be zero unless the message from one of the incident parity nodes is zero.

Further, if the check node belongs to CG_j , it indicates that the parity nodes connected to it belong to one of the groups PG_l , for $l \in [j, j+a]$. Thus, $\varepsilon_{p,j}^{(k)}$ depends on $e_{p,l}^{(k)}$, $l \in [j, j+a]$, the probability of a message from a parity node in PG_l to a check node being zero. It is straightforward to see that

$$\varepsilon_{p,j}^{(k)} = \sum_{l=j}^{j+a} \nu_{j,l} e_{p,l}^{(k)}, \quad (4.15)$$

where $\nu_{j,l} = \frac{v_l}{\sum_{l=j}^{j+a} v_l}$. Therefore, $\varepsilon_p^{(k)} = \sum_{j=1}^a \zeta_j \left(\sum_{l=j}^{j+a} \nu_{j,l} e_{p,l}^{(k)} \right)$.

As shown in Fig. 4.3, an information-to-check-node message and a parity-to-check-node message are determined by the linear sum of the incident check node messages and the

associated channel output messages. The message from a parity node in PG_l to a check node is zero if this parity node is associated with an expunged parity bit and the message received from the other incident check node is zero. Therefore, the probability of zero message is simply expressed as

$$e_{p,l}^{(k)} = \delta_l^{(0)} \psi_{p,l}^{(k-1)}, \quad (4.16)$$

where $\psi_{p,l}^{(k-1)}$ is the probability of a message from a check node to a parity node in PG_l being zero at the $(k-1)$ -th iteration. After some mathematical manipulations, we get

$$\psi_{p,l}^{(k-1)} = \begin{cases} \sum_{j=1}^{(l-1)/2} \Omega_{l,j}^{(k-1)} + \zeta_{l,l} \varepsilon_{p,l}^{(k-1)} & l \text{ is odd and } l \leq a \\ \sum_{j=1}^{l/2-1} \Omega_{l,j}^{(k-1)} + \zeta_{l/2,l/2} \varepsilon_{p,l/2}^{(k-1)} + \zeta_{l,l} \varepsilon_{p,l}^{(k-1)} & l \text{ is even and } l \leq a \\ \sum_{j=l-a}^{(l-1)/2} \Omega_{l,j}^{(k-1)} & l \text{ is odd and } l > a \\ \sum_{j=l-a}^{l/2-1} \Omega_{l,j}^{(k-1)} + \zeta_{l/2,l/2} \varepsilon_{p,l/2}^{(k-1)} & l \text{ is even and } l > a. \end{cases} \quad (4.17)$$

where we use $\Omega_{l,j}^{(k-1)}$ to denote $\left(\zeta_{l,j} \varepsilon_{p,j}^{(k-1)} + \zeta_{l,l-j} \varepsilon_{p,l-j}^{(k-1)} \right)$, and $\zeta_{l,j} = \frac{\zeta_j}{\sum_{j=\min(1,l-a)}^l \zeta_j}$ is normalized fraction of edges connecting the check nodes to the parity bits in PG_l . Note that $\varepsilon_{p,j}^{(0)} = 1$, for $j \in [1, a]$. The derivation of $\psi_{p,l}^{(k-1)}$ is shown in Appendix B.

If none of the parity nodes connected to a check node is expunged, the statistical properties of a check-to-information-node message and a check-to-parity-node message can be approximated as Gaussian and are uniquely determined by their mean values [44]. In this way, we can track the mean values instead of the density functions to determine if the decoding process converges or not. We use random variables u_i , u_p , v_i and $\hat{v}_{p,j}$ to denote a check-to-information-node message, a check-to-parity-node message, an information-to-check-node message and a message from a parity node to a check node in CG_j , respectively.

Using the function $\phi(x)$ defined in [44], the mean values of u_i and u_p are updated by the following relations:

$$\phi(x) = \begin{cases} 1 - \frac{1}{\sqrt{4\pi x}} \int_{\mathbb{R}} \tanh \frac{u}{2} \exp\left(-\frac{(u-x)^2}{4x}\right) du & x > 0 \\ 1 & x = 0, \end{cases} \quad (4.18)$$

the mean value of u_i at the k -th iteration is represented as

$$\begin{aligned} m_{u_i}^{(k)} &= \phi^{-1}\left(1 - E\left[\tanh(u_i^{(k)}/2) \mid u_i^{(k)} \neq 0\right]\right) \\ &= \sum_{j=1}^a \rho_j \phi^{-1}\left(1 - E\left[\tanh(v_i^{(k)}/2)\right]^{j-1} E\left[\tanh(\hat{v}_{p,j}^{(k)}/2) \mid \hat{v}_{p,j}^{(k)} \neq 0\right]^2\right), \end{aligned} \quad (4.19)$$

since the message from a check node in CG_j to one incident information node is a function of the $j - 1$ incoming messages from all the other incident information nodes and the two incoming messages from the incident parity nodes. Similarly, the mean value of u_p at the k -th iteration is represented as

$$\begin{aligned} m_{u_p}^{(k)} &= \phi^{-1} \left(1 - E \left[\tanh (u_p^{(k)} / 2) \mid u_p^{(k)} \neq 0 \right] \right) \\ &= \sum_{j=1}^a \zeta_j \phi^{-1} \left(1 - E \left[\tanh (v_i^{(k)} / 2) \right]^j E \left[\tanh (\hat{v}_{p,j}^{(k)} / 2) \mid \hat{v}_{p,j}^{(k)} \neq 0 \right] \right). \end{aligned} \quad (4.20)$$

Obviously, in order to track the updated mean values, we need the probability density functions (pdfs) of $v_i^{(k)}$ and $\hat{v}_{p,j}^{(k)}$ to calculate $E \left[\tanh (v_i^{(k)} / 2) \right]$ and $E \left[\tanh (\hat{v}_{p,j}^{(k)} / 2) \mid \hat{v}_{p,j}^{(k)} \neq 0 \right]$.

Using Gaussian approximation, the pdf of an information node message $v_i^{(k)}$ at the k -th iteration, $g^{(k)}(v_i)$, is approximated as a mixture of Gaussian probability densities [54]

$$g^{(k)}(v_i) = \sum_{j=2}^{d_i} \lambda_j \sum_{l=0}^{j-1} \chi_{j-1,l}^{(k)} N \left(m_{v_i,l}^{(k)} + m_{v_0}, 2m_{v_i,l}^{(k)} + 2m_{v_0} \right), \quad (4.21)$$

where $\lambda_j = \frac{j \cdot f_j}{\sum_j j \cdot f_j}$ is the fraction of edges connected to the information nodes of degree j . For each information node of degree j , $\chi_{j-1,l}^{(k)} \triangleq C_l^{j-1} (\varepsilon_i^{(k-1)})^{j-1-l} (1 - \varepsilon_i^{(k-1)})^l$ is defined as the probability that $(j - 1 - l)$ out of $(j - 1)$ incoming messages from the incident check nodes are zeros, where C_l^{j-1} is the binomial coefficient. We use $N(m, 2m)$ to denote the Gaussian pdf with mean m and variance $2m$.

The pdf of messages from parity nodes to check nodes in CG_j at the k -th iteration, $f^{(k)}(\hat{v}_{p,j})$, can be decomposed into four terms: $A_1^{(k)} \triangleq \hat{e}_{p,j}^{(k)} \Delta_0(\hat{v}_{p,j})$ corresponds to the a zero message, where $\hat{e}_{p,j}^{(k)}$ is the probability of the message from a parity node to a check node in CG_j at the k -th iteration being zero, which turns out to be equal to $\varepsilon_{p,j}^{(k)}$ after some calculations, as shown in Appendix B; $A_2^{(k)} \triangleq \sum_{l=j}^{j+a} \nu_{j,l} \delta_l^{(0)} N(m_{u_p}^{(k-1)}, 2m_{u_p}^{(k-1)})$ corresponds to the message obtained from the incident check nodes, given that the parity bit is expunged; $A_3^{(k)} \triangleq \sum_{l=j}^{j+a} \nu_{j,l} (1 - \delta_l^{(0)}) \psi_{p,l}^{(k-1)} N(m_{v_0}, 2m_{v_0})$ corresponds to the message from the channel, given that the messages from the incident check nodes are zeros; and $A_4^{(k)} \triangleq \sum_{l=j}^{j+a} \nu_{j,l} (1 - \delta_l^{(0)}) (1 - \psi_{p,l}^{(k-1)}) N(m_{u_p}^{(k-1)} + m_{v_0}, 2m_{u_p}^{(k-1)} + 2m_{v_0})$ corresponds to the message obtained from the combination of the messages from the incident check nodes and the message from the channel. Hence $f^{(k)}(\hat{v}_{p,j}) = \sum_{i=1}^4 A_i^{(k)}$, and the continuous part of it is

denoted by $f_c^{(k)}(\hat{v}_{p,j}) = \sum_{i=2}^4 A_i^{(k)}$. Parameters $\nu_{j,l} = \frac{\nu_l}{\sum_{l=j}^{j+a} \nu_l}$, $l \in [j, j+a]$, are normalized so that the integral of $f^{(k)}(\hat{v}_{p,j})$ equals 1.

4.4.1.3 Derivations of the Recursive Equations for the Updated Mean Values

First we substitute $g^{(k)}(v_i)$ into $E \left[\tanh (v_i^{(k)} / 2) \right]$ and denote it by $M_1^{(k)}$,

$$\begin{aligned} M_1^{(k)} &= \int \tanh (v_i^{(k)} / 2) g^{(k)}(v_i) dv_i \\ &= 1 - \sum_{j=2}^{d_i} \lambda_j \sum_{l=0}^{j-1} \chi_{j-1,l}^{(k)} \phi \left(l m_{u_i}^{(k-1)} + m_{v_0} \right). \end{aligned} \quad (4.22)$$

Next we substitute $f_c^{(k)}(\hat{v}_{p,j})$ into $E \left[\tanh (\hat{v}_{p,j}^{(k)} / 2) | \hat{v}_{p,j}^{(k)} \neq 0 \right]$ and denote it by $M_{2,j}^{(k)}$,

$$\begin{aligned} M_{2,j}^{(k)} &= \int \tanh (\hat{v}_{p,j}^{(k)} / 2) f_c^{(k)}(\hat{v}_{p,j}) d\hat{v}_{p,j} \\ &= 1 - \sum_{l=j}^{j+a} \nu_{j,l} \delta_l^{(0)} \phi(m_{u_p}^{(k-1)}) - \sum_{l=j}^{j+a} \nu_{j,l} (1 - \delta_l^{(0)}) \\ &\quad \cdot \left(\psi_{p,l}^{(k-1)} \phi(m_{v_0}) + (1 - \psi_{p,l}^{(k-1)}) \phi(m_{u_p}^{(k-1)} + m_{v_0}) \right). \end{aligned} \quad (4.23)$$

Then we substitute $M_1^{(k)}$ and $M_{2,j}^{(k)}$ into Eq. (4.19) and Eq. (4.20) and get

$$m_{u_i}^{(k)} = \sum_{s=1}^a \rho_s \phi^{-1} \left(1 - \left(M_1^{(k)} \right)^{s-1} \left(\frac{M_{2,s}^{(k)}}{1 - \hat{e}_{p,s}^{(k)}} \right)^2 \right), \quad (4.24)$$

and

$$m_{u_p}^{(k)} = \sum_{s=1}^a \zeta_s \phi^{-1} \left(1 - \left(M_1^{(k)} \right)^s \left(\frac{M_{2,s}^{(k)}}{1 - \hat{e}_{p,s}^{(k)}} \right) \right). \quad (4.25)$$

With the expression of $g^{(k)}(v_i)$ in hand, we can calculate the bit error probability at the k -th iteration, $P_e^{(k)}$, as a weighted sum of Q functions,

$$P_e^{(k)} = \sum_{j=2}^{d_i} f_j \sum_{l=0}^j \chi_{j,l}^{(k-1)} Q \left(\sqrt{\frac{m_{v_0} + l m_{u_i}^{(k-1)}}{2}} \right). \quad (4.26)$$

For error-free decoding, it is required that $\varepsilon_i^{(k)}$ and $\hat{e}_{p,s}^{(k)}$, $s \in [1, a]$, should converge to zero, which implies that $\chi_{j,l}^{(k)}$ converges to the Kronecker Delta function $\delta_{j,l}$, as k goes to infinity.

Then the mean values are simplified to

$$\begin{aligned} m_{u_i}^{(k)} &= \mathcal{F}_1 \left(m_{u_i}^{(k-1)}, m_{u_p}^{(k-1)} \right) \\ &= \sum_{s=1}^a \rho_s \phi^{-1} \left(1 - \left(M_1^{(k)} \right)^{s-1} \left(M_{2,s}^{(k)} \right)^2 \right) \end{aligned} \quad (4.27)$$

and

$$\begin{aligned} m_{u_p}^{(k)} &= \mathcal{F}_2 \left(m_{u_i}^{(k-1)}, m_{u_p}^{(k-1)} \right) \\ &= \sum_{s=1}^a \zeta_s \phi^{-1} \left(1 - \left(M_1^{(k)} \right)^s M_{2,s}^{(k)} \right), \end{aligned} \quad (4.28)$$

which for error-free decoding should grow to infinity as k goes to infinity.

4.4.1.4 Optimization of Shortening Proportions $\left\{ \delta_l^{(0)} \right\}_{l=1}^{2a}$

We denote a fixed-point solution of $m_{u_i} = \mathcal{F}_1 \left(m_{u_i}, m_{u_p}^{(k-1)} \right)$ by $m_{u_i} \left(m_{u_p}^{(k-1)} \right)$. To avoid having any fixed-point solution other than (∞, ∞) , it is required that

$$\begin{aligned} m_{u_p}^{(k-1)} &< \sum_{s=1}^a \zeta_s \phi^{-1} \left(1 - \left[1 - \sum_{j=2}^{d_l} \lambda_j \phi \left((j-1) m_{u_i} \left(m_{u_p}^{(k-1)} \right) + m_{v_0} \right) \right]^s \right. \\ &\quad \left. \cdot \left(1 - \sum_{l=s}^{s+a} \nu_{s,l} \left[\delta_l^{(0)} \phi \left(m_{u_p}^{(k-1)} \right) + (1 - \delta_l^{(0)}) \phi \left(m_{u_p}^{(k-1)} + m_{v_0} \right) \right] \right) \right). \end{aligned} \quad (4.29)$$

It is inconvenient to get the shortening proportions $\delta_l^{(0)}$, $l \in [1, 2a]$, directly from the inequality above, so we adopt the approach used in [44] and define

$$r_i^{(k)} = \sum_{j=2}^{d_i} \lambda_j \phi \left((j-1) m_{u_i}^{(k-1)} + m_{v_0} \right), \quad (4.30)$$

and

$$r_{p,s}^{(k)} = \sum_{l=s}^{s+a} \nu_{s,l} \left[\delta_l^{(0)} \phi \left(m_{u_p}^{(k-1)} \right) + (1 - \delta_l^{(0)}) \phi \left(m_{u_p}^{(k-1)} + m_{v_0} \right) \right], \quad (4.31)$$

for $s \in [1, a]$, to help solve for the $\delta_l^{(0)}$, $l \in [1, 2a]$. After substituting them into Eq. (4.27) and Eq. (4.28) we have the following equivalent recursive equations:

$$r_i^{(k+1)} = \sum_{j=2}^{d_i} \lambda_j h_j \left(m_{v_0}, r_i^{(k)}, R_p^{(k)} \right), \quad (4.32)$$

and

$$r_{p,s}^{(k+1)} = g(m_{v_0}, r_i^{(k)}, R_p^{(k)}) + \sum_{l=s}^{s+a} \nu_{s,l} \delta_l^{(0)} \left[g(0, r_i^{(k)}, R_p^{(k)}) - g(m_{v_0}, r_i^{(k)}, R_p^{(k)}) \right], \quad (4.33)$$

where $R_p^{(k)}$ denotes $\{r_{p,1}^{(k)}, \dots, r_{p,a}^{(k)}\}$. The $h(\cdot)$ and $g(\cdot)$ functions are expressed as

$$h_j(m, r_i, r_p) = \phi \left(m + (j-1) \sum_{s=1}^a \rho_s \phi^{-1} \left(1 - (1-r_i)^{s-1} (1-r_{p,s})^2 \right) \right), \quad (4.34)$$

and

$$g(m, r_i, R_p) = \phi \left(m + \sum_{t=1}^a \zeta_t \phi^{-1} \left(1 - (1-r_i)^t (1-r_{p,t}) \right) \right). \quad (4.35)$$

We denote the solution of the equation

$$r_i = \sum_{j=2}^{d_i} \lambda_j h_j(m_{v_0}, r_i, R_p) \quad (4.36)$$

by $r_i(R_p)$, for all $r_{p,s} \in (0, \phi(m_{v_0}))$ and $s \in [1, a]$, where R_p denotes $\{r_{p,1}, \dots, r_{p,a}\}$. Now the parity node updated mean inequality can be expressed as

$$r_{p,s} > g(m_{v_0}, r_i(R_p), R_p) + \sum_{l=s}^{s+a} \delta_l^{(0)} \nu_{s,l} \left[g(0, r_i(R_p), R_p) - g(m_{v_0}, r_i(R_p), R_p) \right], \quad (4.37)$$

for $r_{p,s} \in (0, \phi(m_{v_0}))$ and $s \in [1, a]$. In summing $\nu_{s,l}$, $s \in [1, a]$, $l \in [s, s+a]$, we use $\delta_l^{(0)}$ to weight them to satisfy the inequality above while maximizing $\sum_{l=1}^{2a} \alpha_l \delta_l^{(0)}$ subject to the constraint that $\delta_l^{(0)} \in [0, 1]$. Here, linear programming is employed to solve for the shortening proportions $\delta_l^{(0)}$, for $l \in [1, 2a]$.

4.4.2 Analysis of Puncturing IRA Codes

After the shortening step, we get a shortened mother code (n_{sb}, k_{sb}) with the same rate r_b as the original mother code. Note that the left repetition profile (f_1, \dots, f_{d_i}) is preserved (at least approximately) because we expunge the information nodes in each group IG_j , $j \in [1, a]$, in proportion to its cardinality. Each parity node still connects to two check nodes. However, a check node may connect to either one or two parity nodes. We use $\varphi_{j,1}$ and $\varphi_{j,2}$ to denote the fraction of edges incident to the check nodes in CG_j which are connected to one and two parity nodes, respectively, for $j \in [1, a]$.

In the puncturing step, we use the same methodology introduced in [54]. Specifically, we randomly puncture a proportion $\pi_{i,j}^{(0)}$ of information nodes in IG_j , $j \in [2, d_l]$, and a proportion $\pi_{p,l}^{(0)}$ of parity nodes in PG_l , $l \in [1, 2a]$. The puncturing proportions $\left\{ \pi_{i,j}^{(0)} \right\}_{j=2}^{d_l}$, $\left\{ \pi_{p,l}^{(0)} \right\}_{l=1}^{2a}$, $\pi_{i,j}^{(0)}, \pi_{p,l}^{(0)} \in [0, 1]$, are optimized to approach the desired puncturing fraction $p^{(0)}$ while satisfying the convergence condition given below. The puncturing fraction is defined as the ratio of the number of punctured information and parity bits to the shortened mother code length n_{sb} , which is thus expressed as

$$p^{(0)} = \sum_{j=2}^{d_l} \pi_{i,j}^{(0)} f_j + \sum_{l=1}^{2a} \pi_{p,l}^{(0)} \tilde{\alpha}_l, \quad (4.38)$$

where $\tilde{\alpha}_l$, $l \in [1, 2a]$, is the proportion of parity nodes in PG_l after the shortening step. The code rate of the punctured IRA code is $r_p = \frac{r_b}{1-p^{(0)}}$.

4.4.2.1 Probability Density Functions of Messages from Information/Parity Nodes to Check Nodes

Following a similar approach to that described in the shortening step, let $\tilde{\varepsilon}_i^{(k)}$, $\tilde{e}_i^{(k)}$ and $\tilde{e}_{p,j}^{(k)}$ denote the probabilities of a check-to-information-node message, an information-to-check-node message and a message from a parity node to a check node in CG_j being zero, respectively. We can easily obtain $\tilde{\varepsilon}_i^{(k)}$ as

$$\tilde{\varepsilon}_i^{(k)} = 1 - \sum_{j=1}^a \rho_j \left(1 - \tilde{e}_i^{(k)}\right)^{j-1} \left(\sum_{l=1}^2 \varphi_{j,l} \left(1 - \tilde{e}_{p,j}^{(k)}\right)^l \right), \quad (4.39)$$

since a check node in CG_j sends a zero message to an incident information node if the message from at least one of the other $j - 1$ incident information nodes or one of the two incident parity nodes is zero, according to the sum-product decoding algorithm.

Let $\tilde{\varepsilon}_p^{(k)}$ and $\tilde{\varepsilon}_{p,j}^{(k)}$ denote the probabilities of a check-to-parity-node message and a message from a check node in CG_j to a parity node being zeros at the k -th iteration, respectively. It is straightforward to show that

$$\begin{aligned} \tilde{\varepsilon}_p^{(k)} &= \sum_{j=1}^a \tilde{\zeta}_j \tilde{\varepsilon}_{p,j}^{(k)} \\ &= 1 - \sum_{j=1}^a \tilde{\zeta}_j \left(1 - \tilde{e}_i^{(k)}\right)^j \left(\sum_{l=1}^2 \varphi_{j,l} \left(1 - \tilde{e}_{p,j}^{(k)}\right)^{l-1} \right), \end{aligned} \quad (4.40)$$

where $\tilde{\zeta}_j$ is the fraction of edges connecting the check nodes in CG_j to the parity nodes after the shortening step.

According to the sum-product decoding algorithm, the information node message is determined by the linear sum of the incident check node messages(excluding the one from the check node to which this message is sent) , and the channel output message. Thus, a message from an information node of degree j is zero if the $(j - 1)$ incoming messages from the incident check nodes are zeros and the associated bit is punctured. Therefore, the probability of an information-to-check-node message being zero at the k -th iteration, $\tilde{e}_i^{(k)}$, can be expressed as

$$\begin{aligned}\tilde{e}_i^{(k)} &= \sum_{j=2}^{d_i} \lambda_j \tilde{e}_{i,j}^{(k)} \\ &= \sum_{j=2}^{d_i} \lambda_j \pi_{i,j}^{(0)} \left(\tilde{e}_i^{(k-1)} \right)^{j-1},\end{aligned}\quad (4.41)$$

where $\tilde{e}_{i,j}^{(k)}$ is the probability of a message from an information node of degree j being zero at the k -th iteration.

Similarly, the parity node message is determined by the linear sum of one of the incident check node messages, and the channel output message. Thus, the probability of a parity-to-check message being zero at the k -th iteration can be expressed as

$$\begin{aligned}\tilde{e}_p^{(k)} &= \sum_{l=1}^{2a} \tilde{\nu}_l \tilde{e}_{p,l}^{(k)} \\ &= \sum_{l=1}^{2a} \tilde{\nu}_l \pi_{p,l}^{(0)} \tilde{\psi}_{p,l}^{(k-1)},\end{aligned}\quad (4.42)$$

where $\tilde{\nu}_l$ is the fraction of edges connecting parity nodes in PG_l to the check nodes after the shortening step, and $\tilde{e}_{p,l}^{(k)}$ and $\tilde{\psi}_{p,l}^{(k-1)}$ are the probabilities of a message from a parity node in PG_l to a check node at the k -th iteration and a message from a check node to a parity node in PG_l being zeros at the $(k - 1)$ -th iteration, respectively. Analogous to the derivation of $\psi_{p,l}^{(k-1)}$, it can be shown that

$$\tilde{\psi}_{p,l}^{(k-1)} = \begin{cases} \sum_{j=1}^{(l-1)/2} \tilde{\Omega}_{l,j}^{(k-1)} + \tilde{\zeta}_{l,l} \tilde{e}_{p,l}^{(k-1)} & l \text{ is odd and } l \leq a \\ \sum_{j=1}^{l/2-1} \tilde{\Omega}_{l,j}^{(k-1)} + \tilde{\zeta}_{l/2,l/2} \tilde{e}_{p,l/2}^{(k-1)} + \tilde{\zeta}_{l,l} \tilde{e}_{p,l}^{(k-1)} & l \text{ is even and } l \leq a \\ \sum_{j=l-a}^{(l-1)/2} \tilde{\Omega}_{l,j}^{(k-1)} & l \text{ is odd and } l > a \\ \sum_{j=l-a}^{l/2-1} \tilde{\Omega}_{l,j}^{(k-1)} + \tilde{\zeta}_{l/2,l/2} \tilde{e}_{p,l/2}^{(k-1)} & l \text{ is even and } l > a. \end{cases}\quad (4.43)$$

where $\tilde{\Omega}_{l,j}^{(k-1)}$ is used to denote $\left(\tilde{\zeta}_{l,j}\tilde{\epsilon}_{p,j}^{(k-1)} + \tilde{\zeta}_{l,l-j}\tilde{\epsilon}_{p,l-j}^{(k-1)}\right)$, and $\tilde{\zeta}_{l,j} = \frac{\tilde{\zeta}_j}{\sum_{j=\min(1,l-a)}^l \tilde{\zeta}_j}$ is normalized fraction of edges connecting the check nodes to the parity bits in PG_l .

As we mentioned in the shortening step, parity nodes connected to a check node in CG_j must belong to one of the groups PG_l , $l \in [j, j+a]$. Thus, whether a message from a parity node to a check node in CG_j is zero depends on the messages from parity nodes in PG_l , which implies that

$$\tilde{\epsilon}_{p,j}^{(k)} = \sum_{l=j}^{j+a} \tilde{v}_{j,l} \tilde{\epsilon}_{p,l}^{(k)}, \quad (4.44)$$

where $\tilde{v}_{j,l} = \frac{\tilde{v}_l}{\sum_{l=j}^{j+a} \tilde{v}_l}$, $l \in [1, 2a]$.

If none of the information and parity nodes connected to a check node is punctured, the probabilistic characteristics of the check node messages can be approximated as Gaussian [44], and they are uniquely determined by the updated mean values

$$\begin{aligned} m_{u_i}^{(k)} &= \phi^{-1} \left(1 - E \left[\tanh(u_i^{(k)}/2) | u_i^{(k)} \neq 0 \right] \right) \\ &= \sum_{j=1}^a \rho_j \phi^{-1} \left(1 - E \left[\tanh(v_i^{(k)}/2) | v_i^{(k)} \neq 0 \right]^{j-1} \right. \\ &\quad \left. \cdot \sum_{l=1}^2 \varphi_{j,l} E \left[\tanh(\hat{v}_{p,j}^{(k)}/2) | \hat{v}_{p,j}^{(k)} \neq 0 \right]^l \right), \end{aligned} \quad (4.45)$$

and

$$\begin{aligned} m_{u_p}^{(k)} &= \phi^{-1} \left(1 - E \left[\tanh(u_p^{(k)}/2) | u_p^{(k)} \neq 0 \right] \right) \\ &= \sum_{j=1}^a \tilde{\zeta}_j \phi^{-1} \left(1 - E \left[\tanh(v_i^{(k)}/2) | v_i^{(k)} \neq 0 \right]^j \right. \\ &\quad \left. \cdot \sum_{l=1}^2 \varphi_{j,l} \cdot E \left[\tanh(\hat{v}_{p,j}^{(k)}/2) | \hat{v}_{p,j}^{(k)} \neq 0 \right]^{l-1} \right). \end{aligned} \quad (4.46)$$

To calculate $E \left[\tanh(v_i^{(k)}/2) | v_i^{(k)} \neq 0 \right]$ and $E \left[\tanh(\hat{v}_{p,j}^{(k)}/2) | \hat{v}_{p,j}^{(k)} \neq 0 \right]$, we need to track the pdfs of $v_i^{(k)}$ and $\hat{v}_{p,j}^{(k)}$ as we did in the shortening step.

The pdf of an information node message $g^{(k)}(v_i)$ can be factored into four terms: The first term $B_1^{(k)} \triangleq \tilde{\epsilon}_i^{(k)} \Delta_0(v_i)$ corresponds to a zero message; the second term $B_2^{(k)} \triangleq \sum_{j=2}^{d_i} \lambda_j \pi_{i,j}^{(0)} \sum_{l=1}^{j-1} \tilde{\chi}_{j-1,l}^{(k)} N(lm_{u_i}^{(k-1)}, 2lm_{u_i}^{(k-1)})$ corresponds to the message obtained

from the incident check nodes, given that the information bit is punctured; the third term $B_3^{(k)} \triangleq \sum_{j=2}^{d_l} \lambda_j (1 - \pi_{i,j}^{(0)}) N(m_{v_0}, 2m_{v_0})$ corresponds to the message from the channel, given that the messages from the incident check node are zeros; and the last term $B_4^{(k)} \triangleq \sum_{j=2}^{d_l} \lambda_j (1 - \pi_{i,j}^{(0)}) \sum_{l=1}^{j-1} \tilde{\chi}_{j-1,l}^{(k)} N(lm_{u_i}^{(k-1)} + m_{v_0}, 2lm_{u_i}^{(k-1)} + 2m_{v_0})$ corresponds to the message obtained from the combination of the messages from the incident check nodes and the message obtained from the channel. Here, $\tilde{\chi}_{j,l}^{(k)}$ is analogous to $\chi_{j,l}^{(k)}$ defined in the shortening step. Therefore, $g^{(k)}(v_i) = \sum_{i=1}^4 B_i^{(k)}$ and the continuous part of it is denoted by $g_c^{(k)}(v_i) = \sum_{i=2}^4 B_i^{(k)}$.

Similarly, the pdf of a message from a parity node to a check node in CG_j , $f^{(k)}(\hat{v}_{p,j})$, can also be factored into the following four terms:

$$\begin{aligned} C_1^{(k)} &\triangleq \tilde{e}_{p,j}^{(k)} \Delta_0(\hat{v}_{p,j}), \\ C_2^{(k)} &\triangleq \sum_{l=j}^{j+a} \tilde{v}_{j,l} \pi_{p,l}^{(0)} N(m_{u_p}^{(k-1)}, 2m_{u_p}^{(k-1)}), \\ C_3^{(k)} &\triangleq \sum_{l=j}^{j+a} \tilde{v}_{j,l} (1 - \pi_{p,l}^{(0)}) \tilde{\psi}_{p,l}^{(k-1)} N(m_{v_0}, 2m_{v_0}), \\ C_4^{(k)} &\triangleq \sum_{l=j}^{j+a} \tilde{v}_{j,l} (1 - \pi_{p,l}^{(0)}) (1 - \tilde{\psi}_{p,l}^{(k-1)}) N(m_{u_p}^{(k-1)} + m_{v_0}, 2m_{u_p}^{(k-1)} + 2m_{v_0}). \end{aligned}$$

And $f^{(k)}(\hat{v}_{p,j}) = \sum_{i=1}^4 C_i^{(k)}$, with the continuous part being denoted by $f_c^{(k)}(\hat{v}_{p,j}) = \sum_{i=2}^4 C_i^{(k)}$.

4.4.2.2 Derivations of the Recursive Equations for the Updated Mean Values

With the expression of $g_c^{(k)}(v_i)$, we are ready to calculate $E[\tanh(v_i^{(k)}/2) | v_i^{(k)} \neq 0]$ and denote it by $L_1^{(k)}$,

$$\begin{aligned} L_1^{(k)} &= \int \tanh(v_i^{(k)}/2) g_c^{(k)}(v_i) dv_i \\ &= 1 - \sum_{j=2}^{d_l} \left(\lambda_j \pi_{i,j}^{(0)} \sum_{l=1}^{j-1} \tilde{\chi}_{j-1,l}^{(k-1)} \phi(lm_{u_i}^{(k-1)}) \right. \\ &\quad \left. + \lambda_j (1 - \pi_{i,j}^{(0)}) \sum_{l=0}^{j-1} \tilde{\chi}_{j-1,l}^{(k-1)} \phi(lm_{u_i}^{(k-1)} + m_{v_0}) \right). \end{aligned} \tag{4.47}$$

Next we substitute $f_c^{(k)}(\hat{v}_{p,j})$ into $E \left[\tanh(\hat{v}_{p,j}/2) | \hat{v}_{p,j}^{(k)} \neq 0 \right]$ and denote it by $L_{2,j}^{(k)}$,

$$\begin{aligned} L_{2,j}^{(k)} &= \int \tanh(\hat{v}_{p,j}/2) f_c^{(k)}(\hat{v}_{p,j}) d\hat{v}_{p,j} \\ &= 1 - \sum_{n=j}^{j+a} \tilde{\nu}_{j,n} \left(\pi_{p,n}^{(0)} \phi(m_{u_p}^{(k-1)}) + (1 - \pi_{p,n}^{(0)}) \right. \\ &\quad \left. \cdot (\tilde{\psi}_{p,n}^{(k-1)} \phi(m_{v_0}) + (1 - \tilde{\psi}_{p,n}^{(k-1)}) \phi(m_{v_0} + m_{u_p}^{(k-1)})) \right). \end{aligned} \quad (4.48)$$

Then we substitute $L_1^{(k)}$ and $L_{2,j}^{(k)}$ into Eq. (4.45) and Eq. (4.46) and get

$$m_{u_i}^{(k)} = \sum_{j=1}^a \rho_j \phi^{-1} \left(1 - \left(\frac{L_1^{(k)}}{1 - \tilde{\epsilon}_i^{(k)}} \right)^{j-1} \left(\sum_{l=1}^2 \varphi_{j,l} \left(\frac{L_{2,j}^{(k)}}{1 - \tilde{\epsilon}_{p,j}^{(k)}} \right)^l \right) \right) \quad (4.49)$$

and

$$m_{u_p}^{(k)} = \sum_{j=1}^a \tilde{\zeta}_j \phi^{-1} \left(1 - \left(\frac{L_1^{(k)}}{1 - \tilde{\epsilon}_i^{(k)}} \right)^j \left(\sum_{l=1}^2 \varphi_{j,l} \left(\frac{L_{2,j}^{(k)}}{1 - \tilde{\epsilon}_{p,j}^{(k)}} \right)^{l-1} \right) \right) \quad (4.50)$$

Using the approach introduced in [54], we calculated the bit error probability at the k -th iteration $P_e^{(k)}$,

$$\begin{aligned} P_e^{(k)} &= \frac{\sum_{j=2}^{d_l} f_j \pi_{i,j}^{(0)} \left(\tilde{\epsilon}_i^{(k-1)} \right)^j}{2} \\ &\quad + \sum_{j=2}^{d_l} f_j \pi_{i,j}^{(0)} \sum_{l=1}^j \tilde{\chi}_{j,l}^{(k)} Q \left(\sqrt{\frac{l m_{u_i}^{(k-1)}}{2}} \right) \\ &\quad + \sum_{j=2}^{d_l} f_j \left(1 - \pi_{i,j}^{(0)} \right) \left(\tilde{\epsilon}_i^{(k-1)} \right)^j Q \left(\sqrt{\frac{m_{v_0}}{2}} \right) \\ &\quad + \sum_{j=2}^{d_l} f_j \left(1 - \pi_{i,j}^{(0)} \right) \sum_{l=1}^j \tilde{\chi}_{j,l}^{(k)} Q \left(\sqrt{\frac{l m_{u_i}^{(k-1)} + m_{v_0}}{2}} \right), \end{aligned} \quad (4.51)$$

with each term corresponding to $B_i^{(k)}$, $i \in [1, 4]$. If the transmitted coded bits converge to the correct coded bits, the last two terms in the BER expression will converge to 0 as $k \rightarrow \infty$, i.e.,

$$\begin{aligned} \sum_{j=2}^{d_l} f_j \left(1 - \pi_{i,j}^{(0)} \right) \left(\tilde{\epsilon}_i^{(k-1)} \right)^j Q \left(\sqrt{\frac{m_{v_0}}{2}} \right) &\rightarrow 0 \\ \sum_{j=2}^{d_l} f_j \left(1 - \pi_{i,j}^{(0)} \right) \sum_{l=1}^j \tilde{\chi}_{j,l}^{(k)} Q \left(\sqrt{\frac{l m_{u_i}^{(k-1)} + m_{v_0}}{2}} \right) &\rightarrow 0. \end{aligned} \quad (4.52)$$

This is true if and only if $m_{u_i}^{(k)} \rightarrow \infty$ and $\tilde{\varepsilon}_i^{(k)} \rightarrow 0$ as $k \rightarrow \infty$ [54]. If these conditions are satisfied, the first and second terms also converge to 0, i.e., punctured bits converge to correct coded bits, too.

For error-free decoding, it is required that $\tilde{\varepsilon}_i^{(k)}$ and $\hat{e}_{p,j}^{(k)}$ converge to zero indicating that $\tilde{\chi}_{j,l}^{(k)}$ converges to a distribution function $\delta_{j,l}$. Thus, the recursive equations of the updated mean values are simplified as follows:

$$\begin{aligned} m_{u_i}^{(k)} &= \mathcal{F}_1 \left(m_{u_i}^{(k-1)}, m_{u_p}^{(k-1)} \right) \\ &= \sum_{s=1}^a \rho_s \phi^{-1} \left(1 - \left(L_1^{(k)} \right)^{s-1} \left(\sum_{l=1}^2 \varphi_{s,l} \left(L_{2,s}^{(k)} \right)^l \right) \right), \end{aligned} \quad (4.53)$$

and

$$\begin{aligned} m_{u_p}^{(k)} &= \mathcal{F}_2 \left(m_{u_i}^{(k-1)}, m_{u_p}^{(k-1)} \right) \\ &= \sum_{s=1}^a \tilde{\zeta}_s \phi^{-1} \left(1 - \left(L_1^{(k)} \right)^s \left(\sum_{l=1}^2 \varphi_{s,l} \left(L_{2,s}^{(k)} \right)^{l-1} \right) \right), \end{aligned} \quad (4.54)$$

which for error-free decoding should grow to infinity as k goes to infinity.

4.4.2.3 Optimization of Puncturing Proportions $\left\{ \pi_{i,j}^{(0)} \right\}_{j=2}^{d_i}$ and $\left\{ \pi_{p,l}^{(0)} \right\}_{l=1}^{2a}$

Again, we denote a fixed-point solution of $m_{u_i} = \mathcal{F}_1 \left(m_{u_i}, m_{u_p}^{(k-1)} \right)$ by $m_{u_i} \left(m_{u_p}^{(k-1)} \right)$. To avoid having any fixed-point solution other than (∞, ∞) , it is required that

$$m_{u_p}^{(k-1)} < \mathcal{F}_2 \left(m_{u_i} \left(m_{u_p}^{(k-1)} \right), m_{u_p}^{(k-1)} \right). \quad (4.55)$$

Following a similar approach to that described in the shortening step, we define two functions of $m_{u_i}^{(k-1)}, m_{u_p}^{(k-1)}, \left\{ \pi_{i,j}^{(0)} \right\}_{j=2}^{d_i}$ and $\left\{ \pi_{p,l}^{(0)} \right\}_{l=1}^{2a}$ as follows:

$$\tilde{r}_i^{(k)} = \sum_{j=2}^{d_i} \lambda_j \left(\pi_{i,j}^{(0)} \phi \left((j-1)m_{u_i}^{(k-1)} \right) + (1 - \pi_{i,j}^{(0)}) \phi \left((j-1)m_{u_i}^{(k-1)} + m_{v_0} \right) \right), \quad (4.56)$$

and

$$\tilde{r}_{p,s}^{(k)} = \sum_{n=s}^{s+a} \tilde{\nu}_{s,n} \left(\pi_{p,n}^{(0)} \phi \left(m_{u_p}^{(k-1)} \right) + (1 - \pi_{p,n}^{(0)}) \phi \left(m_{v_0} + m_{u_p}^{(k-1)} \right) \right). \quad (4.57)$$

Now we substitute $r_i^{(k)}$ and $r_{p,s}^{(k)}$ into Eq. (4.53) and Eq. (4.54) to solve for the $\{\pi_{i,j}^{(0)}\}_{j=2}^{d_i}$ and $\{\pi_{p,l}^{(0)}\}_{l=1}^{2a}$. And we have the following equivalent recursive equations for $r_i^{(k)}$ and $r_{p,s}^{(k)}$:

$$\tilde{r}_i^{(k+1)} = \sum_{j=2}^{d_i} \lambda_j^{\pi_i} h_j \left(0, \tilde{r}_i^{(k)}, \tilde{R}_p^{(k)} \right) + \sum_{j=2}^{d_i} \lambda_j^{(1-\pi_i)} h_j \left(m_{v_0}, \tilde{r}_i^{(k)}, \tilde{R}_p^{(k)} \right) \quad (4.58)$$

and

$$\begin{aligned} \tilde{r}_{p,s}^{(k+1)} &= g \left(m_{v_0}, \tilde{r}_i^{(k)}, \tilde{R}_p^{(k)} \right) \\ &+ \sum_{n=s}^{s+a} \tilde{\nu}_{s,n} \pi_{p,n}^{(0)} \left(g \left(0, \tilde{r}_i^{(k)}, \tilde{R}_p^{(k)} \right) - g \left(m_{v_0}, \tilde{r}_i^{(k)}, \tilde{R}_p^{(k)} \right) \right), \end{aligned} \quad (4.59)$$

where $\tilde{R}_p^{(k)}$ is used to denote $\{\tilde{r}_{p,1}^{(k)}, \dots, \tilde{r}_{p,a}^{(k)}\}$. The $h(\cdot)$ and $g(\cdot)$ functions are expressed as

$$h_j(m, \tilde{r}_i, \tilde{R}_p) = \phi \left(m + (j-1) \sum_{s=1}^a \rho_s \phi^{-1} \left(1 - (1 - \tilde{r}_i)^{s-1} \left(\sum_{l=1}^2 \varphi_{s,l} (1 - \tilde{r}_{p,s})^l \right) \right) \right), \quad (4.60)$$

and

$$g(m, \tilde{r}_i, \tilde{R}_p) = \phi \left(m + \sum_{t=1}^a \tilde{\zeta}_t \phi^{-1} \left(1 - (1 - \tilde{r}_i)^t \left(\sum_{l=1}^2 \varphi_{t,l} (1 - \tilde{r}_{p,t})^{l-1} \right) \right) \right). \quad (4.61)$$

We denote the solution of the equation

$$\tilde{r}_i = \sum_{j=2}^{d_i} \lambda_j \pi_{i,j}^{(0)} h_j \left(0, \tilde{r}_i, \tilde{R}_p \right) + \sum_{j=2}^{d_i} \lambda_j (1 - \pi_{i,j}^{(0)}) h_j \left(m_{v_0}, \tilde{r}_i, \tilde{R}_p \right) \quad (4.62)$$

by $r_i(\tilde{R}_p)$, for all $r_{p,s} \in (0, \phi(m_{v_0}))$, $s \in [1, a]$. To solve the equation for r_i , we use $\pi_{i,j}^{(0)}$, $j \in [2, d_i]$, to weight the functions $h_j(\cdot)$. Once r_i in the equation above is obtained, the $\pi_{i,j}^{(0)}$, $j \in [2, d_i]$, are fixed as the corresponding weights for the functions $h_j(\cdot)$.

Now the updated mean inequality for a parity node can be expressed as

$$\tilde{r}_{p,s} > g(m_{v_0}, \tilde{r}_i, \tilde{R}_p) + \sum_{n=s}^{s+a} \tilde{\nu}_{s,n} \pi_{p,n}^{(0)} \left(g \left(0, \tilde{r}_i, \tilde{R}_p \right) - g \left(m_{v_0}, \tilde{r}_i, \tilde{R}_p \right) \right), \quad (4.63)$$

for $r_{p,s} \in (0, \phi(m_{v_0}))$, $s \in [1, a]$. In summing $\tilde{\nu}_{s,n}$, $s \in [1, a]$, $n \in [s, s+a]$, we use $\pi_{p,n}^{(0)}$, $n \in [1, 2a]$, to weight them to satisfy the inequality above. Again, linear programming is employed to get the solution for $\pi_{p,l}^{(0)}$, $l \in [1, 2a]$. Specifically, $\sum_{l=1}^{2a} \pi_{p,l}^{(0)} \tilde{\alpha}_l$ are maximized to approach the desired puncturing fraction $p^{(0)}$ subject to the constraint $\pi_{p,l}^{(0)} \in [0, 1]$.

Table 4.1: Distribution of expunged information bits for codes of rate 0.25 and 0.35 using rate 0.2 mother code

degree j	2	3	6	7	10	19	50
f_j	0.1696	0.3092	0.0752	0.1267	0.1088	0.0722	0.1382
$l_{i,j}(\text{rate}=0.25)$	21	39	9	16	14	9	17
$l_{i,j}(\text{rate}=0.35)$	44	77	19	32	27	18	35

Table 4.2: Shortening proportion of parity bits for codes of rate 0.25 and 0.35 using rate 0.2 mother code

degree l	0	1	2	3	4	5	6
$\delta_l^{(0)}(\text{rate}=0.25)$	1	0.4015	0.0726	0.0652	0.0296	0.0566	0.0656
$\delta_l^{(0)}(\text{rate}=0.35)$	1	0.7838	0.0401	0.0351	0.0301	0.0251	0.1751

4.5 Numerical Results

We show the details of the shortening of the rate 0.2, length-10000 mother code to length 9375 and length 8750 in Tables 4.1 and 4.2, respectively. Table 4.1 shows the number of expunged information bits of each node degree, which is proportional to the cardinality of each group. Table 4.2 shows the shortening proportion of parity bits in group PG_l , $l \in [0, 2a]$. The results indicate that parity bits connected to the least number of transmitted information bits are more likely to be expunged. The reason is that the messages are only exchanged among the information bits and the parity bits through the mutually incident check node pair. The reliabilities of the parity nodes connected to the least number of transmitted information bits can hardly be improved by using the extrinsic information provided by the information bits after the first few iterations in the message-passing decoding process.

The results in Table 4.3 and Table 4.4 are the puncturing proportions of the information bits in group IG_j , $\pi_{i,j}^{(0)}$, $j \in [2, d_l]$, and the puncturing proportions of the parity bits in group PG_l , $\pi_{p,l}^{(0)}$, $l \in [1, 2a]$, required to produce rate 0.25 and 0.35 punctured IRA codes using a rate 0.2 mother code. Examining the results, we find that the parity bits are more likely to be punctured than the information bits. We also notice that the puncturing fraction of the information bits $\sum_{j=2}^{d_l} \pi_{i,j}^{(0)} f_j$ hardly increases, even when the overall puncturing fraction increases significantly. The increase of the overall puncturing fraction is mainly reflected in the punctur-

Table 4.3: Puncturing proportion of information bits for codes of rate 0.25 and 0.35 using rate 0.2 mother code

degree j	2	3	6	7	10	19	50
$\pi_{i,j}^{(0)}$ (rate=0.25)	0.0810	0.0383	0.0105	0.0326	0.0363	0.0418	0.0536
$\pi_{i,j}^{(0)}$ (rate=0.35)	0.0892	0.0473	0.0908	0.0951	0.0755	0.0656	0.0583

Table 4.4: Puncturing proportion of parity bits for codes of rate 0.25 and 0.35 using rate 0.2 mother code

degree l	1	2	3	4	5	6
$\pi_{p,l}^{(0)}$ (rate=0.25)	0.3949	0.0905	0.0989	0.0816	0.1664	0.1640
$\pi_{p,l}^{(0)}$ (rate=0.35)	0.9955	0.3304	0.4225	0.4081	0.3686	0.3239

ing fraction of the parity bits $\sum_{l=1}^{2a} \pi_{p,l}^{(0)} \tilde{\alpha}_l$. For example, the puncturing fraction increases from 0.1997 to 0.4234 for rate 0.2 and 0.35 punctured codes, respectively. The puncturing fraction of the information bits only increases from 0.0449 to 0.0696, while the puncturing fraction of the parity bits increases from 0.1548 to 0.3538.

The results in Table 4.5 are obtained by averaging the maximal achievable rate (resp., instantaneous capacity) over a large number of independent channel realizations for an adaptively modulated V-BLAST type system. Block fading is assumed for each realization. The superscript *osd* denotes optimal successive detection. The results in Table 4.5 are also shown in Fig. 4.4. It is obvious that the average sum rate of our adaptively modulated, IRA-encoded system closely approaches the V-BLAST capacity using rate/power adaptation and optimal suc-

Table 4.5: Capacities and maximal achievable rates of a 2×2 V-BLAST system using rate-compatible punctured IRA codes

SNR	R_{p1}	R_{p2}	R_p	C^{osd}
0	0.5732	0.7844	1.3576	1.6982
2	0.7499	1.0524	1.8023	2.2688
4	1.0120	1.3224	2.3345	2.9287
6	1.2496	1.7365	3.0112	3.7092
8	1.5426	2.1825	3.7251	4.5846
10	1.9380	2.6361	4.5741	5.5513

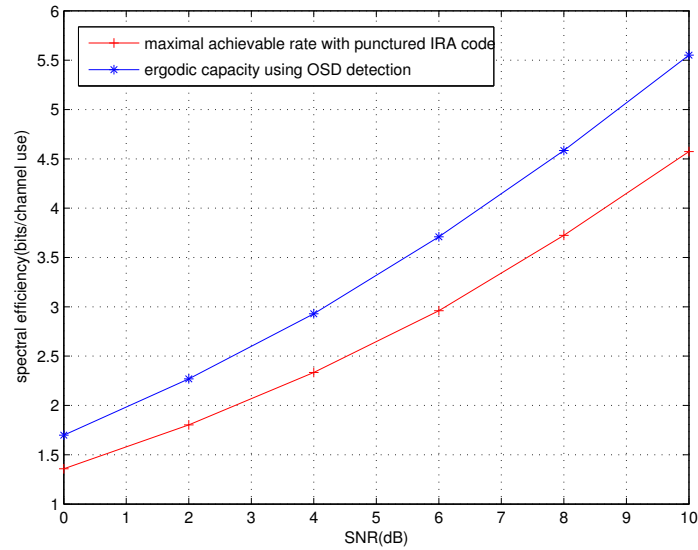


Figure 4.4: Sum rate of an adaptively modulated 2×2 V-BLAST system with rate-compatible punctured IRA codes

cessive detection.

In the BER performance simulation, parallel interference cancellation (PIC) followed by MMSE filtering is adopted to mitigate the effects of error propagation, since in an actual system the perfect feedback assumed in the optimal successive detection is not achieved. Here, we fix the symbol length in each frame to be 2500, and use a rate 0.2 length-10,000 IRA code as the mother code, which has an optimal degree distribution obtained from DE. For each channel realization, the codeword length is determined by the chosen constellation and the target code rate. Then we follow the two-step shortening/puncturing process described in Section 4.4 to get the punctured IRA codes for different realizations. Specifically, in practice, we precalculate the codeword lengths for all the possible combinations among a set of constellations and a set of quantized code rates which are matched to the channel variations. Then we design the shortening and puncturing distributions for these codeword lengths and rates and store them at the transmitter. When the transmitter receives the feedback of the quantized channel state information, it chooses the shortening and puncturing distributions required to produce the code with the appropriate length and rate. The curves in Fig. 4.5 show that the BER of our adaptively

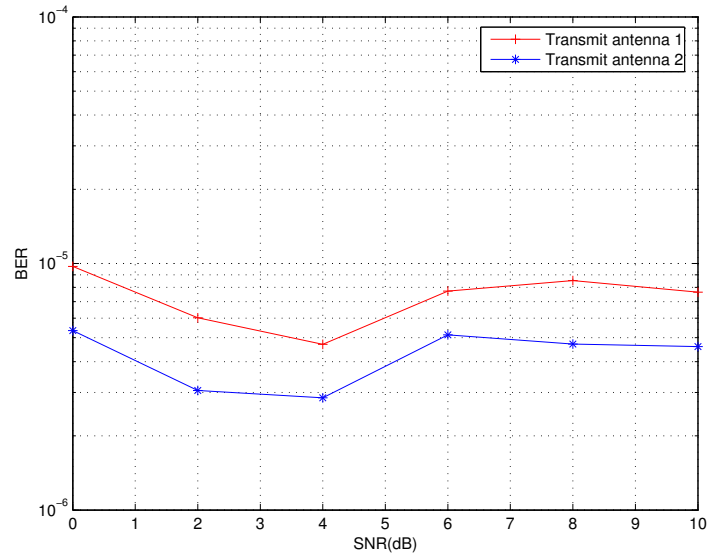


Figure 4.5: BER performance of an adaptively modulated 2×2 V-BLAST system with rate-compatible punctured IRA codes

modulated V-BLAST system using the rate-compatible punctured IRA codes achieves the target BER, 10^{-5} .

4.6 Conclusion

A novel method using both shortening and puncturing of a mother IRA code is proposed to generate a family of codes of different rates and block lengths, with the latter subject to the constraint imposed by the fixed frame length (in symbols), an adaptive modulation alphabet and a target code rate. Good performance of all the derived codes is guaranteed by the design criterion. The key advantage of this method is that one underlying decoder structure for the mother IRA code suffices to decode punctured IRA codes with different rates and block lengths, which is desirable in scenarios using adaptive modulation.

Specifically, a two-step shortening/puncturing process is used to produce a code with the target length and rate from the mother code. In the shortening step, the block length of the mother code is adapted to satisfy the constraint imposed by the chosen modulation alphabet, the length of the frame (in symbols) and the target code rate for each channel realization. The

mother code rate is (at least approximately) preserved by proportionally expunging information bits and parity bits. Furthermore, the optimal degree distribution of the mother code is (at least approximately) preserved by expunging the number of information bits of each node degree in proportion to the cardinality of the corresponding group. Then Gaussian approximation analysis is used to derive the shortening distribution for the parity bits and the puncturing distributions for the information bits and the parity bits. Since the channel code rate is maximized for each channel realization at a given SNR, the target puncturing fraction is accordingly maximized. Therefore, the puncturing distribution is optimized to approach the target puncturing fraction for a higher code rate while satisfying the convergence condition, which guarantees good performance of the code. Finally, we apply our punctured IRA codes to an adaptively modulated V-BLAST system to demonstrate the practicality of our approach.

4.7 Acknowledgement

I would like to thank Joseph Soriaga for suggesting to look at IRA codes to solve the problem we addressed in subsection 4.2.1.

Appendix 4.A Derivation of ρ_j , ζ_j , ν_j and α_j

For the (n_b, k_b) IRA mother code with repetition profile $(f_2, \dots, f_{d_l}; a)$ and rate r_b , the total number of check nodes is $r = \frac{k_b \sum_{i=2}^{d_l} i f_i}{a}$, which also equals the total number of parity bits m_p . Among k_b information bits, some are the data bits to be transmitted, and others are the padded zeros to be expunged in the shortening step. First, we divide the information bits into groups G_j , $j \in [2, d_l]$, according to their node degrees. Similarly, we divide the check nodes into groups CG_j , $j \in [1, a]$, according to the number of incident information nodes which are associated with the transmitted information bits. We also divide the parity nodes into groups PG_j , $j \in [1, 2a]$, according to the total number of connections through the two incident check nodes to the information nodes associated with the transmitted bits.

Suppose that n_s coded bits have to be expunged in the shortening step. It is obvious that the number of expunged information bits in group G_j is $l_{i,j} = \lceil n_s r_b f_j \rceil$, so that the left degree distribution is preserved (at least approximately), where $l_i = \sum_{j=2}^{d_l} l_{i,j}$ is the total number of expunged information bits. The total number of expunged parity bits should be $l_p = \lceil n_s (1 - r_b) \rceil$ in order to maintain the code rate. The total number of edges of the interleaver is $n_e = k_b \sum_{j=2}^{d_l} j f_j$, and $n_{e,i} = \sum_{j=2}^{d_l} j l_{i,j}$ edges are connected to the expunged bits.

If a uniform interleaver is assumed, the probability of each fixed interleaver will be $\frac{1}{N_r}$. Suppose that we connect n_e edges of a fixed interleaver to the check nodes in a fixed order, say, from c_1 to c_r . For c_1 , the total number of connections is

$$N_{1,l} = \binom{n_{e,i}}{a-l} \binom{n_e - n_{e,i}}{l}, \quad (4.64)$$

where c_1 is connected to l edges incident to the information nodes associated with the transmitted bits and $a - l$ edges incident to the information nodes associated with the expunged bits.

If we continue the process, it is straightforward to show that the total number of combinations of $0 \leq l \leq 2a$ edges connected to (c_1, c_2) is

$$N_{2,l} = \sum_{k=\max(0, l-a)}^{\min(l, a)} \binom{n_{e,i}}{a-k} \binom{n_e - n_{e,i}}{k} \binom{n_{e,i} - a + k}{a-l+k} \binom{n_e - n_{e,i} - k}{l-k}, \quad (4.65)$$

where c_1 is connected to k edges incident to the information nodes associated with the transmitted bits and $a - k$ edges incident to the information nodes associated with the expunged bits,

while c_2 is connected to $l - k$ edges incident to the information nodes associated with the transmitted bits and $a - l + k$ edges incident to the information nodes associated with the expunged bits.

Generalizing the results above, we get the total number of connections for the r check nodes

$$N_r = \underbrace{\sum_{k_1} \cdots \sum_{k_r}}_{\sum_{i=1}^r k_i = n_e - n_{e,i}, k_i \in [0, a]} \binom{n_{e,i}}{a - k_1} \binom{n_e - n_{e,i}}{k_1} \cdots \binom{n_{e,i} - (r-1)a + \sum_{i=1}^{r-1} k_i}{a - k_r} \binom{n_e - n_{e,i} - \sum_{i=1}^{r-1} k_i}{k_r}, \quad (4.66)$$

where c_m is connected to k_m edges incident to the information nodes associated with the transmitted bits and $a - k_m$ edges incident to the information nodes associated with the expunged bits, $m \in [1, r]$.

For the i -th interleaver, we count the number of edges connecting check nodes in CG_j to the information nodes associated with the transmitted bits, $n_{I,j}^{(i)}$, and the number of edges connecting check nodes in CG_j to the parity nodes, $n_{P,j}^{(i)}$, for $i = 1, \dots, N_r$ and $j = 1, \dots, a$. Note that the total number of edges connecting check nodes and parity nodes is twice the number of parity bits minus one, because the last parity bit only connects to one check node. In Section 4.4, ρ_j is defined as the fraction of edges connecting information nodes to the check nodes in CG_j , and ζ_j is defined as the fraction of edges connecting parity nodes to the check nodes in CG_j in the shortening step, respectively. Hence, ρ_j and ζ_j are expressed as $\rho_j = \frac{1}{N_r(n_e - n_{e,pi})} \sum_{i=1}^{N_r} n_{I,j}^{(i)}$ and $\zeta_j = \frac{1}{N_r(2m_p - 1)} \sum_{i=1}^{N_r} n_{P,j}^{(i)}$.

In Section 4.4, α_j is defined as the fraction of parity nodes in PG_j in the shortening step, and ν_j is defined as the fraction of edges connecting check nodes to the parity nodes in PG_j . Thus, for the i -th interleaver, we count the number of edges connecting parity nodes in PG_j to the check nodes, $p_{e,j}^{(i)}$, for $i = 1, \dots, N_r$ and $j = 1, \dots, 2a$. At the same time, we get the number of parity bits in PG_j , that is $p_{n,j}^{(i)}$, $j = 0, \dots, 2a$. Hence, ν_j and α_j are expressed as $\nu_j = \frac{1}{N_r(2m_p - 1)} \sum_{i=1}^{N_r} p_{e,j}^{(i)}$ and $\alpha_j = \frac{1}{N_r m_p} \sum_{i=1}^{N_r} p_{n,j}^{(i)}$, for $j = 0, \dots, 2a$.

solid line edges connected to transmitted bits
dotted line edges connected to expunged bits

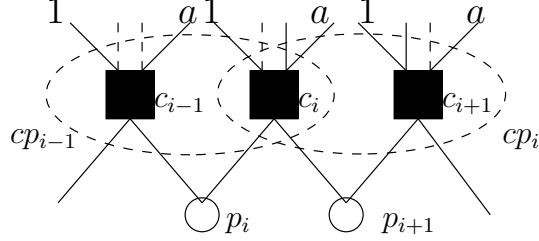


Figure 4.6: Grouping check nodes and parity nodes

Appendix 4.B Derivation of $\psi_{p,l}^{(k-1)}$ and Proof of $\hat{\varepsilon}_{p,j}^{(k)} = \varepsilon_{p,j}^{(k)}$

4.B.1 Derivation of $\psi_{p,l}^{(k-1)}$

In sub-subsection 4.4.2.1, we use $\psi_{p,l}^{(k-1)}$ and $\varepsilon_{p,j}^{(k-1)}$ to denote the probabilities of a message from a check node to a parity node in PG_l and a message from a check node in CG_j to a parity node being zero at the $(k-1)$ -th iteration, respectively.

In Figure 4.6, each check node connects to a information nodes, where the solid-line-edge is incident to an information node associated with a transmitted bit and the dotted-line-edge is incident to an information node associated to an expunged bit. Thus, each parity node connects to at most $2a$ information nodes which are associated with transmitted bits through a check node pair.

For example, if c_{i-1} belongs to CG_j and c_i belongs to CG_{l-j} , then p_i belongs to PG_l . It is obvious that $\psi_{p,l}^{(k-1)}$ is a linear combination of $\varepsilon_{p,j}^{(k-1)}$ and $\varepsilon_{p,l-j}^{(k-1)}$, and the linear parameters depend on the fractions of the edges connecting the check nodes in CG_j and CG_{l-j} to the parity bits in PG_l . Note that j and $l-j$ have to satisfy $j \in [1, a]$ and $(l-j) \in [1, a]$.

If we average all the possible combinations of j and $l-j$, we get

$$\psi_{p,l}^{(k-1)} = \begin{cases} \sum_{j=1}^{(l-1)/2} \Omega_{l,j}^{(k-1)} + \zeta_{l,l} \varepsilon_{p,l}^{(k-1)} & l \text{ is odd and } l \leq a \\ \sum_{j=1}^{l/2-1} \Omega_{l,j}^{(k-1)} + \zeta_{l/2,l/2} \varepsilon_{p,l/2}^{(k-1)} + \zeta_{l,l} \varepsilon_{p,l}^{(k-1)} & l \text{ is even and } l \leq a \\ \sum_{j=l-a}^{(l-1)/2} \Omega_{l,j}^{(k-1)} & l \text{ is odd and } l > a \\ \sum_{j=l-a}^{l/2-1} \Omega_{l,j}^{(k-1)} + \zeta_{l/2,l/2} \varepsilon_{p,l/2}^{(k-1)} & l \text{ is even and } l > a. \end{cases} \quad (4.67)$$

where we use $\Omega_{l,j}^{(k-1)}$ to denote $\left(\zeta_{l,j}\varepsilon_{p,j}^{(k-1)} + \zeta_{l,l-j}\varepsilon_{p,l-j}^{(k-1)}\right)$, and $\zeta_{l,j} = \frac{\zeta_j}{\sum_{j=\min(1,l-a)}^l \zeta_j}$ is normalized fraction of edges connecting the check nodes to the parity bits in PG_l .

4.B.2 Proof of $\hat{e}_{p,j}^{(k)} = \varepsilon_{p,j}^{(k)}$

In sub-subsection 4.4.2.1, we show that the probability of a message from a check node in CG_j to a parity bit being zero at the k -th iteration

$$\varepsilon_{p,j}^{(k)} = \sum_{l=j}^{j+a} \nu_{j,l} e_{p,l}^{(k)},$$

where $e_{p,l}^{(k)}$ is the probability of a message from a parity bit in PG_l to a check node being zero at the k -th iteration, and $\nu_{j,l}$ is the normalized fraction of edges connecting the parity bits to the check nodes in CG_j .

We use $\hat{e}_{p,j}^{(k)}$ to denote the probability of a message from a parity bit to a check node in CG_j being zero at the k -th iteration. Since a parity bit incident to a check node in CG_j must belong to one of the group PG_l , $l \in [j, j+a]$, therefore

$$\begin{aligned} \hat{e}_{p,j}^{(k)} &= \sum_{l=j}^{j+a} \nu_{j,l} e_{p,l}^{(k)} \\ &= \varepsilon_{p,j}^{(k)}. \end{aligned}$$

Appendix 4.C Notation List

- IG_j : a group consisting of information nodes of degree $j \in [2, d_l]$;
- d_l : maximum degree of information nodes;
- CG_j : a group consisting of check nodes connected to $j \in [1, a]$ information nodes which are associated with the transmitted bits;
- PG_l : a group consisting of parity nodes which connect to check node pairs incident to $l \in [0, 2a]$ information nodes associated with the transmitted bits;
- ρ_j : the fraction of edges connecting the check nodes in CG_j to the information nodes;
- ζ_j : the fraction of edges connecting the check nodes in CG_j to the parity nodes;
- ν_l : the fraction of edges connecting the parity nodes in PG_l ;
- α_l : the proportion of parity nodes in PG_l ;
- $\delta_l^{(0)}$: the puncturing proportion of parity nodes in PG_l ;
- $g^{(k)}(v_i)$: the pdf of an information-to-check-node message at the k -th iteration;
- $m_{u_i}^{(k)}$: the mean value of a check-to-information-node message;
- $m_{u_p}^{(k)}$: the mean value of a check-to-parity-node message;
- $f^{(k)}(\hat{v}_{p,j})$: the pdf of the message from a parity node to a check node in group CG_j at the k -th iteration;
- $\varepsilon_i^{(k)}$: the probability of a check-to-information-node message being zero at the k -th iteration;
- $\varepsilon_p^{(k)}$: the probability of a check-to-parity-node message being zero at the k -th iteration;
- $\varepsilon_{p,j}^{(k)}$: the probability of a message from a check node in CG_j to a parity node being zero at the k -th iteration;
- $e_p^{(k)}$: the probability of a parity-to-check-node message being zero at the k -th iteration;

- $e_{p,l}^{(k)}$: the probability of a message from a parity node in PG_l to a check node being zero at the k -th iteration;
- $\hat{e}_{p,j}^{(k)}$: the probability of a message from a parity node to a check node in CG_j being zero at the k -th iteration;
- $\psi_{p,l}^{(k)}$: the probability of a message from a check node to a parity node in PG_l being zero at the k -th iteration;
- $\tilde{\zeta}_j$: the fraction of edges connecting the check nodes in CG_j to the parity nodes after the shortening step;
- $\tilde{\nu}_l$: the fraction of edges connecting the parity nodes in PG_l after the shortening step;
- $\tilde{\alpha}_l$: the proportion of parity nodes in PG_l after the shortening step;
- $\pi_{i,j}^{(0)}$: the puncturing proportion of information bits in IG_j ;
- $\pi_{p,j}^{(0)}$: the puncturing proportion of parity bits in PG_j ;
- $\tilde{\varepsilon}_i^{(k)}$: the probability of a check-to-information-node message being zero in the k -th iteration after the shortening step;
- $\tilde{\varepsilon}_p^{(k)}$: the probability of a check-to-parity-node message being zero at the k -th iteration after the shortening step;
- $\tilde{\varepsilon}_{p,j}^{(k)}$: the probability of a message from a check node in CG_j to a parity node being zero at the k -th iteration after the shortening step;
- $\tilde{e}_p^{(k)}$: the probability of a parity-to-check-node message being zero at the k -th iteration after the shortening step;
- $\tilde{e}_{p,l}^{(k)}$: the probability of a message from a parity node in PG_l to a check node being zero at the k -th iteration after the shortening step;
- $\tilde{\tilde{e}}_{p,j}^{(k)}$: the probability of a message from a parity node to a check node in CG_j being zero at the k -th iteration after the shortening step;
- $\tilde{\psi}_{p,l}^{(k)}$: the probability of a message from a check node to a parity node in PG_l being zero at the k -th iteration after the shortening step.

Chapter 5

Conclusion

In this dissertation, we investigated different space-time processing schemes using multiple antennas at both the transmitter and the receiver. Multiple antennas were employed to achieve diversity gain, reduce multiple access interference, or significantly increase spectral efficiency, depending on the specific scheme.

We first examined a MC-DS-CDMA system employing multiple antennas at both the mobile and the base station. Assuming perfect channel state information at the transmitter, maximal ratio transmission was used to achieve the diversity for combating fading. The receive antenna array was assumed to be correlated, and adaptive beamforming reception was used to achieve the maximum received SINR for the desired user, optimally in terms of joint fading reduction and MAI suppression. The conditional SINR was analytically derived and the average BER was investigated via simulation. By varying the number of transmit antennas, receive antennas and subcarriers, we found a tradeoff between obtaining diversity gain against fading and MAI suppression. Considering the intensive computation involved in optimizing SINR, we designed the adaptive beamformer maximizing the received SNR instead of SINR. To validate this simplification, we compared the performance obtained by using the two criteria in different scenarios and conjectured scenarios to which this simplification was appropriate.

Next, we designed a per-antenna-coded V-BLAST type system with adaptive modulation and powerful channel codes, specifically, LDPC codes, to approach the V-BLAST capacity with rate and/or power adaptations. The data stream corresponding to each transmit antenna was independently encoded and mapped to an M-ary QAM symbol. The received signal was the superposition of the signals from all the transmit antennas and the noise. Optimal succes-

sive detection (OSD) was used to detect individual signals, followed by LDPC decoding. The constellation size of each transmission was selected according to the channel state, i.e., the instantaneous capacity based on OSD. Then the density evolution was used to design the LDPC code with the maximal achievable rate for each antenna in each channel realization. Finally, we calculated the average sum rate of this adaptively modulated LDPC-encoded V-BLAST type system by averaging the instantaneous sum rate over a large number of channel realizations. Comparing the average sum rate to the ergodic capacity, we demonstrated that the proposed system is a practical implementation to approach the V-BLAST capacity with rate and/or power adaptations. For systems with delay constraints, outage capacity instead of ergodic capacity is the appropriate theoretical measure. Using the methodology introduced above, we also designed LDPC codes to approach the outage capacity of a 2×2 V-BLAST system for a specified outage probability.

In the coded-V-BLAST scheme, the optimally designed LDPC codes may have distinct degree distributions, meaning that distinct parity check matrices and Tanner graph structures may be needed, leading to significant implementation complexity for encoding and decoding. To alleviate this problem, we used rate-compatible punctured LDPC codes instead of the individually designed optimal LDPC codes. The code with the lowest rate was chosen as the mother code and all the other higher rate codes were obtained by puncturing the mother code. Linear programming was used to derive the puncturing distributions of the variable nodes so that the optimal puncturing fraction, determined by the relation between the mother code rate and the optimal code rate for a given channel realization, was approached.

We simulated the BER performance for the designed system using optimal LDPC codes as well as punctured LDPC codes. Since the assumption of perfect feedback in OSD does not hold in a real system, error propagation can significantly degrade the system performance. Decoding delay is another drawback of the OSD algorithm in a real setting. Thus, we chose PIC followed by MMSE filtering to suppress the interference. Simulation results showed that the BERs of both systems using the optimal LDPC codes and punctured LDPC codes are very close to the target BER.

We explored the possibility of designing rate-compatible punctured codes for a practical system with physical constraints, such as fixed symbol frame length. Since the constellation size and channel code rate have to be adapted to match the channel state, the lengths of the transmitted frames measured in bits can be different. Thus, the parity check matrices for encoding

and decoding the punctured codes are different, which makes using punctured codes from one mother code highly impractical. However, irregular repeat accumulate (IRA) codes, make it possible to design punctured codes of different rates and different block lengths from one underlying mother code. In IRA codes, the information nodes and parity nodes are explicitly designated. Exploiting this property, we proposed a two-step shortening/puncturing process that makes it possible to get all the necessary rate-compatible punctured IRA codes of different lengths from a single mother code.

First we expunged the information bits and parity bits proportionally so that the code rate was (at least approximately) preserved and the length of the shortened code satisfies the constraint imposed by the target modulation alphabet, code rate and frame length (in symbols) for each channel realization. Specifically, we expunged the information bits of each degree in proportion to the respective cardinality to (at least approximately) maintain the optimal information node degree distribution. Then, we used Gaussian approximation (GA) analysis to obtain the shortening distribution of the parity bits to approach the desired shortening fraction while satisfying the convergence condition. The resulting distributions showed that it is preferable to expunge parity bits connected to the least number of information nodes which are associated with the transmitted bits.

After obtaining the shortened mother code, we once again used GA analysis to derive the optimal puncturing distributions of the information bits and the parity bits. Examining the puncturing distributions, we found that the puncturing proportion of the parity bits is much greater than that of the information bits for a given puncturing fraction. We also noticed that the puncturing fraction of the information bits hardly increased even when the overall puncturing fraction increased significantly. Actually the increase of the overall puncturing fraction was mainly reflected in that of the puncturing fraction of the parity bits.

Finally, we calculated the average sum rate of the adaptively modulated IRA-encoded V-BLAST systems using rate-compatible punctured IRA codes, which was very close to the ergodic capacity of the V-BLAST system using OSD. Performance simulation result confirmed that the designed system achieved the target BER.

Bibliography

- [1] H. Huang, H. Viswanathan, and G. J. Foschini, "Multiple antennas in cellular cdma systems: Transmission, detection and spectral efficiency," *IEEE Transactions on Wireless Communications*, vol. 1, pp. 383–392, July 2002.
- [2] A. Lozano and A. M. Tulino, "Capacity of multiple-transmit multiple-receive antenna architectures," *IEEE Transactions on Information Theory*, vol. 48, pp. 3117–3128, December 2002.
- [3] J. Winters, "On the capacity of radio communication systems with diversity in a Rayleigh fading environment," *IEEE Journal on Selected Areas in Communications*, vol. 5, pp. 871–878, June 1987.
- [4] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multiple antennas," *The Bell Labs Technical Journal*, vol. 1, pp. 41–59, August 1996.
- [5] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, pp. 585–595, November 1999.
- [6] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Transactions on Information Theory*, vol. 45, pp. 139–157, January 1999.
- [7] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, pp. 1451–1458, October 1998.
- [8] N. Al-Dhahir, C. Fragouli, A. Stamoulis, W. Younis, and R. Calderbank, "Space-time processing for broadband wireless access," *IEEE Communications Magazine*, vol. 50, pp. 2477–2488, October 2002.
- [9] T. L. Marzetta and B. M. Hochwald, "Unitary space-time modulation for multiple-antenna communications in Rayleigh flat fading," *IEEE Transactions on Information Theory*, vol. 46, pp. 543–564, March 2000.

- [10] B. Hassibi and B. Hochwald, "Cayley differential unitary space-time codes," *IEEE Transactions on Information Theory*, vol. 48, pp. 1485–1503, June 2002.
- [11] M. Hochwald, T. L. Marzetta, T. J. Richardson, W. Sweldens, and R. Urbanke, "Systematic design of unitary space-time constellations," *IEEE Transactions on Information Theory*, vol. 46, pp. 1962–1973, September 2000.
- [12] G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communication," *IEEE Transactions on Communications*, vol. 46, pp. 357–366, March 1998.
- [13] G. J. Foschini, D. Chizhik, M. Gans, C. Papadias, and R. A. Valenzuela, "Analysis and performance of some basic spacetime architectures," *IEEE Journal on Selected Areas in Communications, Special Issue on MIMO Systems, part I*, vol. 21, pp. 303–320, April 2003.
- [14] H. Böleskei, D. Gesbert, and A. J. Paulraj, "On the capacity of OFDM-based spatial multiplexing systems," *IEEE Transactions on Communications*, vol. 10, pp. 1926–1934, February 2002.
- [15] A. Lozano and C. Papadias, "Layered space-time receivers for frequency-selective wireless channels," *IEEE Transactions on Communications*, vol. 50, pp. 65–73, January 2002.
- [16] J. Winters, "Smart antennas for wireless systems," *IEEE Personal Communications*, vol. 5, pp. 23–27, February 1998.
- [17] W. C. Y. Lee, "Spectrum and technology of a wireless local loop system," *IEEE Personal Communications*, vol. 5, pp. 49–54, February 1998.
- [18] A. F. Naguib and A. Paulraj, "A base-station antenna array receiver for cellular DS/CDMA with M-ary orthogonal modulation," in *Proceedings Asilomar Conference on Signals, Systems & Computers*, (Asilomar, CA, USA), pp. 858–862, October – November 1995.
- [19] R. Kohno, "Spatial and temporal communication theory using adaptive antenna array," *IEEE Personal Communications*, vol. 5, pp. 28–35, February 1998.
- [20] F. Rashid-Farrokhi, K. R. Liu, and L. Tassiulas, "Transmit beamforming and power control for cellular wireless systems," *IEEE Journal on Selected Areas in Communications*, vol. 16, pp. 1437–1450, October 1998.
- [21] S. Simon and A. Moustakas, "Optimizing mimo antenna systems with channel covariance feedback," *IEEE Journal on Selected Areas in Communications*, vol. 21, pp. 406–417, April 2003.
- [22] V. Tarokh, A. F. Naguib, N. Seshadri, and A. R. Calderbank, "Combined array processing and space-time coding," *IEEE Transactions on Information Theory*, vol. 45, pp. 1121–1128, May 1999.

- [23] S. Shamai and S. Verdú, "Space-time block codes: A capacity perspective," *IEEE Communications Letters*, vol. 4, pp. 384–386, December 2000.
- [24] R. S. Blum, "Some analytical tools for the design of space-time convolutional codes," *IEEE Transactions on Communications*, vol. 50, pp. 1593–1599, October 2002.
- [25] H. El Gamal and A. R. Hammons, Jr., "On the design and performance of space-time codes for BPSK and QPSK modulation," *IEEE Transactions on Communications*, vol. 50, pp. 907–913, June 2002.
- [26] L. Zheng and D. Tse, "Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels," *IEEE Transactions on Information Theory*, vol. 49, pp. 1073–1096, May 2003.
- [27] C. N. Chuah, J. M. Kahn, G. J. Foschini, R. A. Valenzuela, D. Chizhik, and J. Ling, "Capacity growth of multi-element arrays in indoor and outdoor wireless channels," in *IEEE Wireless Communications and Networking Conference*, (Chicago, IL, USA), pp. 23–28, IEEE, September 2000.
- [28] S. Hochwald, M. ten Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Transactions on Communications*, vol. 51, pp. 389–399, March 2003.
- [29] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *IEEE Wireless Personal Communications*, vol. 6, pp. 311–335, March 1998.
- [30] S. Kondo and L. B. Milstein, "Performance of multicarrier DS CDMA system," *IEEE Transactions on Communications*, vol. 44, pp. 238–246, February 1996.
- [31] J. Salz and J. H. Winters, "Effects of fading correlation on adaptive arrays in digital wireless communications," in *Proceedings IEEE International Conference on Communications*, (Geneva, Switzerland), pp. 1768–1774, November 1993.
- [32] T. K. Y. Lo, "Maximum ratio transmission," *IEEE Transactions on Communications*, vol. 47, pp. 1458–1461, October 1999.
- [33] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: an architecture for realizing very high data rates over the rich-scattering wireless channel," in *Proceedings of International Symposium on Signals, Systems, and Electronics*, (Pisa, Italy), pp. 295–300, September 1998.
- [34] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE Journal on Selected Areas in Communications*, vol. 21, pp. 684–702, June 2003.
- [35] A. Goldsmith and S. G. Chua, "Adaptive coded modulation for fading channels," *IEEE Transactions on Communications*, vol. 46, pp. 595–602, May 1998.

- [36] M. K. Varanasi and T. Guess, "Optimum decision feedback multiuser equalization with successive decoding achieves the total capacity of the Gaussian multi-access channel," in *Proceedings Asilomar Conference on Signals, Systems & Computers*, (Monterey, CA, USA), pp. 1405–1409, November 1997.
- [37] D. N. C. Tse and S. V. Hanly, "Multiaccess fading channels—part I: Polymatroid structure, optimal resource allocation and throughput capacities," *IEEE Transactions on Information Theory*, vol. 44, pp. 2796–2815, November 1998.
- [38] T. J. Richardson, M. A. Shokrollahi, and R. L. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Transactions on Information Theory*, vol. 27, pp. 619–637, February 2001.
- [39] S. T. Chung, A. Lozano, H. C. Huang, A. Sutivong, and J. M. Cioffi, "Approaching the MIMO capacity with a low-rate feedback channel in V-BLAST," *EURASIP Journal on Applied Signal Processing*, pp. 762–771, May 2004.
- [40] X. Wang and H. V. Poor, "Iterative (turbo) soft interference cancellation and decoding for coded CDMA," *IEEE Transactions on Communications*, vol. 47, pp. 1046–1061, July 1999.
- [41] E. Biglieri, J. Proakis, and S. S. Shitz, "Fading channels: Information theoretic and communication aspects," *IEEE Transactions on Information Theory*, vol. 44, pp. 2619–2692, October 1998.
- [42] A. I. Vila Casado, W. Y. Weng, and R. D. Wesel, "Multiple rate low-density parity-check codes with constant blocklength," in *Proceedings 42th Annual Allerton Conference on Communication, Control, and Computing*, (Monticello, IL, USA), pp. 2010–2014, October 2004.
- [43] H. Jin, A. Khandekar, and R. J. McEliece, "Irregular repeat-accumulate codes," in *Proceedings International Symposium on Turbo Codes & Related Topics*, (Brest, France), September 2000.
- [44] S. Chung, T. J. Richardson, and R. L. Urbanke, "Analysis of sum-product decoding of low-density parity-check codes using a Gaussian approximation," *IEEE Transactions on Information Theory*, vol. 47, pp. 657–670, February 2001.
- [45] H. El Gamal and A. R. Hammons, Jr., "A new approach to layered space-time coding and signal processing," *IEEE Transactions on Information Theory*, vol. 47, pp. 2321–2334, September 2001.
- [46] W. Rhee and J. M. Cioffi, "On the asymptotic optimality of beamforming in multi-antenna gaussian multiple access channels," in *Proceedings IEEE Global Telecommunications Conference*, vol. 2, (San Antonio, TX, USA), pp. 891–895, November 2001.

- [47] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Transactions on Information Theory*, vol. 44, pp. 744–765, March 1998.
- [48] B. D. V. Veen and K. M. Buckley, "Beamforming: a versatile approach to spatial filtering," *IEEE Acoustics, Speech, and Signal Processing Magazine*, vol. 36, pp. 4–24, April 1988.
- [49] S. Catreux, P. F. Driessen, and Greenstein, "Data throughputs using multiple-input multiple-output (MIMO) techniques in a noise-limited cellular environment," *IEEE Transactions on Wireless Communications*, vol. 1, pp. 226–235, April 2002.
- [50] S. T. Chung, A. Lozano, and H. C. Huang, "Approaching eigenmode BLAST channel capacity using V-BLAST with rate and power feedback," in *Proceedings IEEE Vehicular Technology Conference*, vol. 2, (Atlantic City, NJ, USA), pp. 915–919, October 2001.
- [51] A. J. Goldsmith and P. P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Transactions on Information Theory*, vol. 43, pp. 1986–1992, November 1997.
- [52] T. J. Richardson and R. L. Urbanke, "The capacity of low-density parity check codes under message-passing decoding," *IEEE Transactions on Information Theory*, vol. 47, pp. 599–618, February 2001.
- [53] J. Hou, P. H. Siegel, L. B. Milstein, and H. D. Pfister, "Capacity-approaching bandwidth-efficient coded modulation schemes based on low-density parity-check codes," *IEEE Transactions on Information Theory*, vol. 49, pp. 2141–2155, September 2003.
- [54] J. Ha, J. Kim, and S. W. McLaughlin, "Rate-compatible puncturing of low-density parity-check codes," *IEEE Transactions on Information Theory*, vol. 50, pp. 2824–2836, November 2004.
- [55] T. Guess, H. Zhang, and T. V. Kottchiv, "The outage capacity of BLAST in MIMO channels," in *Proceedings IEEE International Conference on Communications*, vol. 4, (Anchorage, AK, USA), pp. 2628–2632, May 2003.
- [56] T. Richardson and V. Novichkov, "Methods and apparatus for decoding LDPC codes." U.S. Patent 6,633,856, October 2003.
- [57] H. Jin, *Analysis and Design of Turbo-like Codes*. PhD thesis, Caltech, May 2001.
- [58] A. Roumy, S. Guemghar, G. Carie, and S. Verdú, "Design methods for irregular repeat-accumulate codes," *IEEE Transactions on Information Theory*, vol. 50, pp. 1711–1721, August 2004.