CONCATENATED CODES FOR BLOCK-FADING CHANNELS

THÈSE N° 3030 (2004)

PRÉSENTÉE À LA FACULTÉ D'INFORMATIQUE ET COMMUNICATIONS

Institut Eurécom

SECTION DES SYSTÈMES DE COMMUNICATION

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Albert GUILLÉN i FÀBREGAS
ingénieur de télécommunications, Universitat Politècnica de Catalunya, Espagne
dottore in Ingegneria Elettronica, Politecnico di Torino, Italie
et de nationalité espagnole

accepté sur proposition du jury:

Prof. G. Caire, directeur de thèse
Prof. J. Boutros , rapporteur
Prof. R. Knopp , rapporteur
Prof. G. Taricco , rapporteur
Prof. R. Urbanke, rapporteur

Lausanne, EPFL
2004

Version Abrégée

Nous étudions dans cette thèse des techniques et des stratégies de codage de canal adaptées aux canaux à évanouissements par bloc dans un système mono-utilisateur. Le canal à évanouissements par bloc est un modèle simplifié du canal de communication sans fil qui varie lentement dans le temps avec de fortes contraintes sur le délai de transmission. Ce modèle est particulièrement intéressant pour les applications pratiques lorsque la communication utilise un nombre *fini* de degrés de liberté du canal. Ce nombre est habituellement appelé *ordre de diversité*. Du point de vue de l'ingénieur, le but final est de concevoir des stratégies de transmission qui exploitent efficacement les degrés de liberté disponibles pour améliorer la qualité et la fiabilité des communications. Les lents sauts en fréquence et en temps, typiques des sytèmes GSM ou EDGE, et la modulation à porteuses multiples utilisant le multiplexage par fréquences orthogonales (OFDM) sont des exemples pratiques de tels systèmes de communication. Contrairement aux canaux de communication standards, le canal à évanouissements par bloc a une capacité nulle au sens de Shannon parce qu'il subsiste une probabilité, appelée probabilité de coupure "*outage*", que le débit d'information ne soit pas supporté par le canal. Par conséquent, pour des blocs à grande longueur, la probabilité d'erreur, c'est à dire, la probabilité que l'information émise soit mal décodée par le récepteur, sera au moins aussi grande que la probabilité de coupure.

Dans le cas où émetteur et récepteur ont tous les deux une seule antenne pour communiquer, nous examinons les limites théoriques de tels canaux à évanouissements par bloc. En étudiant le comportement asymptotique de la probabilité de coupure, nous montrons que les performances des schémas de codage physiquement réalisables, construits par des constellations discrètes de signaux de modulation, sont déterminées par une borne fondamentale sur l'ordre de diversité atteignable. Cela aboutit au compromis débit-diversité optimal. Ce compromis optimal induit le concept d'*expansion de la constellation des signaux* étant donné qu'il peut être avantageux de transmettre avec de grandes constellations pour atteindre de meilleures performances. Nous introduisons une nouvelle famille de codes concaténés qui approchent les limites mentionnées ci-dessus pour des constellations arbitraires en utilisant des algorithmes de décodage probabilistique itératif avec une complexité constante par bit d'information. Nous analysons les performances d'ensemble sous décodage optimal à maximum de vraissemblance et nous montrons que le décodage itératif a des performances très proches du décodeur optimal, contrairement au cas des canaux gaussiens sans évanouissements et ceux avec évanouissements parfaitement entrelacés. Nous montrons que les codes proposés ont des performances proches de la probabilité de coupure du canal pour toutes les longueurs de bloc alors que les codes standards basés sur les treillis et spécialement conçus pour le canal à évanouissements par bloc présentent une dégradation importante quand la longueur de bloc augmente. De plus, nous montrons que la structure de codage proposée dépasse en performance les codes concaténés parallèles ou séries conventionnels.

Nous étudions aussi des stratégies de codage pour les canaux à évanouissements par bloc avec des antennes multiples. Sur de tels canaux, selon les critères traditionnels, la construction de schémas de codage efficaces atteignant la diversité maximale est difficile et conduit habituellement à des algorithmes de décodage complexes. Nous proposons une conception plus simple basée sur une construction pragmatique qui conduit à des algorithmes de décodage itératif à faible complexité. Nous nous concentrons sur l'impact de l'expansion de la constellation sur l'ordre maximal de diversité. En particulier, nous étudions deux méthodes différentes pour atteindre la diversité maximale en adaptant la taille de la constellation. Nous considérons l'expansion classique dans le plan complexe selon Ungerböck et l'expansion multi-dimensionnelle basée sur des réseaux de points. Au moyen de différents décodeurs échangeant de messages souples, nous comparons les deux approches et nous montrons que l'expansion multi-dimensionnelle est, en général, toujours avantageuse en raison de sa grande flexibilité de conception. Nous donnons une évaluation complète des performances ainsi qu'une comparaison avec les schémas classiques d'Alamouti et V-BLAST pour le cas intéressant d'un canal OFDM à antennes multiples, typique des standards de la prochaine génération de réseaux locaux sans-fil.

Abstract

In this thesis we study coding strategies for single-user block-fading channels. The block-fading channel is a simplified model of slowly varying wireless communication channels with stringent delay constraints. This model is particularly important in practical applications where communication spans a *finite* number of degrees of freedom of the channel. The number of such degrees of freedom is usually referred to as *diversity*. From an engineering perspective, the ultimate goal is to design transmission strategies that efficiently exploit the available degrees of freedom in order to communicate more reliably. Slow time-frequency hopping, typical of GSM or EDGE systems, and multi-carrier modulation employing orthogonal division multiplexing (OFDM), are practical examples of such communication systems. As opposed to standard communication channels, the block-fading channel has zero capacity in the strict Shannon sense, since there is an irreducible probability that the transmitted data rate is not supported by the channel, namely the *information outage probability*. Therefore, for large block length, the probability of error, i.e., the probability that the transmitted information is decoded in error at the receiver, will be at least as large as the outage probability.

In the case of both transmitter and receiver having one single antenna to communicate, we examine the theoretical limits of such block-fading channels. By studying the asymptotic behavior of the outage probability, we show that the performance of practical coding schemes, constructed over discrete signal constellations, is determined by a fundamental bound on the achievable diversity. This gives rise to the optimal rate-diversity trade-off. This optimal trade-off induces the concept of *signal constellation expansion*, since, in order to achieve better performance it may be advantageous to transmit with large signal sets. We introduce a new family of concatenated codes that approach the aforementioned limits for arbitrary signal constellations with constant decoding complexity per information bit using message passing iterative decoding algorithms. We analyze the ensemble performance under optimal maximum-likelihood decoding and we show that iterative message passing decoding performs very close to the optimal decoder, differently from non-faded AWGN and fully-interleaved fading channels. We show that the proposed codes perform close to the the outage probability of the channel for any block length, while standard trellis-based codes specifically designed for the block-fading channel exhibit significant degradation as the block length gets large. Moreover, we show that the proposed coding structure significantly outperforms conventional parallel or serial concatenated codes.

We also study coding strategies in multiple-antenna block-fading channels. For such multiple-antenna channels, constructing efficient coding schemes that achieve full diversity based traditional code design criteria is difficult and usually leads to complex decoding algorithms. We propose simpler designs based on pragmatic constructions that lead to low-complexity iterative message passing decoders. We focus on the impact of signal constellation expansion on the achievable diversity. In particular, we study two different approaches of achieving full diversity by appropriately expanding the signal constellation. We consider classical complex-plane Ungerböck's style expansion and multidimensional lattice-like expansion. By means of several message passing decoders, we compare both approaches and we show that, in general, multidimensional expansion is always advantageous due to its higher design flexibility. We report extensive performance evaluation and comparison of the proposed coding schemes with classical schemes, namely Alamouti and V-BLAST, for the relevant case of an OFDM multiple-antenna channel arising in next generation wireless local-area network standards.

Acknowledgment

First and foremost, I am most grateful to my advisor Prof. Giuseppe Caire for his supervision, guidance and support throughout the duration of this work. His passion and dedication for research, musician thoroughness and great insight have been of enormous value to this work and also to myself. It has been a real pleasure to work with him. I have learned a lot from all discussions we had along my stay at Eurecom and I really hope that our collaboration will continue in the future. I am also very grateful to Prof. Giuseppe Caire for his support and encouragement before and during my visit to EPFL.

I am also very grateful to Prof. Joseph Boutros, Prof. Raymond Knopp, Prof. Giorgio Taricco, Prof. Emre Telatar and Prof. Rüdiger Urbanke for accepting to be part of my thesis comitee, take time to read the first version of the thesis and give detailed comments. It has been a great pleasure and a real honor to have them in my thesis defense comitee. I have very much appreciated Prof. Joseph Boutros' feedback also on smoothing the French abstract of the thesis.

I would like to thank my colleagues and friends at Eurecom for the good time I had there and for the interesting discussions. In particular, I would like to thank Alessandro Nodrio for telling me some of his best programming tricks, Prof. Raymond Knopp for his comments on some papers and interesting discussions we had and Prof. Dirk Slock, Maxime Guillaud and Abdelkader Medles for the interesting discussions we had while working for the ANTIPODE project. I also would like to thank Chris Fischer, Souad Guemghar, Maxime Guillaud, Jussi Kangasharju, Mari Kobayashi, Max Lenardi, David Mary, Abdelkader Medles, Navid Nikaein, Alessandro Nordio, Cristina Russo, Stefania Sesia, Daniela Tuninetti and Samuli Visuri for friendship and kindness during all this time. I thank also all people at Eurecom for providing such a good athmosphere. I am also indebted to Silvia Almagia, Iñaki Gimeno, Jordi Vilanova (and lately the little Ainhoa), my *family* in the French Riviera. I have no words to express my enormous gratitude to them. I also would like to thank Carolina Pinart for the personalized coffee cup (even though I have mostly drunk tea on it) and her unconditional enthusiastic support before I started this PhD.

I am very grateful to Prof. Rüdiger Urbanke for hosting me in his lab during the last stage of this work at EPFL. It has been an experience of enormous value to me, from both professional and personal points of view. I would like to thank all

LTHC and LTHI members for the warm welcome I had, specially Abdelaziz Amraoui, Muriel Bardet, Denice Deatrich, Olivier Leveque, Cyril Measson, Sibi Raj, Aslan Tchamkerten and Emre Telatar. I also with to thank Daniela Tuninetti, from LCM, for her friendship, support and enormous help. I have very much appreciated the feedback from Abdelaziz Amraoui, Cyril Measson and Sibi Raj on earlier versions of this repport. Cyril's help for writing the first version of the French abstract of this thesis is enormously appreciated.

I would like to thank Riccardo De Gaudenzi and Alfonso Martinez for stimulating discussions that originated during my stay in The Netherlands and continued until present, even when not strictly related to the main research topic presented in this work.

I also would like to acknowledge Prof. Ezio Biglieri and Prof. Sergi Verdú for their constant support as well as Prof. Gregori Vázquez for his enthusiastic encouragement and help.

Last but not least, I would like to thank Maria and my family for love, support and constant encouragement.

CONTENTS

1 Introduction and Thesis Outline 3

2 The Block-Fading Channel: Background 7

2.1 The SISO Block-Fading Channel 7

2.1.1 System Model 8

2.1.2 Fundamental Limits 9

2.1.3 Code Design 14

2.2 The MIMO Block-Fading Channel 16

2.2.1 System Model 17

2.2.2 Fundamental Limits 18

2.2.3 Code Design 21

3 Coded Modulation in the SISO Block-Fading Channel 25

3.1 Introduction 26

3.2 System model 28

3.3 Optimal rate-diversity trade-off 30

3.4 Blockwise concatenated coded modulation 34

3.4.1 Code construction 34

3.4.2 Upper bounds on ML decoding error probability 37

3.4.3 Weak goodness of BCC ensembles 43

3.4.4 On code optimization 49

3.5 Conclusions 52

4 Coded Modulation in the MIMO Block-Fading Channel 55

4.1 Pragmatic Space-Time Codes 56

4.2 Pragmatic Concatenated Space-Time Codes 58

4.2.1 Extension to General Block-Fading Channels 60

4.3 Message Passing Decoding 62

4.3.1 Belief Propagation 64

4.3.2 Soft-Input Soft-Output Sphere Decoding 66

4.3.3 Lower-Complexity Approximations 67

4.4 Numerical Examples 71

4.4.1 Block Diversity vs. Rank Diversity 71
4.4.2 Impact of Signal Constellation Expansion 72
4.5 Conclusions 75

5 Application to OFDM 77
5.1 OFDM with Cyclic Prefix 78
5.2 HIPERLAN/2 Physical Layer 79
5.2.1 Environment Description 81
5.3 BICM NSTC for OFDM 82
5.4 Performance Comparison with Other Schemes 83
5.4.1 Comparison in the Independent BRAN-A Channel 85
5.4.2 Comparison in the Realistic NLOS Indoor Channel 86
5.5 Conclusions 88

6 Conclusions and Future Work 93
6.1 Areas for Further Research 94

A Error Probability of BICM with the Gaussian Approximation 97
A.1 Introduction and Motivation 97
A.2 System model 98
A.3 The Bhattacharyya Union Bound 100
A.4 The Gaussian Approximation 101
A.5 Approximations on BICM-ML Error Probability 103
A.6 BICM-ML Thresholds for Turbo-coded BICM 108
A.7 Conclusions 111

B Multivariate Weight Enumerators 113
B.1 MWEF of BCCs 113
B.2 Asymptotic MWEF of BCCs 114
B.3 PWEF of BCCs 117

C Proofs 119
C.1 Proof of Lemma 1 119
C.2 Proof of Theorem 1 122
C.3 Proofs for section 3.4.3 127

D Tangential-Sphere Bound for Block-Fading Channels 131

E Approximated Density Evolution for MIMO IC receivers 137

ACRONYMS

Here we list the main acronyms used in this document. The meaning of the acronym is usually indicated once, the first time it appears in the text.

ADSL	Asymmetric Digital Subscriber Line
ANTIPODE	ANTennes Intelligentes POur le haut DEbit avec HIPERLAN/2
APP	A Posteriori Probability
ATM	Asynchronous Transfer Mode
AWGN	Additive White Gaussian Noise
BCC	Blockwise Concatenated Code
BCCC	Blockwise Concatenated Convolutional Code
BCJR	Bahl, Cocke, Jelinek, Raviv
BER	Bit Error Rate
BI	Binary-Input
BICM	Bit-Interleaved Coded Modulation
BP	Belief Propagation
BPL	Byun, Park, Lee
BPSK	Binary Phase-Shift Keying
BRAN	Broadband Radio Access Networks
CC	Convolutional Code
CDMA	Code-Division Multiple Access
CSI	Channel State Information
CSIR	Channel State Information at the Receiver
CSIT	Channel State Information at the Transmitter
DAB	Digital Audio Broadcasting
DE	Density Evolution
DVB	Digital Video Broadcasting
EDGE	Enhanced Data GSM Environment
ETSI	European Telecommunications Standard
EXIT	EXtrinsic Information Transfer
EXT	EXTrinsic information

FFT	Fast Fourier Transform
GA	Gaussian Approximation
GR	Gray
GSM	Global System for Mobile communications
HIPERLAN/2	HIgh PERformance Local Area Network
IC	Interference Cancellation
IFFT	Inverse Fast Fourier Transform
i.i.d.	independent and identically distributed
IOWEF	Input-Output Weight Enumeration Function
ISI	Inter-Symbol Interference
LAN	Local Area Network
LD	Linear Dispersion
LDPC	Low-Density Parity Check
LLR	Log-Likelihood Ratio
MAP	Maximum A Posteriori
MDS	Maximum Distance Separable
MF	Matched Filter
MFB	Matched Filter Bound
MIMO	Multiple-Input Multiple-Output
ML	Maximum-Likelihood
M&L	Malkamaki & Leib
MMSE	Minimum Mean Squared Error
MRC	Maximal Ratio Combining
MSE	Mean Squared Error
MWEF	Multivariate Weight Enumeration Function
NLOS	Non-Line-of-Sight
NSTC	Natural Space-Time Code
OFDM	Orthogonal Frequency Division Multiplexing
pdf	probability density function
PSK	Phase-Shift Keying
PWEF	Product Weight Enumeration Function

QAM	Quadrature-Amplitude Modulation
QPSK	Quadrature Phase-Shift Keying
QRA	Quasi Repeat and Accumulate
RA	Repeat and Accumulate
RBA	Repeat and Blockwise Accumulate
RHS	Right Hand Side
RNRT	Réseau National de Recherche en Télécommunications
SCCC	Serial Concatenated Convolutional Code
SD	Sphere Decoding
SISO	Single-Input Single-Output
SNR	Signal-to-Noise Ratio
SP	Set-Partitioning
TAST	Threaded Algebraic Space-Time
TSTC	Threaded Space-Time Code
TSB	Tangential-Sphere Bound
UB	Union Bound
V-BLAST	Vertical Bell Labs Layered Space-Time
WEF	Weight Enumeration Function
WER	Word Error Rate

NOTATION

We here list the main notations and symbols used throughout the thesis.

$\mathbb{1}\{\mathcal{E}\}$	indicator function of the event $\mathcal{E}$
a	scalar variable
$\mathbf{a}$	vector variable
A_w	weight enumeration function
$a(\omega)$	asymptotic growth rate of the weight enumeration function
$\mathbf{A}$	matrix variable
B	Bhattacharyya coefficient
$\mathcal{C}$	binary code
$\mathbb{C}$	complex numbers
d	Euclidean distance
$\mathrm{diag}(\mathbf{a})$	diagonal matrix with diagonal elements the components of $\mathbf{a}$
d_B	block-diversity
$d_B^\star$	SNR reliability function
d_r	rank diversity
d_{SB}	Singleton Bound on the block diversity
δ	Squared Euclidean distance
Δ_p	product weight
$\mathbb{E}[.]$	expectation operator
$\frac{E_b}{N_0}$	information bit signal-to-noise ratio
$F(\omega)$	growth rate of the weight enumeration function
$\mathbb{F}_2$	$GF(2) = \{0, 1\}$
$\mathcal{F}$	spatial modulation function
$\Gamma(x)$	Gamma function
$\bar{\Gamma}(a, x)$	normalized incomplete Gamma function
$h(x)$	binary entropy function
$\mathbf{I}_m$	identity matrix of dimensions $m \times m$
i	information Hamming weight
ι	normalized information Hamming weight
$J_{\mathcal{X}}$	mutual information of the signal constellation $\mathcal{X}$

$J_{\mathcal{X},BICM}$	BICM mutual information of the signal constellation $\mathcal{X}$
L	number of complex dimensions per block
$\mathcal{L}$	log-likelihood ratio
$M = \log_2 \lvert\mathcal{X}\rvert$	number of bits/symbol of the constellation $\mathcal{X}$
$\mathcal{M}$	coded modulation scheme
μ	binary labeling rule
N_B	number of fading blocks
N_C	number of sub-carriers
N_R	number of receive antennas
N_T	number of transmit antennas
$\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\sigma})$	Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\sigma}$
$\mathcal{N}_{\mathbb{C}}(\boldsymbol{\mu}, \boldsymbol{\sigma})$	complex Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\sigma}$
P_e	error probability
P_{out}	outage probability
$\mathcal{P}$	parsing rule
$\boldsymbol{\pi}$	interleaver permutation
$Q(x)$	Gaussian tail function
r	rate of binary code
R	transmission rate in bit per complex dimension
R	multidimensional sphere radius
$\mathbb{R}$	real numbers
$\rho = \frac{E_s}{N_0}$	signal-to-noise ratio
$\mathcal{S}$	space-time code
$\text{vec}(\mathbf{A})$	matrix $\mathbf{A}$ written in vector form
w	output Hamming weight
ω	normalized output Hamming weight
$\mathcal{X}$	signal constellation
$\mathbb{Z}$	integer numbers
$\sim$	distributed as
$\doteq$	exponential equality
$\dot{\geq}$, $\dot{\leq}$	exponential inequalities
$\lvert\lvert.\rvert\rvert_F$	Frobenius norm
$\lfloor x \rfloor$	largest integer greater than or equal to x

LIST OF FIGURES

2.1 Block diagram of the SISO block-fading channel. 9

2.2 Outage probability with Gaussian inputs for a SISO block-fading channel with $N_B = 2, 4, 16$ and 100 fading blocks per codeword as a function of R for $\frac{E_b}{N_0} = 10$ dB. 12

2.3 Outage probability with Gaussian inputs for a SISO block-fading channel with $N_B = 2, 4, 16$ and 100 fading blocks per codeword as a function of SNR for $R = 2$ bit/s/Hz. 13

2.4 Outage probability with Gaussian inputs for a SISO block-fading channel with $N_B = 4$ fading blocks per codeword as a function of SNR for $R = 1, 2, 3$ and 4 bit/s/Hz. 14

2.5 Multiple transmit and multiple receive antenna environment. . . . 17

2.6 Block diagram of the MIMO block-fading channel model (2.23). . 19

2.7 Outage probability with Gaussian inputs for a MIMO block-fading channel with $N_T = 2$ transmit antennas, $N_R = 2$ receive antennas, $N_B = 1, 2$ and 16 fading blocks per codeword as a function of SNR for $R = 2$ bit/s/Hz. 20

2.8 Optimal diversity vs. multiplexing trade-off in MIMO block-fading channels with $N_T = 4$, $N_R = 4$ and $N_B = 1, 2$. 21

3.1 SNR reliability function and random coding exponents $d_B^{(r)}(R)$ for $N_B = 8$ and $M = 4$. 33

3.2 Outage probability for $N_B = 8$, $R = 1, 1.5, 2$ bit/complex dimension, Gaussian inputs, 8-PSK and 16-QAM modulations. Thick solid lines correspond to Gaussian inputs, thin solid lines to 8-PSK, dashed lines to 8-PSK with BICM (Gray mapping), dashed-dotted lines to 16-QAM and dotted lines to 16-QAM with BICM (Gray mapping). 34

3.3 The general encoder for Blockwise Concatenated Coding. 35

3.4 WER obtained by BP decoding (simulation with 10 iterations) of binary RBA, RA, BCCC and SCCC of rate $R = 1/2$ for $N_B = 2$ and $K = 1024$. 36

3.5 FER $r = 1/2$ BCCCs and CCs mapped over $N_B = 8$ fading blocks. 37

3.6 WER obtained by BP decoding simulation with 10 iterations and ML bounds and approximations for binary RBA of $R = 1/2$ and $K = 256$ over $N_B = 2$ blocks. 39

3.7 WER obtained by BP decoding simulation with 10 iterations and ML bounds for binary RBA of $R = 1/2$ and 1024 information bits per frame, over $N_B = 2$ blocks. 40

3.8 WER obtained by BP decoding simulation with 10 iterations and ML bounds and approximations for RBA with BICM of $r = 1/2$ over $N_B = 2$ blocks with 8-PSK and 16-QAM with Gray mapping. 42

3.9 WER vs. block length (information bits) at $E_b/N_0 = 8$dB for binary BCC, RBA and trellis terminated CCs obtained by simulation (10 iterations of BP decoding for the BCCs and ML decoding for the CCs using the Viterbi algorithm). 44

3.10 Asymptotic error probability (3.28) for a binary rate $r = 1/2$ RBA code mapped over $N_B = 2$ fading blocks and corresponding BP decoding simulation with 30 iterations and $K = 100, 1000$ and 10000 information bits per codeword. 47

3.11 Asymptotic error probability (3.29) for a rate $R = 2$ RBA code mapped over $N_B = 2$ fading blocks with 16-QAM and Gray mapping (BICM) and corresponding BP decoding simulation with 30 iterations for $K = 100, 1000$ and 10000 information bits per codeword. 48

3.12 WER obtained by BP decoding simulation with 10 iterations and BPL approximations for RBA with rate $R = 1/4$ and 100 information bits per frame, over $N_B = 4$ fading blocks. 50

3.13 PWEF growth rate for RBA of rate $R = 1/4$ with 20 (a) and 100 (b) information bits per frame, over $N_B = 4$ blocks. 51

3.14 WER (simulation of BP decoding with 40 iterations) of several BCCs of rate $R = 1/4$ over BPSK, for $N_B = 4$ fading blocks with $K = 1024$ input bits per codeword. 52

3.15 WER (simulation of BP decoding with 40 iterations) of several BCCs of rate $R = 1$ over 16-QAM (BICM with Gray mapping), for $N_B = 4$ fading blocks with $K = 1024$ input bits per codeword. 53

4.1 Transmission scheme of BICM NSTC. 57

4.2 Transmission scheme of pragmatic concatenated space-time code. 59

4.3 Transmission scheme of BICM TAST. 60

4.4 TAST threading (4.10) for $N_T = 4$, $N_L = 4$ and $T = 4$. The numbers indicate the thread index. 61

4.5 TAST threading for $N_T = 4$, $N_B = 2$, $N_L = 4$ and $T = 4$. The numbers indicate the thread index. 61

4.6 Parsing rule for $N_T = 4$, $N_B = 2$, $N = 48$ and $M = 1$. 62

4.7 Dependency graph of BICM NSTC for $r = 1/2$, $N_T = 2$, $M = 2$, $N = 8$ and $L = 2$. 63
4.8 Iterative decoder with the sum-product algorithm for BICM NSTC. 65
4.9 Iterative decoder with the sum-product algorithm for BICM TAST. 65
4.10 Interference cancellation and filtering. 68
4.11 Iterative decoder with MMSE interference cancellation for BICM NSTC. 70
4.12 Iterative decoder with MMSE interference cancellation for BICM TAST. 70
4.13 WER for $N_T = 4$, $N_R = 4$ and $R = 1$bit/s/Hz, with the $(5, 7, 7, 7)_8$ convolutional code and BPSK modulation. 72
4.14 WER for $N_T = 2$ and $N_R = 2$ with $(5, 7)_8$ convolutional code, QPSK and 16-QAM with Gray mapping. 73
4.15 WER for $N_T = 2$, $N_R = 2$ and $R = 3$bit/s/Hz, with 4 states convolutional codes, QPSK and 8-PSK with Gray mapping. . . . 74
4.16 WER for $N_T = 4$, $N_R = 4$ with $(5, 7, 7, 7)_8$ convolutional code, 16-QAM and 64-QAM with Gray mapping, with MMSE-IC decoding. 75

5.1 Normalized autocorrelation of the OFDM channel of BRAN-A and Ray Tracing impulse responses. 80
5.2 Basic transmission scheme of HIPERLAN/2. 80
5.3 WER for $N_T = N_R = 2$, $(5, 7)_8$ code with BPSK, OFDM BRAN-A and Ray Tracing models, for several N_{VB}. 84
5.4 WER for $N_T = N_R = 2$, $R = 1$ bit/s/Hz with OFDM BRAN-A channel. 86
5.5 WER for $N_T = N_R = 2$, $R = 2$ bit/s/Hz with OFDM BRAN-A channel. 87
5.6 WER for $N_T = N_R = 4$, $R = 2$ bit/s/Hz with OFDM BRAN-A channel. 87
5.7 WER for $N_T = N_R = 4$, $R = 4$ bit/s/Hz with OFDM BRAN-A channel. 88
5.8 WER for $N_T = N_R = 2$, $R = 1$ bit/s/Hz with the realistic NLOS indoor OFDM channel generated with the ray tracing program simulator. 89
5.9 WER for $N_T = N_R = 2$, $R = 2$ bit/s/Hz with the realistic NLOS indoor OFDM channel generated with the ray tracing program simulator. 89
5.10 WER for $N_T = N_R = 4$, $R = 2$ bit/s/Hz with the realistic NLOS indoor OFDM channel generated with the ray tracing program simulator. 90

5.11 WER for $N_T = N_R = 4$, $R = 4$ bit/s/Hz with the realistic NLOS indoor OFDM channel generated with the ray tracing program simulator. 90

A.1 Capacity for 16-QAM inputs and BICM capacity with Gray mapping. 100
A.2 Signal-to-noise ratio scaling of the equivalent BI-AWGN channel for BICM with 16-QAM and Gray and Set-Partitioning mapping. 103
A.3 Simulated pdf of the LLR given that a 0 was transmitted for BICM with 16-QAM and Gray mapping at $\rho = 5$dB. 104
A.4 Simulated pdf of the LLR given that a 0 was transmitted for BICM with 16-QAM and Gray mapping at $\rho = 10$dB. 104
A.5 Simulated pdf of the LLR given that a 0 was transmitted for BICM with 16-QAM and Gray mapping at $\rho = 20$dB. 105
A.6 BER ML BICM Bounds for the 64 states convolutional code of rate 1/2 with 16-QAM and Gray mapping. 107
A.7 BER ML BICM Bounds for the a repeat-and-accumulate code of rate 1/4 and $K = 512$ with 16-QAM with Gray mapping. 108
A.8 BER ML BICM Bounds for the a RA code of rate 1/4 and $K = 1024$ with 16-QAM with Gray mapping. 109
A.9 BER ML BICM Bounds for the a QRA code of rate 1/4 and $K = 1024$ with 16-QAM with Gray mapping. 109

B.1 Growth rate of the MWEF, $F^{RBA}_{\omega_1,\ldots,\omega_{N_B}}$ for a rate $r = 1/2$ RBA code of blocklength 100 information bits per frame mapped over $N_B = 2$ fading blocks. 116
B.2 Asymptotic growth rate $a^{RBA}_{\omega_1,\ldots,\omega_{N_B}}$ for a rate $r = 1/2$ RBA code mapped over $N_B = 2$ fading blocks. 116

D.1 Illustration of the geometry of the tangential sphere bound for the block-fading channel. 133

E.1 BER evolution with the iterations for a fixed $\mathbf{H}$ for $E_b/N_0 = 0$dB and $(5, 7, 7, 7)_8$ CC. 140
E.2 Snapshot for fixed $\mathbf{H}$ (the same of Figure E.1) of the residual interference variance evolution for $E_b/N_0 = 0$dB and $(5, 7, 7, 7)_8$ code. 140
E.3 Average BER and WER of BICM NSTC and NSTC $(5, 7, 7, 7)_8$ code. 141

LIST OF TABLES

5.1 Transmission modes for HIPERLAN/2. 81
5.2 Delay intensity profile of the ETSI BRAN-A channel model. 82
5.3 Delay intensity profile of the normalized BRAN-A channel model assuming a zero-hold order transmission filter. 83

A.1 Upper bounds on the BICM-ML decoding $\left|\frac{E_b}{N_0}\right|_{\mathrm{th}}$ thresholds (A.20) using the Gaussian approximation for 16-QAM with Gray mapping compared to the BICM capacity limit (A.4). 110

1

Introduction and Thesis Outline

Wireless communication systems are characterized by time varying communication channels. These channel time variations are mainly due to multipath propagation, user mobility and changes in the environment. Such variations are usually called *fading*. If the transmitted signal bandwidth is large, the different frequency subbands undergo significantly different fading attenuation. Equivalently, the transmitted signal is spread in the time domain, i.e., the channel induces inter-symbol interference (ISI). These effects usually referred to as frequency selectivity or time dispersion. Moreover, the relative motion between the transmitter and the receiver determines the time variation of the channel characteristics.

In this thesis, we study data transmission over wireless local area network systems, for which, very large spectral efficiencies are required in multiple transmission modes. In such an environment, the motion of the user is very limited, as we can think of the user with a portable computer sitting on a desk communicating with the access point. Therefore, we shall regard such channels as low time diversity-channels with some degree of frequency selectivity. In particular, we will consider a classical indoor networks setting, namely, orthogonal frequency division multiplexing (OFDM) modulation. Under simple assumptions, OFDM reduces the frequency-selective channel to a set of non-interfering parallel channels in frequency, each of which, suffering a single (flat) fading attenuation. In such indoor OFDM environments it is common practice to assume that during the transmission of one packet the channel remains essentially constant. This is a very simple but interesting case of a channel that critically depends on *bad* fading values, since, when the fading is such that most of the parallel channels are severely attenuated, the transmitted packet may not be correctly recovered at the receiver.

Historically OFDM was introduced in [1], but it was not really considered an interesting option due to technological limitations until recently where it has been adopted as the main transmission strategy in many standards. In particular, it has been selected for digital audio and video broadcasting (DAB / DVB) [2, 3], for high-speed modem transmission using Asymmetric Digital Subscriber Line (ADSL) [4] and for the wireless Local Area Networks (LAN) standards HIPERLAN/2 [5] in Europe, IEEE802.11a in North America and ARIB HiSWAN in Japan. In this work, we shall study the HIPERLAN/2 application, even though the physical layers of these wireless LAN systems are compatible.

The above systems have motivated the introduction of the block-fading channel, first presented in [6], to model such small temporal fading variations. In particular, with this model, the transmitted packets span a fixed and usually small number of fading realizations. As we will see in Chapter 5, this simple model can be used to efficiently deal with the frequency variations in OFDM systems with reduced temporal variation as in wireless indoor networks. Moreover, the study of the block-fading channel is itself a subject of significant theoretical relevance, due to the non-ergodic nature of the channel since the transmitted codewords will only span a finite number of degrees of freedom. Strictly speaking the block-fading channel has zero capacity, meaning that, at a given fixed rate, we cannot make the error probability arbitrarily small. The relevant fundamental limits in such non-ergodic fading channels are then the outage capacity and the outage probability [6, 7].

This thesis studies coded modulation schemes in single user slowly varying block-fading channels with applications to high rate indoor wireless systems using orthogonal frequency division multiplexing (OFDM). In particular, we study the single-input single-output (SISO) channel that models slow frequency hopping or OFDM and the multiple-input multiple-output (MIMO) channel that models the transmission over multiple-transmit multiple-receive antenna channels.

The outline of the thesis is as follows:

- In Chapter 2 we first review some basic material on the block-fading channel. In particular, we describe the SISO and MIMO block-fading channels and we review its information-theoretic limits. In particular, we introduce the notions of outage probability, outage capacity and the diversity versus multiplexing trade-off. From this, we give some perspectives for code design, that will motivate the material presented in the rest of the thesis.

- Chapter 3 focuses on the SISO block-fading channel and studies its information theoretic limits with discrete signal constellations and discusses code design criteria. In particular, we rigurously prove that coded modulations over a finite signal set achieve the optimal rate-diversity trade-off given by the Singleton bound on the block-diversity of the code. Up to present, there has only been nuerical evidence of this fact. We introduce the novel family of blockwise concatenated codes and we analyze it by means of maximum-likelihood decoding error probability upper bounds and tight approximations. We show that blockwise concatenated codes achieve the optimal rate-diversity trade-off. Moreover, we show that belief propagation iterative decoding performs very close to maximum-likelihood on the block-fading channel, even for relatively short block lengths. We also show that the proposed codes perform close to the information outage probability for any block length. This is the dual threshold behavior of capacity achieving codes in ergodic channels in block-fading channels. This represents a major improvement to existing schemes based on trellis terminated convolutional codes, for which the error probability grows almost linearly with the block length. The material of this chapter is reported in part in [8, 9, 10, 11].

- Chapter 4 studies the impact of signal constellation expansion on the achievable diversity in MIMO block-fading channels by comparing two different signal constellation expansion methods: a) expansion of the complex QAM constellation; b) multidimensional expansion through lattice-based linear constellations. The role of the block diversity as the relevant design parameter when using iterative receivers is highlited. In particular, we compare such approaches under message passing iterative decoding. We study several low-complexity decoders and we focus on their rate versus decoding complexity trade-off in order to achieve the maximal diversity. We show that multidimensional signal constellation expansion that inherently achieves full diversity, can be advantageous over complex-plane expansion due to its inherent design flexibility. The material of this chapter is reported in part in [12, 13, 14].

- Chapter 5 is devoted to the application of the results of the previous chapters to slowly varying indoor channels that use OFDM. In particular, we compare the BICM NSTC approach described in Chapter 4 with other more classical approaches for space-time coding that use orthogonal and quasi-orthogonal precoders, full-rate linear rotations, and standard V-BLAST, over MIMO OFDM channels. We perform this comparison under the framework of HIPERLAN/2, and we will consider two channel environments: a typical office environment for NLOS propagation conditions with independent antenna channels, and a more realistic indoor channel generated by a ray tracing program. We show that, under both scenarios, with very simple and low-complexity iterative decoders, BICM NSTCs remarkably outperform all other approaches at the same spectral efficiencies. The material of this chapter is reported in part in [14, 15, 16].

Then, conclusions, future work perspectives and open problems are given in Chapter 6. We report some complementary material in the Appendices. Specifically, Appendix A, introduces a very simple and powerful method for accurately evaluating the error probability of bit-interleaved coded modulation based on a Gaussian approximation of the tail of the average log-likelihood ratio distribution. The material of this appendix is reported in part in [17, 18]. This method will be used in the maximum-likelihood analysis of the codes proposed in Chapter 3. Next, Appendix B reviews the computation of the average multivariate weight enumerator for the turbo-like code ensembles proposed in Chapter 3. It also provides the corresponding asymptotic extension and describes a way to compute the product weight enumerator of a given code. This will be useful again in Chapter 3 for the maximum-likelihood analysis of blockwise concatenated codes. In Appendix C we provide the proofs of some of the results presented in Chapter 3. Appendix D is devoted to the extension of the tangential-sphere bound for the block-fading channel. Finally, Appendix E introduces a density evolution algorithm for iterative interference cancellation decoder analysis in MIMO channels, which complements some of the results presented in Chapter 4. The material of this appendix is reported in part in [13].

2

The Block-Fading Channel: Background

This chapter reviews some background material on the block-fading channel. The block-fading channel was introduced in [6] (see also [7]) in order to model slowly-varying fading, where codewords span only a fixed number N_B of fading degrees of freedom, regardless of the code block length. This model is particularly relevant in wireless communication systems involving slow time-frequency hopping (e.g., GSM, EDGE) or multicarrier modulation using orthogonal frequency division multiplexing (OFDM). Despite its extreme simplification, it serves as a useful model to develop coding design criteria which turn out to be useful in more general settings of correlated slowly-varying fading. In particular, we consider single-output (SISO) and multiple-input multiple-output (MIMO) single user point-to-point block-fading channels. For both channels, we first review the system model, where the main assumptions used throughout the thesis are set. Then, we review their the information-theoretic limits and also give some perspectives for code design.

2.1 The SISO Block-Fading Channel

In this section we will review some basic material on the SISO block-fading channel used in this thesis. Of particular interest are the notions of diversity, outage probability and outage capacity. We will review these aspects and highlight the relevant code design criteria that will be subsequently used throughout the thesis.

2.1.1 System Model

We consider a single-user SISO block-fading channel with N_B fading blocks per codeword, where each block has L complex dimensions. A whole codeword spans then $N_B L$ complex dimensions. We assume that the fading is flat, constant over a block and i.i.d. from block to block and from one codeword to the next. Then, the discrete-time complex baseband equivalent channel is given by,

$$\mathbf{y}_b = \sqrt{\rho}\, h_b \mathbf{x}_b + \mathbf{z}_b, \quad b = 1, \ldots, N_B \tag{2.1}$$

where $\mathbf{y}_b \in \mathbb{C}^L$ denotes the received signal at block b, $\mathbf{x}_b \in \mathbb{C}^L$ is the portion of the transmitted signal assigned to block b, h_b represents the b-th *scalar* fading coefficient and $\mathbf{z}_b$ is the b-th vector of i.i.d. complex circularly symmetric Gaussian noise samples $\sim \mathcal{N}_{\mathbb{C}}(0,1)$. We assume that codewords are constructed over a signal constellation $\mathcal{X} \subseteq \mathbb{C}$. We will also assume that the signal constellation is normalized in energy, i.e., $\mathbb{E}[|x|^2] = 1$. Without any loss of generality, we also assume that the fading is also normalized such that $\mathbb{E}[|h_b|^2] = 1$. Then, the *average* signal-to-noise ratio (SNR) is given by ρ while the *instantaneous* SNR of block b is $|h_b|^2\rho$. For the sake of notation simplicity we will define $\gamma_b \triangleq |h_b|^2$ as the fading *power gain.* We can express channel 2.1 in a more compact form as,

$$\mathbf{Y} = \sqrt{\rho}\, \mathbf{H}\mathbf{X} + \mathbf{Z} \tag{2.2}$$

where $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_{N_B}]^T \in \mathbb{C}^{N_B \times L}$ and $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_{N_B}]^T \in \mathbb{C}^{N_B \times L}$ are obtained by rowwise stacking the corresponding block signals, $\mathbf{H} = \operatorname{diag}(\mathbf{h}) \in \mathbb{C}^{N_B \times N_B}$, with $\mathbf{h} = (h_1, \ldots, h_{N_B})$ and $\mathbf{Z} = [\mathbf{z}_1, \ldots, \mathbf{z}_{N_B}]^T \in \mathbb{C}^{N_B \times L}$.

Throughout this thesis, we shall assume that channel state information at the receiver (CSIR) is perfectly available, i.e., the fading coefficients of $\mathbf{h}$ have been perfectly estimated at the receiver and *ideal* coherent detection is possible. On the other hand, we assume that no channel state information at the transmitter (CSIT) is available.

The collection of all possible transmitted codewords $\mathbf{X}$ forms a coded modulation scheme $\mathcal{M}$ over $\mathcal{X}$ whose transmission rate (in bits per complex dimension) is given by,

$$R = \frac{1}{N_B L} \log_2 |\mathcal{M}| = \frac{K}{N_B L} \tag{2.3}$$

where $K = \log_2 |\mathcal{M}|$ is the number of information bits conveyed in one codeword assuming that all codewords are equally likely. Let $\phi : \{1, \ldots, |\mathcal{M}|\} \to \mathcal{X}^{N_B \times L}$ such that $\phi(\mathsf{m}) = \mathbf{X}$, and $\varphi : \mathbb{C}^{N_B \times L} \to \{1, \ldots, |\mathcal{M}|\}$ such that $\varphi(\mathbf{Y}) = \widehat{\mathsf{m}}$ be the encoding and decoding functions associated to the coded modulation $\mathcal{M}$, where m and $\widehat{\mathsf{m}}$ denote the transmitted message and its corresponding decoder output estimate. We define the error probability as $P_e(\rho) \triangleq \Pr(\widehat{\mathsf{m}} \neq \mathsf{m})$. The block diagram of the coding scheme described above is illustrated in Figure 2.1.

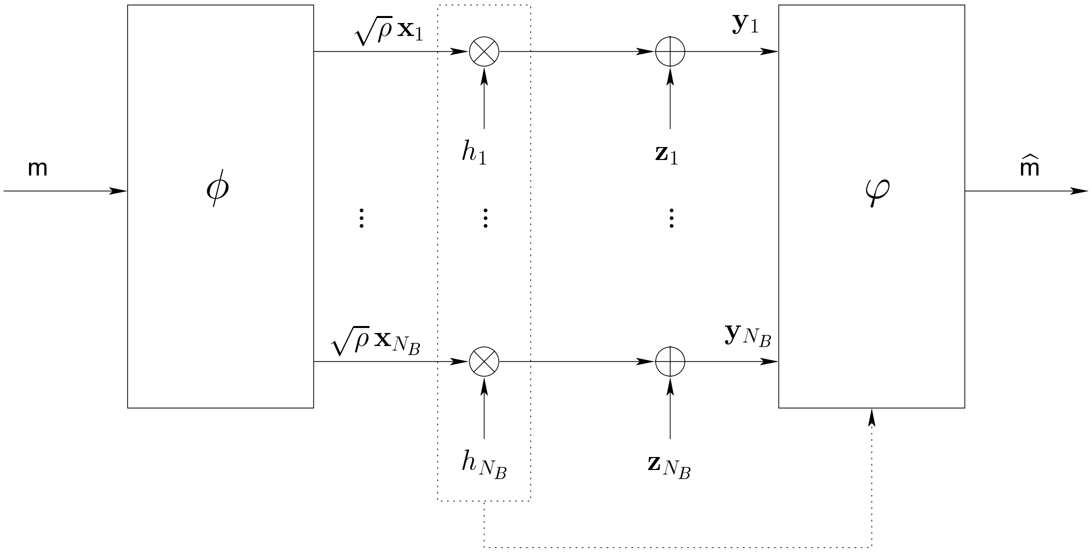


Figure 2.1: Block diagram of the SISO block-fading channel.

Conditioned on the channel parameters $\mathbf{H}$ and the transmitted codeword $\mathbf{X}$, the channel transition is given by

$$p(\mathbf{Y}|\mathbf{X},\mathbf{H}) = \frac{1}{\pi^{N_B L}}\, e^{-\|\mathbf{Y}-\sqrt{\rho}\mathbf{H}\mathbf{X}\|_F^2} \tag{2.4}$$

where $\|\mathbf{A}\|_F^2 = \mathrm{trace}(\mathbf{A}^H\mathbf{A}) = \sum_n \sum_m |a_{n,m}|^2$ is the squared Frobenius norm of the matrix $\mathbf{A}$. We say that φ represents the *optimal* maximum-likelihood (ML) decoding function if φ selects the decoder output message $\widehat{\mathsf{m}}$ according to,

$$\begin{aligned} \widehat{\mathsf{m}} &= \arg \max_{\mathsf{m}\in\{1,\dots,|\mathcal{M}|\}} p(\mathbf{Y}\,|\,\phi(\mathsf{m}),\mathbf{H}) \\ &= \arg \min_{\mathsf{m}\in\{1,\dots,|\mathcal{M}|\}} ||\mathbf{Y}-\sqrt{\rho}\,\mathbf{H}\,\phi(\mathsf{m})||_F^2 \ . \end{aligned} \tag{2.5}$$

Throughout this thesis, we consider a family of codes for which ML decoding is only possible through exhaustive enumeration of the candidate messages m. We shall then consider other decoding rules φ that approximate the ML performance with much less complexity.

2.1.2 Fundamental Limits

This section reviews well known results on the limiting performance of communication schemes in block-fading channels [6, 7, 19, 20]. In particular, we review the notions of *outage* and *diversity*.

Let $I(P_{\mathcal{X}}, \mathbf{H}) \triangleq \frac{1}{N_B L} I(\mathbf{X}; \mathbf{Y}|\mathbf{H})$ denote the mutual information (per complex dimension) between input and output, for a given realization of the fading coefficients $\mathbf{H}$ and $N_B L$-dimensional input probability assignment $P_{\mathcal{X}}$, satisfying the input power constraint $\mathbb{E}[|x|^2] = 1$. Since $\mathbf{H}$ is random, $I(P_{\mathcal{X}}, \mathbf{H})$ is in general a random variable that represents the *instantaneous* mutual information of the channel, whose cumulative distribution function is defined as $F_I(z) \triangleq \Pr(I(P_{\mathcal{X}}, \mathbf{H}) \leq z)$. Let also

$$\mathcal{A}(z) = \left\{ \mathbf{h} \in \mathbb{C}^{N_B} \ : \ I(P_{\mathcal{X}}, \mathbf{H}) \leq z \right\} \tag{2.6}$$

denote the set of fading states such that $I(P_{\mathcal{X}}, \mathbf{H}) \leq z$. Obviously, $F_I(z) = \Pr(\mathcal{A}(z))$. Then, the channel ϵ-capacity (as a function of the SNR ρ), or outage capacity, is given by [21],

$$C_\epsilon(\rho) = \sup_{P_{\mathcal{X}}} \ \sup \left\{ z \in \mathbb{R} \ : \ F_I(z) \leq \epsilon \right\} \tag{2.7}$$

and the Shannon channel capacity,

$$C(\rho) = \lim_{\epsilon \downarrow 0} C_\epsilon(\rho). \tag{2.8}$$

For fading distributions such that $P(|h| < \delta) > 0$, for any $\delta > 0$ (e.g., Rayleigh or Rice fading), we have $C(\rho) = 0$ for all $\rho \in \mathbb{R}_+$, implying that no positive rate is achievable. Therefore, it seems natural to take as the relevant performance measure over this block-fading channel the optimal error probability, or optimal *outage* probability, which is given by,

$$\epsilon(\rho) = \inf_{P_{\mathcal{X}}} F_I(R) \tag{2.9}$$

namely, for large block length L, it is the probability (minimized over the input distributions) that the transmission rate R is not supported by that channel realization. In other words, it represents the probability that the channel parameters belong to the *outage region* $\mathcal{A}(R)$.

Notice that for short block length L it may be possible to find codes with error probability smaller than $\epsilon(\rho)$ given in (2.9). However, in the limit of large L and fixed coding rate R, no code has error probability smaller than $\epsilon(\rho)$. A lower bound to the word error rate (WER) of any code for any finite length L is provided by Fano Inequality and reads [22],

$$P_e(\rho) \geq \inf_{P_{\mathcal{X}}} \mathbb{E}\left[\max\left\{ 1 - \frac{1}{R} I(P_{\mathcal{X}}, \mathbf{H}) - \frac{1}{R N_B L}, 0 \right\} \right] \tag{2.10}$$

that converges to $\epsilon(\rho)$ as $L \to \infty$. In many cases, the input distribution is fixed by some system constraint. Hence, it is customary to define the *information outage probability* [7, 6] as $P_{\text{out}}(\rho, R) \triangleq F_I(R)$ for given $P_{\mathcal{X}}$, ρ and R. The goodness of a coding scheme for the block-fading channel is then measured by its ability to approach the outage probability for large block length L.

Since the channel model given in (2.1) corresponds to a set of N_B non-interfering parallel channels, each of which is used a fraction $\frac{1}{N_B}$ of time, we can express the instantaneous mutual information as,

$$I(P_{\mathcal{X}}, \mathbf{H}) = \frac{1}{N_B} \sum_{b=1}^{N_B} J_{\mathcal{X}}(\rho\gamma_b) \tag{2.11}$$

where $J_{\mathcal{X}}(\rho\gamma_b)$ is the mutual information of a standard additive white Gaussian noise (AWGN) channel with input distribution $P_{\mathcal{X}}$ over $\mathcal{X}$ with SNR $\rho\gamma_b$. Therefore, when the entries of the codewords of $\mathcal{M}$ are $\sim \mathcal{N}_{\mathbb{C}}(0,1)$ we have that,

$$I(P_{\mathcal{X}}, \mathbf{H}) = \frac{1}{N_B} \sum_{b=1}^{N_B} \log_2(1 + \rho\gamma_b). \tag{2.12}$$

Figure 2.2 shows the outage probability as a function of the transmission rate R, for a SISO block-fading channel with Gaussian inputs with $N_B = 2, 4, 16$ and 100 fading blocks per codeword under Rayleigh fading, i.e., $h_b \sim \mathcal{N}_{\mathbb{C}}(0,1)$. The SNR is $\rho = R\frac{E_b}{N_0}$, where $\frac{E_b}{N_0}$ is the signal-to-noise ratio per information bit. We fix $\frac{E_b}{N_0} = 10$ dB. We clearly observe that there is no rate with error probability strictly zero, but rather, the more blocks we have, the more ergodic the channel looks, and the steeper the outage curve is. Indeed, for the ergodic channel, i.e., $N_B \to \infty$, the outage probability is a step function: for rates below the channel capacity the outage probability is strictly zero, and for rates above the channel capacity, the outage probability is one.

Let now define the *SNR reliability function* $d_B^\star$ as the maximum achievable SNR exponent of error probability for codes in a given family of interest, namely,

$$d_B^\star \triangleq \sup_{\mathcal{M} \in \mathcal{F}} \lim_{\rho \to \infty} -\frac{\log P_e(\rho, \mathcal{M})}{\log \rho} \tag{2.13}$$

where ρ denotes the channel SNR, $P_e(\rho, \mathcal{M})$ is the error probability of a given coded modulation scheme $\mathcal{M}$, and the supremum is taken over all coded modulation schemes in the family $\mathcal{F}$. In other words, $d_B^\star \in \{1, 2, \ldots, N_B\}$ is the optimal asymptotic slope of the error probability with SNR. This is commonly referred to as the optimal *diversity* gain of the coded modulation family $\mathcal{F}$. We will then be interested in characterizing the diversity as a function of the rate, i.e., the function $d_B^\star(R)$. We also define $d_B^{(r)}(R)$ as the random coding SNR exponent, i.e., the optimal asymptotic slope of the error probability curve of random codes.

In the case of independent Rayleigh fading, the fading power gains γ_b are Chi-squared with two degrees of freedom, i.e., $\gamma_b \sim f_\gamma(z) = e^{-z} \mathbb{1}\{z \geq 0\}$, where $\mathbb{1}\{\mathcal{E}\}$ denotes the indicator function of the event $\mathcal{E}$, then, the following result can be obtained from [22],

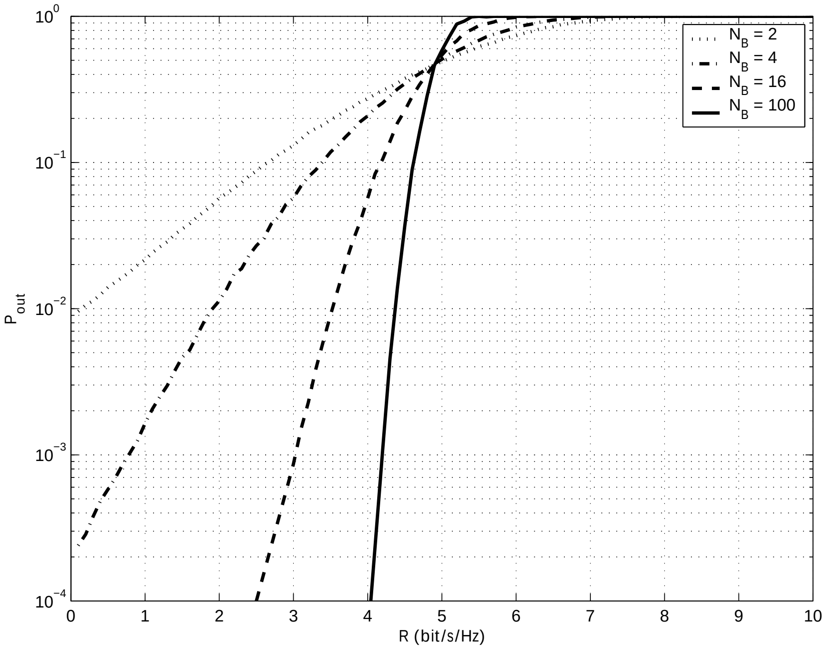


Figure 2.2: Outage probability with Gaussian inputs for a SISO block-fading channel with $N_B = 2, 4, 16$ and 100 fading blocks per codeword as a function of R for $\frac{E_b}{N_0} = 10$ dB.

Lemma 1 *Consider the block-fading channel* (2.1) *with i.i.d. Rayleigh fading, under the average input power constraint* $\frac{1}{N_B L}\sum_{b=1}^{N_B}\mathbb{E}[|\mathbf{x}_b|^2] \leq 1$. *The SNR reliability function for any block length* $L \geq 1$ *and fixed rate* R *is given by* $d_B^\star(R) = N_B$, *and it is achieved by Gaussian random codes, i.e., the random coding SNR exponent* $d_B^{(r)}(R)$ *of the Gaussian i.i.d. ensemble for any* $L \geq 1$ *is also equal to* N_B.

Proof. Although Lemma 1 follows as a corollary of [22, Th. 2], we provide an explicit proof in Appendix C for the sake of completeness and because it is instructive to illustrate the proof technique used for the results presented in Chapter 3. The proof of Lemma 1 deals with the more general case of coding schemes with rate increasing with SNR as $R(\rho) = r\log\rho$, where $r \in [0,1]$, and shows that[1] $P_{\text{out}}^G(\rho, r\log\rho) \doteq \rho^{-N_B(1-r)}$ where the optimal SNR exponent $N_B(1-r)$ can be achieved by coding schemes of any block length $L \geq 1$. The details are given in Appendix C.1. $\square$

The assumption of Rayleigh fading can be relaxed by noticing that in the proof of Lemma 1 only the near-zero behavior of the fading power gain distribution is relevant. For Rayleigh fading, we have $\Pr(\gamma_b \leq \delta) \approx \delta$, for small $\delta > 0$. Hence, the above results hold for all block-fading channels with i.i.d. fading with power gain distribution with this behavior. More in general, as argued in [22], for a fading distribution with near-zero behavior $\Pr(\gamma_b \leq \delta) \approx \delta^D$, the SNR reliability function is given by $D\delta_B(R)$. For example, this is the case of independent Rayleigh fading with a D antenna receiver using D-fold maximal-ratio combining [23].

---

[1] The exponential equality and inequalities notation $\doteq$, $\dot{\geq}$ and $\dot{\leq}$ were introduced in [22]. We write $f(z) \doteq z^d$ to indicate that that $\lim_{z\to\infty}\frac{\log f(z)}{\log z} = d$. $\dot{\geq}$ and $\dot{\leq}$ are used similarly.

Figures 2.3 and 2.4 clearly illustrate the operational meaning of Lemma 1. We first notice, that for a fixed rate R, as the number of blocks increases, the outage probability curve gets steeper. Again, as $N_B \to \infty$ it will converge to the step function indicating the capacity threshold, i.e., channel parameter above which, we can transmit as reliably as desired. Notice that, Lemma 1 states that Gaussian codes achieve $d^\star_B(R) = N_B$, which is indeed independent of R. This is clearly seen in Figure 2.4, where for different rates, the outage probability curves are all parallel with the same asymptotic slope. While this is true for Gaussian inputs, as we shall see in Chapter 3 this will not be the case with coded modulation schemes constructed over a discrete signal constellation. In fact, in this case, $d^\star_B(R)$ will depend on R, and specifically, for a fixed constellation, the higher R is, the lower $d^\star_B(R)$ will be. This will induce the concept of signal constellation *expansion*, as we shall see further on.

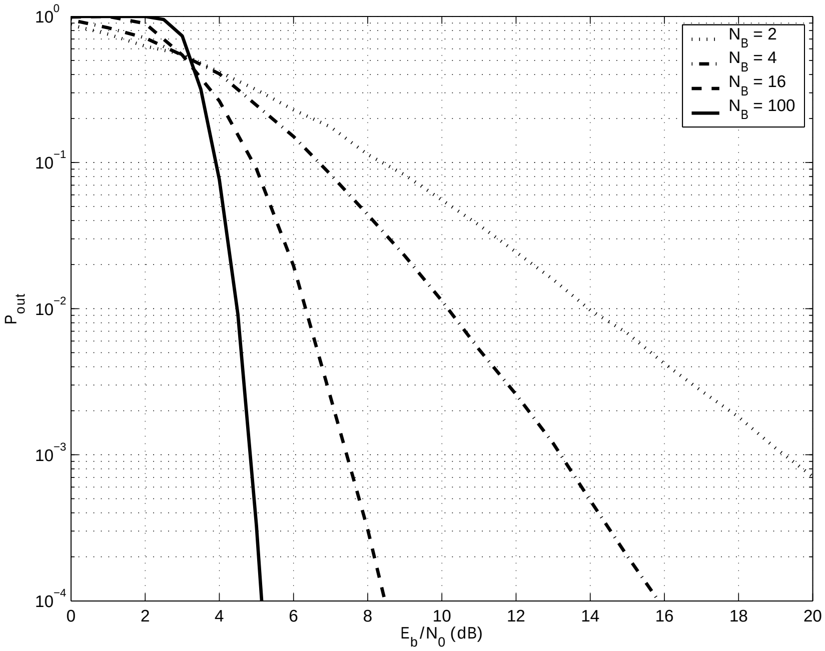


Figure 2.3: Outage probability with Gaussian inputs for a SISO block-fading channel with $N_B = 2, 4, 16$ and 100 fading blocks per codeword as a function of SNR for $R = 2$ bit/s/Hz.

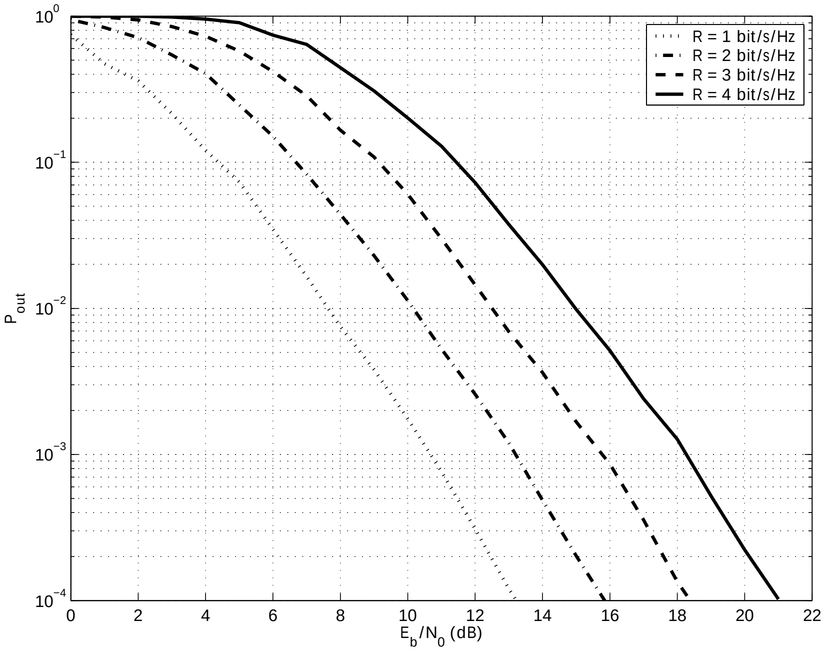


Figure 2.4: Outage probability with Gaussian inputs for a SISO block-fading channel with $N_B = 4$ fading blocks per codeword as a function of SNR for $R = 1, 2, 3$ and 4 bit/s/Hz.

We shall then compare different coded modulation schemes by their corresponding SNR gap to the outage probability. We say that a coded modulation scheme $\mathcal{M}$ is *good* if, as L gets large, its error probability approaches the outage probability limit. Similarly, a coded modulation is said to be *weakly good* if its error probability, as L gets large, becomes independent of L and thus, shows a fixed SNR gap to the outage probability as the slope becomes the same.

2.1.3 Code Design

Coding for the block-fading channel has been considered in a number of recent works (e.g., [24, 19, 25, 26] and references therein). The design criteria for codes over the block-fading channel differ significantly with respect to the standard design criteria for codes over the AWGN channel or over the fully-interleaved fading channel. The key difference, as remarked in the previous sections is that the block-fading channel is not information stable [21, 20].

We will be interested in constructing efficient coded modulation schemes over discrete signal constellations $\mathcal{X}$ of cardinality $|\mathcal{X}| = 2^M$, where M is the number of bits that can be conveyed per constellation symbol. Let us now evaluate the probability of error of such codes. Conditioned on the fading states, and under

ML decoding, the union bound on the error probability is given by,

$$P_e(\rho|\mathbf{H}) \leq \sum_{\mathbf{X}} \sum_{\mathbf{X}' \neq \mathbf{X}} P(\mathbf{X}) P(\mathbf{X} \to \mathbf{X}'|\mathbf{H}) \tag{2.14}$$

where $P(\mathbf{X} \to \mathbf{X}'|\mathbf{H})$ is the conditional pairwise error probability, the probability of deciding in favor of $\mathbf{X}'$ when $\mathbf{X}$ has been transmitted, conditioned on the fading realization. Then, the average error probability is,

$$\begin{aligned} P_e(\rho) &= \mathbb{E}[P_e(\rho|\mathbf{H})] \\ &\leq \sum_{\mathbf{X}} \sum_{\mathbf{X}' \neq \mathbf{X}} P(\mathbf{X}) P(\mathbf{X} \to \mathbf{X}') \end{aligned} \tag{2.15}$$

where $P(\mathbf{X} \to \mathbf{X}') \triangleq \mathbb{E}[P(\mathbf{X} \to \mathbf{X}'|\mathbf{H})]$ denotes the average (over the fading states) pairwise error probability. Consider for example that $\mathcal{M}$ is binary, i.e., it is obtained by simply concatenating a binary code $\mathcal{C} \in \mathbb{F}_2^N$ with BPSK modulation. Then, the pairwise error probability can be written as,

$$\begin{aligned} P(\mathbf{X} \to \mathbf{X}') &= \mathbb{E}[P(\mathbf{X} \to \mathbf{X}'|\mathbf{H})] && (2.16) \\ &= \mathbb{E}\left[Q\left(\sqrt{\frac{\rho}{2}\delta}\right)\right] && (2.17) \end{aligned}$$

where

$$Q(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-\frac{t^2}{2}} dt \tag{2.18}$$

is the Gaussian tail function and $\delta = 4\sum_{b=1}^{N_B} w_b \gamma_b$ is the squared Euclidean distance corresponding to a pairwise error event that associates Hamming weight w_b to the portion of the codeword assigned to block b. Then, (2.17) can be upperbounded by,

$$\begin{aligned} P(\mathbf{X} \to \mathbf{X}') &\leq \frac{1}{2} \mathbb{E}\left[e^{-\rho \sum_{b=1}^{N_B} w_b \gamma_b}\right] && (2.19) \\ &\leq \frac{1}{2} \prod_{b=1}^{N_B} \frac{1}{1 + \rho w_b}. && (2.20) \end{aligned}$$

We define the *block diversity* of a binary code $\mathcal{C}$ mapped onto N_B blocks as the blockwise Hamming distance,

$$d_B \triangleq \min_{\mathbf{X} \in \mathcal{M}} |\{b \in [1, \ldots, N_B] \ : \ w_b \neq 0\}|. \tag{2.21}$$

Then, as apparent from (2.20), at high SNR the slope of $P_e(\rho)$ is dominated by d_B, i.e. the number of nonzero rows of $\mathbf{X} - \mathbf{X}'$. If we now look at $\mathcal{M}$ as a code of length N_B over the alphabet $\mathcal{X}^L$ with cardinality $|\mathcal{X}|^L$, d_B represents nothing but the minimum distance of the code under this representation. Then, any bound on the minimum distance applies here. In particular, the Singleton Bound yields that [24, 19],

$$d_B(R) \leq d_{SB} \triangleq 1 + \left\lfloor N_B \left(1 - \frac{R}{M}\right)\right\rfloor \tag{2.22}$$

where R is defined in (2.3) and $\lfloor x \rfloor$ denotes the largest integer smaller than or equal to x. Then, (2.20) suggests that one should search for codes attaining the Singleton bound with equality, namely, maximum distance separable (MDS) codes. In [27, Th. 1], it is shown that, for binary codes, the Singleton bound is tight for any $R \in (0, 1]$. The achievability proof in [27, Th. 1] is based on the existence of maximum distance separable (MDS) codes over $\mathbb{F}_{2^L}$ (e.g., Reed-Solomon codes).

However, it is well known that, as opposed to ergodic channels, the union bound can be very loose in the block-fading channel, and especially for small N_B, it may not even converge in the SNR region of practical interest [24, 19, 25, 26]. Therefore, it is natural to ask whether or not it is possible to do better by looking at a more meaningful criteria. This question is answered in Chapter 3, where, by analyzing the asymptotic behavior of the outage probability with discrete signal sets, we prove that the optimal rate-diversity trade-off $d_B^\star(R)$ is indeed given by the Singleton bound. Numerical examples in [24, 19] showed evidence of this fact, but there was no formal proof of this result so far.

2.2 The MIMO Block-Fading Channel

Multiple-antenna or MIMO channels model transmission systems where either the transmitter, the receiver or both have multiple antennas available for transmission / reception (see Figure 2.5). Multiple antenna transmission has emerged as a key technology to achieve high spectral and power efficient communication in fading channels ever since the landmark works by Telatar [28] and by Foschini and Gans [29], which illustrate huge advantages from an information-theoretic perspective. Coded modulation schemes that are able to exploit the degrees of freedom in both space and time/frequency were named *space-time codes* after [30]. Since [30] and [31], multiple antennas have been used to enhance transmission robustness by exploiting the available degrees of freedom to increase diversity, i.e., the number of channels across which the transmitted symbols go through. This is somewhat in contrast with the ideas exposed in [28, 29], for which multiple antennas are used to communicate higher rates. In [22], the authors link both concepts and pose a fundamental trade-off between the achievable diversity and the increase in transmission rate, named as multiplexing gain, in multiple antenna systems. This is usually referred to as the *diversity versus multiplexing* trade-off. In this section we will review some basic material on the MIMO block-fading channel.

In particular, we will briefly review the extension of the outage and diversity concepts, and will present the aforementioned results of [22].

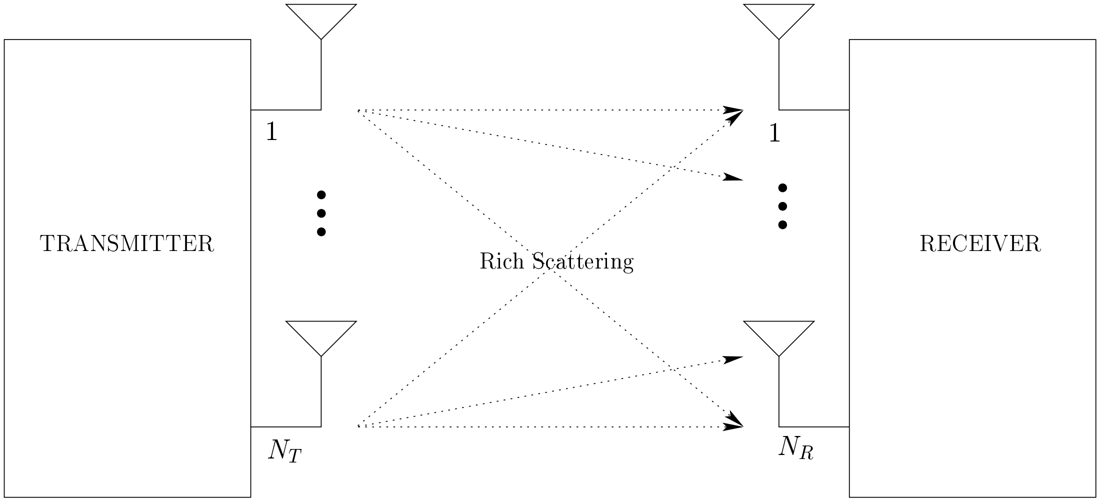


Figure 2.5: Multiple transmit and multiple receive antenna environment.

2.2.1 System Model

We consider a MIMO block-fading channel with N_T transmit and N_R receive antennas respectively and N_B fading blocks per codeword. Again, a codeword spans N_BL complex dimensions. We assume that the fading coefficients are flat, constant during a block and i.i.d., from block to block and from one codeword to the next. Then, the discrete-time complex baseband equivalent channel is given by,

$$\mathbf{Y}_b = \sqrt{\rho}\,\mathbf{H}_b\mathbf{X}_b + \mathbf{Z}_b \quad b = 1, \dots, N_B \tag{2.23}$$

where $\mathbf{Y}_b \in \mathbb{C}^{N_R \times L}$ is the received signal matrix corresponding to the b-th block, $\mathbf{H}_b \in \mathbb{C}^{N_R \times N_T}$ is the fading channel matrix corresponding to block b with i.i.d. entries $\sim \mathcal{N}_{\mathbb{C}}(0,1)$, i.e, Rayleigh fading, $\mathbf{X}_b \in \mathbb{C}^{N_T \times L}$ is the b-th block of the transmitted codeword, $\mathbf{Z}_b \in \mathbb{C}^{N_R \times L}$ is the corresponding noise matrix with i.i.d. samples $\sim \mathcal{N}_{\mathbb{C}}(0,1)$. We assume that the symbols transmitted from a given antenna have unit energy, i.e., $\mathbb{E}[|x|^2] = 1$. The channel is normalized such that $\frac{1}{N_T N_R}\text{trace}(\mathbb{E}[\mathbf{H}_b\mathbf{H}_b^H]) = 1$, for $1 \le b \le N_B$, so the average SNR per receive antenna is $N_T\rho$ and the average SNR per transmit antenna is ρ. The simple case of quasistatic fading is obtained by letting $N_B = 1$ in (2.23). Again, we assume perfect CSIR and no CSIT. Notice that now, we can again express (2.23) using

(2.2), where now, rowwise stacking the signals corresponding to each block, yields

$$\mathbf{X}=\begin{bmatrix}\mathbf{X}_1\\ \vdots\\ \mathbf{X}_{N_B}\end{bmatrix}\in\mathbb{C}^{N_B N_T\times L},\quad \mathbf{Y}=\begin{bmatrix}\mathbf{Y}_1\\ \vdots\\ \mathbf{Y}_{N_B}\end{bmatrix}\in\mathbb{C}^{N_B N_R\times L},\quad \mathbf{Z}=\begin{bmatrix}\mathbf{Z}_1\\ \vdots\\ \mathbf{Z}_{N_B}\end{bmatrix}\in\mathbb{C}^{N_B N_R\times L} \tag{2.24}$$

and

$$\mathbf{H}=\mathrm{diag}(\mathbf{H}_1,\dots,\mathbf{H}_{N_B})=\begin{bmatrix}\mathbf{H}_1 & \dots & \mathbf{0}\\ \vdots & \ddots & \vdots\\ \mathbf{0} & \dots & \mathbf{H}_{N_B}\end{bmatrix}\in\mathbb{C}^{N_B N_R\times N_B N_T} \tag{2.25}$$

is the block diagonal channel matrix.

Similarly to the SISO case, we say that the collection of all possible transmitted codewords $\mathbf{X}$ forms a *space-time coding* scheme $\mathcal{S}$ over $\mathcal{X}\subseteq\mathbb{C}$ whose transmission rate (in bits per complex dimension) is given by,

$$R=\frac{1}{N_B L}\log_2|\mathcal{S}|=\frac{K}{N_B L} \tag{2.26}$$

where $K=\log_2|\mathcal{S}|$ is the number of information bits conveyed in one codeword. Let $\phi:\{1,\dots,|\mathcal{S}|\}\to\mathcal{X}^{N_B N_T\times L}$ such that $\phi(\mathsf{m})=\mathbf{X}$, and $\varphi:\mathbb{C}^{N_B N_R\times L}\to\{1,\dots,|\mathcal{S}|\}$ such that $\varphi(\mathbf{Y})=\widehat{\mathsf{m}}$ be the encoding and decoding functions associated to the space-time code $\mathcal{S}$, where m and $\widehat{\mathsf{m}}$ denote again the information message and its corresponding decoder output estimate. The error probability is $P_e(\rho)\triangleq\Pr(\widehat{\mathsf{m}}\neq\mathsf{m})$. The corresponding block diagram is shown in Figure 2.6. The ML decoder is again given by (2.5), where now, $\mathbf{X},\mathbf{Y},\mathbf{Z}$ and $\mathbf{H}$ are given in (2.24) and (2.25) respectively.

2.2.2 Fundamental Limits

In the MIMO block-fading channel, the outage probability is a relevant fundamental limit. Strictly speaking MIMO block-fading channels have Shannon capacity zero, due to its non-ergodic nature, and all the discussion in the previous section is valid also here. However, there are some differences that we shall tackle in this section.

Notice that the mutual information $I(P_{\mathcal{X}},\rho)$, can be again expressed as (2.11), since the channel (2.23) represents a set of parallel MIMO channels each of them used a fraction $\frac{1}{N_B}$ of the time. When the entries of the codewords of $\mathcal{S}$ are i.i.d. $\sim\mathcal{N}_{\mathbb{C}}(0,1)$, we have that,

$$I(P_{\mathcal{X}},\rho)=\frac{1}{N_B}\sum_{b=1}^{N_B}\log_2\det(\mathbf{I}_{N_R}+\mathbf{H}_b\mathbf{Q}\mathbf{H}_b^H) \tag{2.27}$$

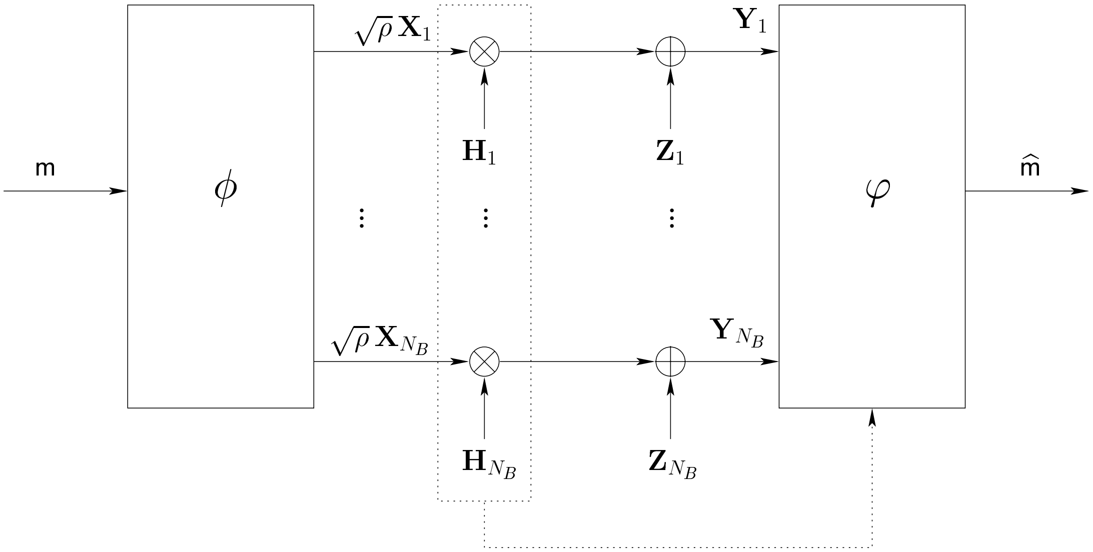


Figure 2.6: Block diagram of the MIMO block-fading channel model (2.23).

where $\mathbf{Q} = \text{diag}(\mathbf{q})$ is the input covariance. At large SNR, $\mathbf{Q} = \rho\mathbf{I}_{N_T}$ is optimal. However, at low SNR the optimal input covariance matrix is still an open problem. Recent results seem to indicate that the optimal input covariance, i.e., the covariance matrix that minimizes the outage probability, is in the line with that conjectured by Telatar in [28], namely, it uniformly distributes power across a set of transmit antennas whose size depends on SNR and R. However, no formal proofs are available and it is still an open problem in general. Thus, we shall consider the simple case of $\mathbf{Q} = \rho\mathbf{I}_{N_T}$ for all the SNR and rate ranges. Since in this work we are mostly concerned by achieving diversity, this simple covariance matrix is optimal.

Figure 2.7 shows the outage probability for a MIMO block-fading channel with Gaussian inputs with $N_T = 2$ transmit antennas, $N_R = 2$ receive antennas, $N_B = 1, 2$ and 16 fading blocks per codeword, with $R = 2$ bit/s/Hz. We observe that again, as the number of blocks increases, the outage curve becomes steeper, and eventually converge for large N_B to the step function of the channel capacity threshold.

Let us now examine the diversity vs. multiplexing trade-off of [22]. The capacity of the ergodic MIMO channel, i.e., a MIMO channel for which the channel matrix changes at every time instant, is given by [28],

$$C(\rho) = \mathbb{E}[\log_2 \det(\mathbf{I}_{N_R} + \rho\widetilde{\mathbf{H}}\widetilde{\mathbf{H}}^H)] \tag{2.28}$$

where $\widetilde{\mathbf{H}} \in \mathbb{C}^{N_R \times N_T}$ has i.i.d. $\sim \mathcal{N}_{\mathbb{C}}(0, 1)$ entries. For large enough SNR, (2.28) behaves like $\min\{N_T, N_R\} \log_2(\rho)$, in contrast to the classical $\log_2(\rho)$ behavior of single-antenna systems. We can then communicate at a rate up to $\min\{N_T, N_R\}$

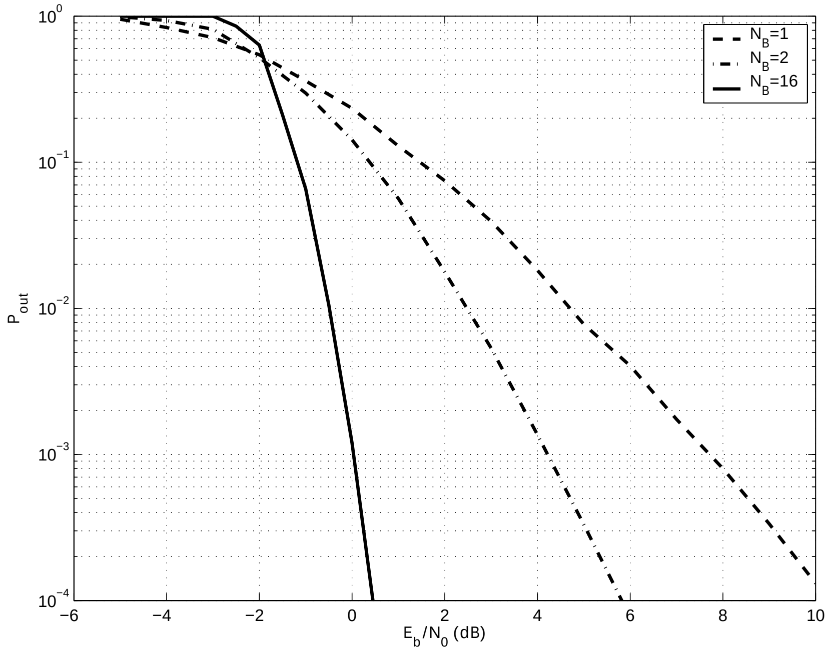


Figure 2.7: Outage probability with Gaussian inputs for a MIMO block-fading channel with $N_T = 2$ transmit antennas, $N_R = 2$ receive antennas, $N_B = 1, 2$ and 16 fading blocks per codeword as a function of SNR for $R = 2$ bit/s/Hz.

times larger than in the SISO case. This rate scaling is usually referred to as the *multiplexing gain*, and represents the number of parallel channels between transmitter and receiver. In [22] the authors considered a family of space-time codes $\{\mathcal{S}(\rho)\}$ indexed by their operating SNR ρ, such that the code $\mathcal{S}(\rho)$ transmits at a rate $R(\rho)$ and has error probability $P_e(\rho, \mathcal{S}(\rho))$. Notice that for such codes, the rate *does* depend on SNR. For such family of space-time codes, the multiplexing gain is defined as,

$$r \triangleq \lim_{\rho\to\infty} \frac{R(\rho)}{\log(\rho)} \tag{2.29}$$

while the diversity gain is defined as,

$$d \triangleq \lim_{\rho\to\infty} -\frac{\log P_e(\rho, \mathcal{S}(\rho))}{\log \rho}. \tag{2.30}$$

In [22], the authors determined that the optimal trade-off $d^\star(r)$ is given by the piecewise linear function,

$$d^\star(r) = N_B(N_T - r)(N_R - r) \tag{2.31}$$

where r is an integer between 0 and $\min\{N_T, N_R\}$. Qualitatively, the diversity vs. multiplexing trade-off says that, both diversity and multiplexing gains can be achieved simultaneously, but that, the more we *invest* in maximizing one of them, will come at the expense of reducing the other. This is clearly seen in Figure 2.8 where we show the optimal diversity vs. multiplexing trade-off in MIMO block-fading channels with $N_B = 1, 2$ and $N_T = 4$, $N_R = 4$.

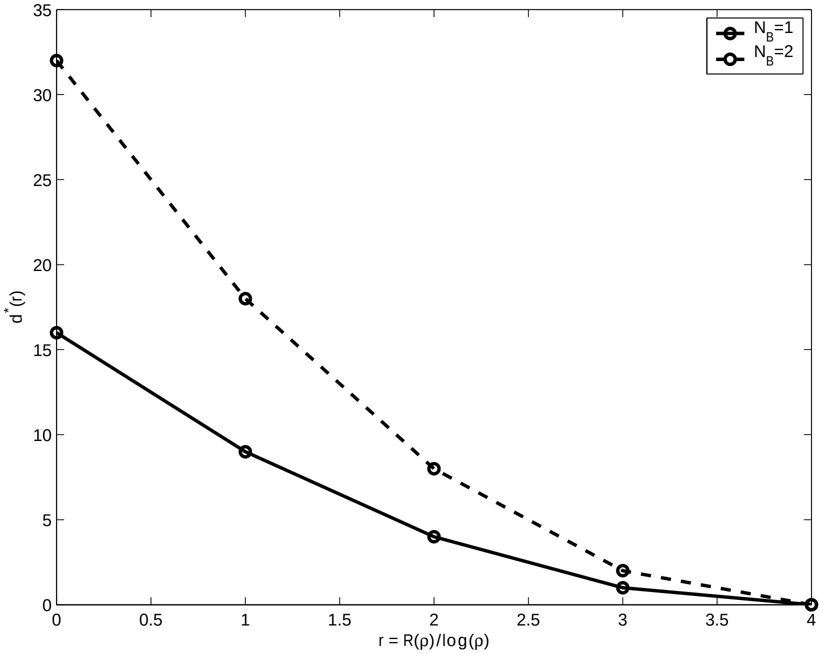


Figure 2.8: Optimal diversity vs. multiplexing trade-off in MIMO block-fading channels with $N_T = 4$, $N_R = 4$ and $N_B = 1, 2$.

For practical space-time coding schemes $\mathcal{S}$, we have the following [32]:

Proposition 1 *Let $\mathcal{S}(R)$ be a space-time coding scheme that supports an arbitrary rate R, and let $P_e(\rho, \mathcal{S}(R))$ denote its error probability. Then, $\mathcal{S}$ achieves the optimal diversity versus multiplexing trade-off if*

$$P_e(\rho, \mathcal{S}(R)) = \alpha P_{\text{out}}(\rho, R), \quad \forall R, \rho \in \mathbb{R}_+ \tag{2.32}$$

where α is a constant such that $\lim_{\rho\to\infty} \alpha / \log \rho = 0$.

Proposition 1 implies that the SNR *gap* between the actual performance of the optimal space-time coding scheme $\mathcal{S}(R)$ and the outage probability should be *independent* of the transmission rate and the SNR. This is an empirical way of verifying if the optimal trade-off is achieved. Notice that in Proposition 1 the constant α is arbitrary, i.e., it does not capture the coding gain of $\mathcal{S}(R)$.

2.2.3 Code Design

Code design for the MIMO block-fading channel has been studied in several works. Tarokh *et al.* in [30] and Guey *et al.* in [33], derived design criteria for the quasistatic fading case. In a more recent work [34], El Gamal and Hammons, generalized the previous design to the general block-fading channel. These works, analyzed the worst case pairwise error probability using ML decoding. As we have seen, the error probability can be upperbounded by,

$$P_e(\rho) \leq \sum_{\mathbf{X}} \sum_{\mathbf{X}' \neq \mathbf{X}} P(\mathbf{X}) P(\mathbf{X} \to \mathbf{X}') \tag{2.33}$$

where $P(\mathbf{X} \to \mathbf{X}') = \mathbb{E}[P(\mathbf{X} \to \mathbf{X}'|\mathbf{H})]$ is the average (over the channel states) pairwise error probability. The conditional pairwise error probability for the MIMO Rayleigh case can be upperbounded as as,

$$\begin{aligned} P(\mathbf{X} \to \mathbf{X}'|\mathbf{H}) &= Q\left(\sqrt{\frac{\rho \|\mathbf{H}\mathbf{D}\|_F^2}{2}}\right) && (2.34) \\ &\leq \prod_{b=1}^{N_B} \exp\left(-\frac{\rho}{4}\,\mathrm{trace}(\mathbf{H}_b \mathbf{D}_b \mathbf{D}_b^H \mathbf{H}_b^H)\right) && (2.35) \end{aligned}$$

where

$$\mathbf{D} = \begin{bmatrix} \mathbf{D}_1 \\ \vdots \\ \mathbf{D}_{N_B} \end{bmatrix} \triangleq \mathbf{X} - \mathbf{X}' = \begin{bmatrix} \mathbf{X}_1 - \mathbf{X}'_1 \\ \vdots \\ \mathbf{X}_{N_B} - \mathbf{X}'_{N_B} \end{bmatrix} \tag{2.36}$$

is the codeword difference matrix. Then, using the decomposition $\mathbf{D}_b \mathbf{D}_b^H = \mathbf{U}_b \mathbf{\Lambda}_b \mathbf{U}_b^H$ and the fact that $\widetilde{\mathbf{H}}_b = \mathbf{H}_b \mathbf{U}_b \sim \mathbf{H}_b$,

$$\begin{aligned} P(\mathbf{X} \to \mathbf{X}') &\leq \mathbb{E}\left[\prod_{b=1}^{N_B} \prod_{t=1}^{N_T} \exp\left(-\frac{\rho}{4}\left(\sum_{r=1}^{N_R} |\widetilde{h}_{b,r,t}|^2\right)\lambda_{b,t}\right)\right] && (2.37) \\ &= \prod_{b=1}^{N_B} \prod_{t=1}^{N_T} \left(\frac{1}{1 + \frac{\rho}{4}\lambda_{b,t}}\right)^{N_R} && (2.38) \end{aligned}$$

where $\lambda_{b,t}$ is the t-th eigenvalue of $\mathbf{D}_b \mathbf{D}_b^H$. Then, as ρ gets large, it is clear that the SNR exponent of (2.38) is dominated by the rank of $\mathbf{D}_b \mathbf{D}_b^H$, i.e., number of nonzero eigenvalues of $\mathbf{D}_b \mathbf{D}_b^H$. Therefore, for large SNR, the error probability behaves as

$$P(\mathbf{X} \to \mathbf{X}') \leq G_c \rho^{d_R N_R} \tag{2.39}$$

where d_R is the *rank diversity* of $\mathcal{S}$ defined as,

$$d_R \triangleq \min_{\mathbf{X}, \mathbf{X}' \in \mathcal{S}} \mathrm{rank}(\mathbf{X} - \mathbf{X}') \tag{2.40}$$

and

$$G_c \triangleq \prod_{\substack{b,t \\ \lambda_{b,t} > 0}} (4\lambda_{b,t})^{-N_R} \tag{2.41}$$

gives the asymptotic coding gain. Thus, we would search for code constructions that maximize both the rank diversity and the product of eigenvalues, which give the asymptotic slope and coding gain of the pairwise error probability. These are the so-called *rank diversity* and the *eigenvalue product* criteria. The difficulty in constructing codes satisfying the above criteria is that the rank diversity is usually hard to evaluate and it is barely related to the algebraic properties of the code, with the exception of binary codes over BPSK or QPSK [35, 34].

A simple upper bound on the rank diversity is given by the block diversity described before. This argument comes from the assumption of a genie-aided decoder that produces observables of the transmitted symbols of one antenna, assuming that the symbols from all other antennas are known. Under this assumption, the channel (2.23) decomposes into a set of $N_T N_B$ parallel channels, i.e., the problem of code design over MIMO channels reduces to the SISO case, where the Singleton bound again plays a key role. We shall study this in more detail in Chapter 4, where we will show that very simple coding strategies using this criterion can achieve full-diversity with no explicit verification of the rank criterion. Notice that such a pragmatic design strongly depends on the ability to design *good* receivers, i.e., receivers that are able to perform close to the genie-aided decoder. As we shall see also in Chapter 4, this is possible by using low-complexity iterative strategies. The main focus of Chapter 4 will be on constructing full-diversity space-time codes, and specifically, on the role of signal constellation expansion on the achievable diversity in MIMO channels.

3

Coded Modulation in the SISO Block-Fading Channel

In this chapter we consider coded modulation schemes for the SISO block-fading channel. We show that coded modulations of rate R bit/complex dimension, over a finite signal set $\mathcal{X} \subset \mathbb{C}$ of size 2^M, achieve the optimal rate-diversity trade-off given by the Singleton bound for $R \in (0, M]$. Furthermore, we show also that bit-interleaved coded modulation achieves the same optimal rate-diversity trade-off. We present a novel coded modulation construction based on blockwise concatenation that systematically yields Singleton-bound achieving turbo-like codes defined over an arbitrary signal set $\mathcal{X} \subset \mathbb{C}$. The proposed blockwise concatenation significantly outperforms conventional serial and parallel turbo codes in the block-fading channel. We analyze the *ensemble* average performance under Maximum-Likelihood (ML) decoding of the proposed codes by means of upper bounds and tight approximations. We show that, differently from the AWGN and fully-interleaved fading cases, Belief-Propagation iterative decoding performs very close to ML on the block-fading channel, even for relatively short block lengths. We also show that, at constant decoding complexity per information bit, the proposed codes perform close to the information outage probability for any block length, while standard block codes (e.g., obtained by trellis-termination of convolutional codes) have a gap from outage that increases with the block length: this is a different and more subtle manifestation of the so-called “interleaving gain” of turbo codes.

3.1 Introduction

Coding for the block-fading channel has been considered in a number of recent works (e.g., [24, 19, 25, 26] and references therein). The design criteria for codes over the block-fading channel differ significantly with respect to the standard design criteria for codes over the AWGN channel or over the fully-interleaved fading channel. Using union bound arguments [24, 19, 25, 26] and error exponent calculations [27], it was shown that in Rayleigh fading the error probability behaves like $O(\text{SNR}^{-d_B})$ for large SNR. The exponent d_B, an integer in $[0, N_B]$, is referred to as the code *block diversity* and is given by the minimum number of blocks on which any two distinct codewords differ (block-wise Hamming distance). If the code is constructed over a finite alphabet (signal set), there exists a trade-off between the achievable block diversity and the coding rate. More precisely, a code over an alphabet $\mathcal{X}$ of cardinality $|\mathcal{X}|$, partitioned into N_B blocks of length L, can be seen as a code over the alphabet $\mathcal{X}^L$ of cardinality $|\mathcal{X}|^L$ with block length N_B. Hence, any upper bound on the minimum Hamming distance of $|\mathcal{X}|^L$-ary codes of length N_B and size A yields an upper bound on the achievable block diversity for a given coding rate $R = \frac{1}{N_B L}\log_2 A$. In [27, Th. 1], it is shown that for binary codes the Singleton bound is tight for any $R \in (0, 1]$. The achievability proof in [27, Th. 1] is based on the existence of maximum distance separable (MDS) codes over $\mathbb{F}_{2^L}$ (e.g., Reed-Solomon codes).

Consider the SNR reliability function $d_B^\star$ defined in (2.13) as the maximum achievable SNR exponent of error probability for codes in a given family of interest,

$$d_B^\star \triangleq \sup_{\mathcal{M}\in\mathcal{F}} \lim_{\rho\to\infty} \frac{-\log P_e(\rho, \mathcal{M})}{\log\rho} \tag{3.1}$$

where ρ denotes the channel SNR, $P_e(\rho, \mathcal{M})$ is the error probability of a given coding scheme $\mathcal{M}$, and the supremum is taken over all coding schemes in the family $\mathcal{F}$.

As illustrated in Chapter 2, in [22], a MIMO block-fading channel with $N_B = 1$ fading blocks is considered and no restriction is imposed on the family of coding schemes (other than the classical average input power constraint). In [22], a coding scheme is defined as a sequence of codes indexed by their operating SNR, with rate $R(\rho) = r\log\rho$ for some $r > 0$, and $d_B^\star$ as a function of r is determined for codes satisfying the input power constraint and given finite block length.

In this work, we consider a SISO block-fading channel with arbitrary (but fixed) number N_B of fading blocks. We are interested in the ensemble of coded modulations, i.e., codes over some given finite signal set $\mathcal{X} \subset \mathbb{C}$ with fixed rate R that, obviously, cannot be larger than $M = \log_2|\mathcal{X}|$ bit/complex dimension. We study the SNR exponent (3.1) as a function of the coding rate, denoted by $d_B^\star(R)$. This represents the optimal rate-diversity trade-off for the given family of codes. We prove that $d_B^\star(R)$ is indeed given by the Singleton bound, and we find an explicit expression for the random-coding SNR error exponent, denoted by $d_B^{(r)}(R)$, which lowerbounds $d_B^\star(R)$ and is tight for all R provided that the code block length grows rapidly enough with respect to $\log(\rho)$: namely, the code block length must be superlinear in the channel SNR expressed in dB. Furthermore, we show that the popular pragmatic Bit-Interleaved Coded Modulation (BICM) scheme [36] achieves the same $d_B^{(r)}(R)$ (and hence $d_B^\star(R)$, subject to the same condition on the block length growth rate).

Then, we focus on the systematic construction of codes achieving the optimal SNR exponent and we introduce a turbo-like code construction suited to the block-fading channel. Notice that standard code ensemble analysis and optimization techniques based on Density Evolution [37] and on various approximations thereof, such as the ubiquitous EXtrinsic Information Transfer (EXIT) functions [38], are not useful over the block-fading channel. In fact, these techniques aim at finding the iterative decoding threshold, defined as the minimum SNR at which the bit error rate (BER) vanishes after infinitely many iterations of the Belief-Propagation (BP) iterative decoder, for a given code ensemble in the limit of infinite block length. In our case, since the block-fading channel is affected by a finite number N_B of fading coefficients that do not average out as the block length grows to infinity, the iterative decoding threshold is a random variable that depends on the channel realization. Hence, one should optimize the distribution of the fixed points of the Density Evolution with respect to the code ensemble: clearly, a very difficult and mostly impractical task.

For our codes we provide upper bounds and tight approximations to the error probability under maximum-likelihood (ML) decoding. While ML decoding is generally infeasible due to its complexity, we show by simulation that the iterative Belief-Propagation (BP) "turbo" decoder performs very close to the ML error probability. This fact stands in stark contrast with the typical behavior of turbo and LDPC codes on the AWGN and fully interleaved fading channels [39, 40, 41, 42, 43], where ML bounds are able to predict accurately the "error floor region" but are quite inaccurate in the "waterfall region" of the BER curve. Hence, our bounds and approximations are relevant, in the sense that they indeed provide very accurate performance evaluation of turbo-like coded modulation in the block-fading channel under BP iterative decoding.

The proposed coded modulation schemes outperform standard turbo-coded or LDPC-coded modulation and outperform also previously proposed trellis codes for the block-fading channel [24, 25, 26]. In particular, by using asymptotic weight enumerator techniques, we argue that the word-error rate (WER) of our codes is almost independent of the block length, while the component encoders are fixed, i.e., the decoding complexity of the BP decoder is linear with the block length. On the contrary, in the case of block codes obtained by trellis termination of trellis codes, the WER increases (roughly linearly) with the block length for linear decoding complexity. We interpret this fact as another manifestation of the so-called “interleaving gain” typical of turbo codes, even though, in block-fading, no “waterfall” behavior of the error curve is visible, even for very large block length.

The rest of the chapter is organized as follows. Section 3.2 recalls the system model introduced in Chapter 2 and presents some of the basic assumptions that we will use throughout the chapter. Section 3.3 presents the coding theorems for the rate-diversity trade-off of coded modulation and BICM. In Section 3.4 we present our novel turbo-like coded modulation scheme, we provide useful upper bounds and approximations of its error probability under ML decoding and we show that the error probability is (asymptotically) independent of the block length. Also, several examples of code constructions and performance comparisons are provided. Section 3.5 summarizes the conclusions of this work. Proofs and computation details of the error bounds and approximations are reported in the appendices.

3.2 System model

Recall the discrete-time complex baseband equivalent channel model introduced in the previous chapter,

$$\mathbf{Y} = \sqrt{\rho}\,\mathbf{H}\mathbf{X} + \mathbf{Z} \tag{3.2}$$

where $\mathbf{Y} \in \mathbb{C}^{N_B \times L}$, is the received signal matrix, $\mathbf{X} \in \mathbb{C}^{N_B \times L}$, is the transmitted signal matrix, $\mathbf{H} = \mathrm{diag}(h_1, \dots, h_{N_B}) \in \mathbb{C}^{N_B \times N_B}$ is the matrix of channel coefficients and $\mathbf{Z} =\in \mathbb{C}^{N_B \times L}$ is the noise matrix with i.i.d. complex circularly-symmetric Gaussian entries $\sim \mathcal{N}_{\mathbb{C}}(0,1)$. We consider codes constructed over a complex signal-set $\mathcal{X}$ (e.g., QAM/PSK) of cardinality 2^M, i.e., the components of $\mathbf{X}$ are points in the constellation $\mathcal{X}$. The overall codeword block length is $N_B L$ (complex dimensions). Without loss of generality, we assume normalized fading, such that $\mathbb{E}[|h_b|^2] = 1$ and unit-energy signal set $\mathcal{X}$ (i.e., $2^{-M}\sum_{x\in\mathcal{X}} |x|^2 = 1$). Therefore, ρ denotes the *average* received SNR and the *instantaneous* SNR on block b is given by $\gamma_b \rho$, where $\gamma_b = |h_b|^2$ denotes the fading *power gain.*

The collection of all possible transmitted codewords $\mathbf{X}$ forms a coded modulation scheme over $\mathcal{X}$. We are interested in schemes $\mathcal{M}(\mathcal{C}, \mu, \mathcal{X})$ obtained by concatenating a binary linear code $\mathcal{C}$ of length $N_B L M$ and rate r bit/symbol with a memoryless one-to-one symbol mapper $\mu : \mathbb{F}_2^M \to \mathcal{X}$. The resulting coding rate (in bit/complex dimension) is given by $R = rM$.

In this work we assume that the vector of fading coefficients $\mathbf{h} = (h_1, \dots, h_{N_B})$ is perfectly known at the receiver and not known at the transmitter. It is worthwhile to notice that in the limit of $L \to \infty$ and fixed N_B, the capacity and, more generally, the outage capacity, of the block-fading channel does not depend on the assumption of perfect channel knowledge at the receiver [7]. Therefore, in this limit such assumption is not optimistic.

Let $\mathsf{m} \in \{1, \dots, |\mathcal{M}|\}$ denote the information message and $\phi(\mathsf{m}) = \mathbf{X}$ denote the codeword corresponding to m. We shall then consider the following decoders:

1. As shown in the previous chapter, the ML decoder is given by

$$\widehat{\mathsf{m}} = \arg \min_{\mathsf{m}=1,\dots,|\mathcal{M}|} \|\mathbf{Y} - \sqrt{\rho}\,\mathbf{H}\phi(\mathsf{m})\|_F^2 \,. \tag{3.3}$$

2. We consider also the suboptimal decoder that consists of producing, for each received symbol, the posterior probabilities of the binary coded symbols in its label (defined by the symbol mapper μ), and then feeding these probabilities to a ML decoder for the binary code $\mathcal{C}$ over the resulting binary-input continuous-output channel. Since this scheme is particularly effective if used in conjunction with BICM [36], we shall refer to it as the BICM-ML decoder (even though it can also be used without an explicit bit-interleaver between $\mathcal{C}$ and μ). For a coded binary symbol mapped to the m-th label position of the k-th modulation symbol of block b, the "bit-wise" posterior log-probability ratio is given by

$$\mathcal{L}_{b,k,m} = \log \frac{\displaystyle\sum_{x \in \mathcal{X}_0^m} \exp\left(-|y_{b,k} - \sqrt{\rho}\,h_b x|^2\right)}{\displaystyle\sum_{x \in \mathcal{X}_1^m} \exp\left(-|y_{b,k} - \sqrt{\rho}\,h_b x|^2\right)} \tag{3.4}$$

where $\mathcal{X}_a^m$ denotes the signal subset of all points in $\mathcal{X}$ whose label has value $a \in \{0,1\}$ in position m. Let again m denote the information message and let $\phi_{\mathcal{C}}(\mathsf{m}) = \mathbf{C} \in \mathcal{C}$ denote the codeword of $\mathcal{C}$ corresponding to m, where $\phi_{\mathcal{C}} : \{1, \dots, |\mathcal{M}|\} \to \mathbb{F}_2^N$, is the corresponding encoding function. The component of $\mathbf{C}$ mapped to the m-th label position of the k-th modulation symbol of block b is denoted by $c_{b,k,m}(\mathsf{m})$. Then, the BICM-ML decoding rule is given by

$$\widehat{\mathsf{m}} = \arg \max_{\mathsf{m}=1,\dots,|\mathcal{M}|} \sum_{b=1}^{N_B} \sum_{k=1}^{L} \sum_{m=1}^{M} (1 - 2c_{b,k,m}(\mathsf{m}))\mathcal{L}_{b,k,m} \tag{3.5}$$

In all cases, the average word-error rate (WER) as a function of SNR, averaged over the fading, is defined as $P_e(\rho) = \Pr(\mathsf{m} \neq \widehat{\mathsf{m}})$ where a uniform distribution of the messages is assumed.

As it will be clear in the following, both the ML and the BICM-ML decoders are practically infeasible for the class of coded modulation schemes proposed in this work. Hence, the suboptimal turbo decoder based on Belief-Propagation (BP) will be used instead. Nevertheless, the two decoders defined above are easier to analyze and provide a benchmark to compare the performance of the BP decoder. Since BP iterative decoding is standard and well-known, for the sake of space limitation we shall omit the detailed BP decoder description. The reader is referred to e.g. [44] for details.

3.3 Optimal rate-diversity trade-off

For the ensemble $\mathcal{M}(\mathcal{C}, \mu, \mathcal{X})$ where $\mathcal{C}$ is a random binary linear code, $P_{\mathcal{X}}$ is the uniform i.i.d. distribution over $\mathcal{X}$. Under this probability assignment, we have that

$$I(P_{\mathcal{X}}, \mathbf{h}) = \frac{1}{N_B} \sum_{b=1}^{N_B} J_{\mathcal{X}}(\gamma_b \rho) \tag{3.6}$$

where

$$J_{\mathcal{X}}(s) \triangleq M - 2^{-M} \sum_{x \in \mathcal{X}} \mathbb{E}\left[\log_2 \sum_{x' \in \mathcal{X}} e^{-|\sqrt{s}(x-x')+Z|^2 + |Z|^2}\right] \tag{3.7}$$

is the mutual information of an AWGN channel with input $X \sim \text{Uniform}(\mathcal{X})$ and SNR s (expectation in (3.7) is with respect to $Z \sim \mathcal{N}_{\mathbb{C}}(0,1)$).

We define the BICM channel associated to the original block-fading channel by including the mapper μ, the modulator $\mathcal{X}$ and the BICM-ML posterior log-probability ratios computer (3.4) as part of the channel and not as a part of a (suboptimal) encoder and decoder. Following [36], the associated BICM channel can be modeled as set of M binary-input symmetric-output channels, where the input and output of the m-th channel over the b-th fading block are given by $\{c_{b,k,m} \;:\; k = 1, \ldots, LM\}$ and $\{\mathcal{L}_{b,k,m} \;:\; k = 1, \ldots, LM\}$, respectively. The resulting mutual information is given by

$$J_{\mathcal{X},\text{BICM}}(s) \triangleq M - 2^{-M} \sum_{m=1}^{M} \sum_{a=0}^{1} \sum_{x \in \mathcal{X}_a^m} \mathbb{E}\left[\log_2 \frac{\sum_{x' \in \mathcal{X}} e^{-|\sqrt{s}(x-x')+Z|^2}}{\sum_{x' \in \mathcal{X}_a^m} e^{-|\sqrt{s}(x-x')+Z|^2}}\right] \tag{3.8}$$

Notice that the expectation over $Z \sim \mathcal{N}_{\mathbb{C}}(0,1)$ in (3.7) and (3.8) can be easily evaluated by using the Gauss-Hermite quadrature rules which are tabulated in [45] and can be computed using for example the algorithms described in [46].

The information outage probabilities of the block-fading channel with i.i.d. input $X \sim \mathcal{N}_{\mathbb{C}}(0,1)$, $X \sim \text{Uniform}(\mathcal{X})$ and that of the associated BICM channel are denoted by $P_{\text{out}}^{G}(\rho, R)$, $P_{\text{out}}^{\mathcal{X}}(\rho, R)$ and by $P_{\text{out}}^{\mathcal{X},\text{BICM}}(\rho, R)$, respectively. From the data processing inequality and the fact that the proper complex Gaussian distribution maximizes differential entropy [47], we obtain that

$$P_{\text{out}}^{G}(\rho, R) \leq P_{\text{out}}^{\mathcal{X}}(\rho, R) \leq P_{\text{out}}^{\mathcal{X},\text{BICM}}(\rho, R) \tag{3.9}$$

for all R and ρ.

By evaluating the outage probability for a given signal set $\mathcal{X}$ we can assess the performance loss incurred by the suboptimal coded modulation ensemble $\mathcal{M}(\mathcal{C}, \mu, \mathcal{X})$. Furthermore, by evaluating the outage probability of the BICM channel, we can assess the performance loss incurred by the suboptimal BICM-ML decoder with respect to the ML decoder.

For the sake of simplicity, we consider independent Rayleigh fading, i.e., the fading coefficients h_b are i.i.d., $\sim \mathcal{N}_{\mathbb{C}}(0,1)$ and the fading power gains γ_b are Chi-squared with two degrees of freedom, i.e., $\gamma_b \sim f_\gamma(z) = e^{-z}\mathbb{1}\{z \geq 0\}$, where $\mathbb{1}\{\mathcal{E}\}$ denotes the indicator function of the event $\mathcal{E}$. This assumption will be discussed and relaxed at the end of this section.

We are interested in the SNR reliability function (3.1) of the block-fading channel. In the previous chapter we have shown that Lemma 1, that follows as a corollary of the analysis in [22], yields the SNR reliability function subject to the average input power constraint. Now, for the considered coded modulation ensemble, we have the following result:

Theorem 1 *Consider the block-fading channel (3.2) with i.i.d. Rayleigh fading and input signal set $\mathcal{X}$ of cardinality 2^M. The SNR reliability function of the channel is upperbounded by the Singleton bound*

$$d_B^\star(R) \leq d_{SB}(R) \triangleq 1 + \left\lfloor N_B\left(1 - \frac{R}{M}\right)\right\rfloor \tag{3.10}$$

The random coding SNR exponent of the coded modulation ensemble $\mathcal{M}(\mathcal{C}, \mu, \mathcal{X})$ defined previously, with block length $L(\rho)$ satisfying $\lim_{\rho\to\infty} \frac{L(\rho)}{\log\rho} = \beta$ and rate R, is lowerbounded by

$$d_B^{(r)}(R) \geq \beta N_B M \log(2)\left(1 - \frac{R}{M}\right), \tag{3.11}$$

for $0 \leq \beta < \frac{1}{M\log(2)}$ and

$$d_B^{(r)}(R) \geq d_{SB}(R) - 1 + \min\left\{1, \beta M\log(2)\left[N_B\left(1 - \frac{R}{M}\right) - d_{SB}(R) + 1\right]\right\} \tag{3.12}$$

for $\frac{1}{M\log(2)} \leq \beta < \infty$.

Furthermore, the SNR random coding exponent of the associated BICM channel satisfies the same lower bounds (3.11) and (3.12).

Proof. See Appendix C.2. □

An immediate consequence of Theorem 1 is the following

Corollary 1 *The SNR reliability function of the block-fading channel with input $\mathcal{X}$ and of the associated BICM channel is given by $d_B^\star(R) = d_{SB}(R)$ for all $R \in (0, M]$, except for the N_B discontinuity points of $d_{SB}(R)$, i.e., for the values of R for which $N_B(1 - R/M)$ is an integer.*

Proof. We let $\beta \to \infty$ in the random coding lower bound (3.11) and we obtain

$$d_{SB}(R) \geq d^{\star}_B(R) \geq d^{(r)}_B(R) \geq \left\lceil N_B \left(1 - \frac{R}{M}\right)\right\rceil$$

where the rightmost term coincides with $d_{SB}(R)$ for all points $R \in (0, M]$ where $d_{SB}(R)$ is continuous. □

The following remarks are in order:

1. The codes achieving the optimal diversity order $d^{\star}_B(R)$ in Theorem 1 are found in the ensemble $\mathcal{M}(\mathcal{C}, \mu, \mathcal{X})$ with block length that increases with SNR faster than $\log(\rho)$. This is due to the fact that, differently from the Gaussian ensemble (Lemma 1), for a given discrete signal set $\mathcal{X}$ there is a non-zero probability that two codewords are identical, for any finite length L. Hence, we have to make L increase with ρ rapidly enough such that this probability does not dominate the overall probability of error. Nevertheless, it is easy to find explicit constructions achieving the optimal Singleton bound block-diversity $d_{SB}(R)$ for several cases of N_B and finite L [24, 25]. Typically, the WER of diversity-wise optimal codes behaves like $\mathcal{K}\rho^{-d_{SB}(R)}$ for large ρ. The coefficient $\mathcal{K}$ yields a horizontal shift of the WER vs. SNR curve (in a log-log chart) with respect to the outage probability curve $P^{\mathcal{X}}_{\text{out}}(\rho, R)$ that we refer to as "gap from outage".

   Codes found in previous works [24, 19, 25, 26] have a gap from outage that increases with the block length L. On the contrary, the gap from outage of the class of codes proposed in this work is asymptotically independent of the block length. We say that a code ensemble is *good* if it achieves *vanishing* gap from outage as $L \to \infty$. We say that a code ensemble is *weakly good* if it achieves *constant* gap from outage as $L \to \infty$. In Section 3.4.3 we show that the proposed codes are weakly good under ML decoding.

2. For any given coding rate R, we can achieve "full diversity" $d_{SB}(R) = N_B$ by considering a signal set large enough. In fact, by letting M sufficiently large in (3.10) we have $d_{SB}(R) = N_B$ for any desired rate R. This corresponds to the intuitive argument that large signal sets approach Gaussian codes.[1]

Again, we can relax the assumption of Rayleigh fading by noticing that in the proof of Theorem 1 only the near-zero behavior of the fading power gain distribution has a role. For Rayleigh fading, we have $\Pr(\gamma_b \leq \delta) \approx \delta$, for small $\delta > 0$. Hence, the above results hold for all block-fading channels with i.i.d.

---

[1] For finite SNR, expanding the signal set without proper shaping incurs shaping loss. However, in terms of SNR exponent this effect is not seen as shaping involves only a fixed gap from outage. Using the definition introduced above, we might say that codes found in our ensemble of coded modulation schemes over larger and larger QAM complex constellations can be weakly good, but cannot be good due to the inherent shaping loss.

fading with power gain distribution with this behavior. More in general, as argued in [22], for a fading distribution with near-zero behavior $\Pr(\gamma_b \leq \delta) \approx \delta^D$, the SNR reliability function is given by $Dd_{SB}(R)$. For example, this is the case of independent Rayleigh fading with a D antenna receiver using D-fold maximal-ratio combining [23].

Figure 3.1 shows $d_{SB}(R)$ (Singleton bound) and the random coding lower bounds for the two cases $\beta M \log(2) = 1/2$ and $\beta M \log(2) = 2$, in the case $N_B = 8$ and $M = 4$ ($\mathcal{X}$ is a 16-ary signal set). It can be observed that as β increases (for fixed M), the random coding lower bound coincides over a larger and larger support with the Singleton upper bound. However, in the discontinuity points it will never coincide.

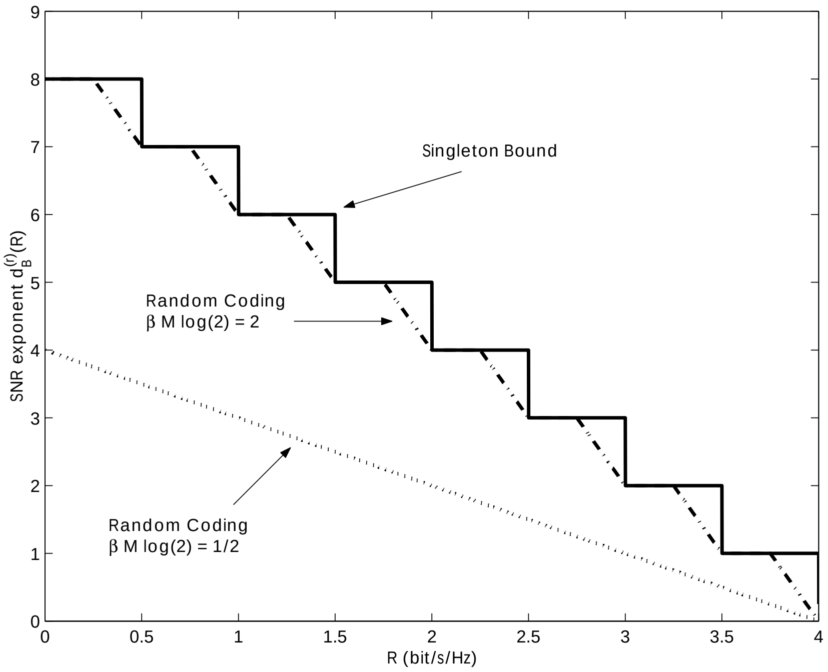


Figure 3.1: SNR reliability function and random coding exponents $d_B^{(r)}(R)$ for $N_B = 8$ and $M = 4$.

In order to illustrate the operational meaning of the above results and motivate the code construction in the following section, we show in Figure 3.2 the outage probability curves as a function of SNR with i.i.d. Rayleigh fading for the case of $N_B = 8$ and spectral efficiencies $R = 1, 1.5, 2$ bit/complex dimension for Gaussian inputs, 8-PSK and 16-QAM constellations and the corresponding Gray-mapped BICM [36]. In these log-log charts, the SNR exponent determines the slope of the outage probability curve at high SNR (small outage probability). We notice that Gaussian inputs always show the steepest slope and that this is independent of R for high SNR (in agreement with Lemma 1). For $R = 1$, we observe that both modulations (and the corresponding BICM curves) have the same slope (we have $d_{SB}(1) = 7$ for both modulations). However, for $R = 1.5$, the curves show different slopes since $d_{SB}(1.5) = 5$ for 8-PSK while $d_{SB}(1.5) = 6$ for 16-QAM. This effect is even more evident for $R = 2$, where $d_{SB}(2) = 4$ for 8-PSK and $d_{SB}(2) = 5$ for 16-QAM. Notice also that, in all cases, the SNR loss incurred by BICM-ML decoding is very small.

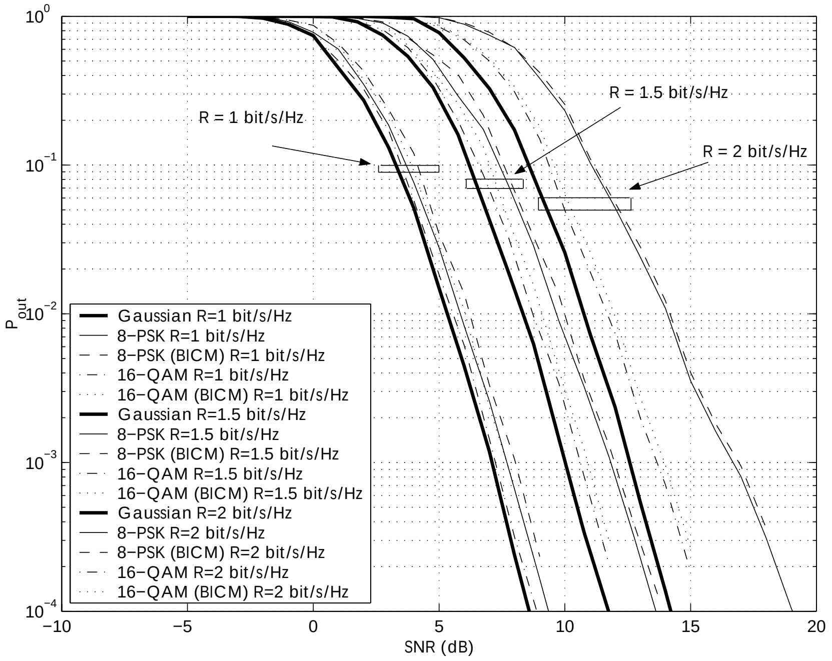


Figure 3.2: Outage probability for $N_B = 8$, $R = 1, 1.5, 2$ bit/complex dimension, Gaussian inputs, 8-PSK and 16-QAM modulations. Thick solid lines correspond to Gaussian inputs, thin solid lines to 8-PSK, dashed lines to 8-PSK with BICM (Gray mapping), dashed-dotted lines to 16-QAM and dotted lines to 16-QAM with BICM (Gray mapping).

3.4 Blockwise concatenated coded modulation

In this section we introduce a general construction for MDS coded modulation schemes $\mathcal{M}(\mathcal{C}, \mu, \mathcal{X})$ for the the block-fading channel and we provide bounds and approximations to their error probability under ML and BICM-ML decoding.

3.4.1 Code construction

Figure 3.3 shows the proposed encoder structure for $\mathcal{M}(\mathcal{C}, \mu, \mathcal{X})$ that we refer to as *Blockwise Concatenated Coding* (BCC). The binary linear code $\mathcal{C}$ is formed by the concatenation of a binary linear outer code $\mathcal{C}^O$ of rate r_O and block length $N_B L_\pi$, partitioned into N_B blocks of length L_π. The blocks are separately interleaved by the permutations $(\pi_1, \ldots, \pi_{N_B})$ and the result is fed into the N_B encoders for the inner code $\mathcal{C}^I$ of rate r_I and length $L_B = LM$. Thus, the total length of $\mathcal{C}$ is $N_B L_B$ (binary symbols). Finally, the output of each component inner code is mapped onto a sequence of signals in $\mathcal{X}$ by the one-to-one symbol mapper μ. We denote by K the number of information bits per codeword. In particular, the codes considered make use of bit-interleaving between the inner encoder and the mapper [36], denoted in Figure 3.3 by the permutations $(\pi_1^\mu, \ldots, \pi_{N_B}^\mu)$. However, we hasten to say that mapping through interleavers is not necessary for the construction and more general mappings could be envisaged. The rate of the resulting blockwise concatenated code is $R = r_O r_I M$.

When the outer code is a simple repetition code of rate $r_O = 1/N_B$ and the inner codes are rate-one accumulators [48], the resulting BCC is referred to as Repeat and *Blockwise* Accumulate (RBA) code. Since interleavers and inner encoding are performed on a blockwise basis, the block diversity of the concatenated code coincides with the block diversity of the outer code. For example, a RBA code has always full diversity $d_B = N_B$. When both outer and inner codes are convolutional codes, we will refer to the resulting structure as *blockwise* concatenated convolutional codes (BCCC).

As anticipated in the Introduction, practical decoding of BCC resorts to the well-known BP iterative decoding algorithm over the code graph [44]. In particular, when either $\mathcal{C}^O$ or $\mathcal{C}^I$ are convolutional codes, the well-known forward-backward decoding algorithm is used over the subgraph representing the corresponding trellis [49].

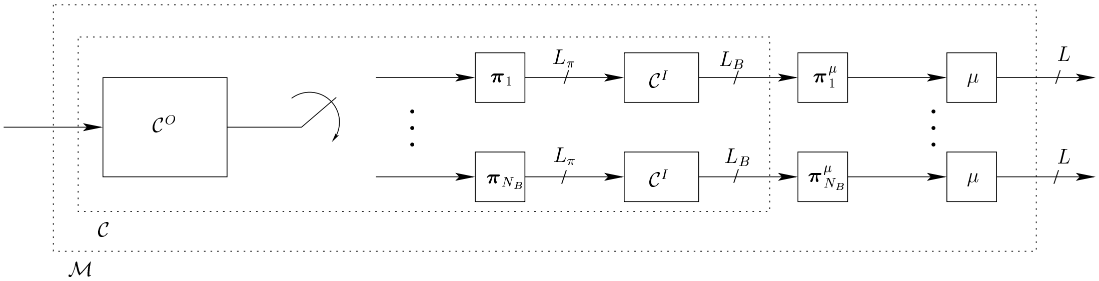


Figure 3.3: The general encoder for Blockwise Concatenated Coding.

Figure 3.4 illustrates the effectiveness of blockwise concatenation with respect to standard turbo-like codes designed for ergodic channels. In particular, we compare the WER of a binary $R = 1/2$ RBA and BCCC (with convolutional $(5,7)_8$ outer code and inner accumulators) with that of their standard counterparts (namely, a Repeat and Accumulate (RA) code and a Serially Concatenated

Convolutional Code (SCCC)), mapped over $N_B = 2$ fading blocks with 10 decoding iterations of the BP decoder. In all cases, the block length corresponds to $K = 1024$ information bits per codeword. We observe a significant difference in the slope of the WER curve, due to the fact that blockwise concatenation preserves d_B of the outer code while standard concatenation does not.

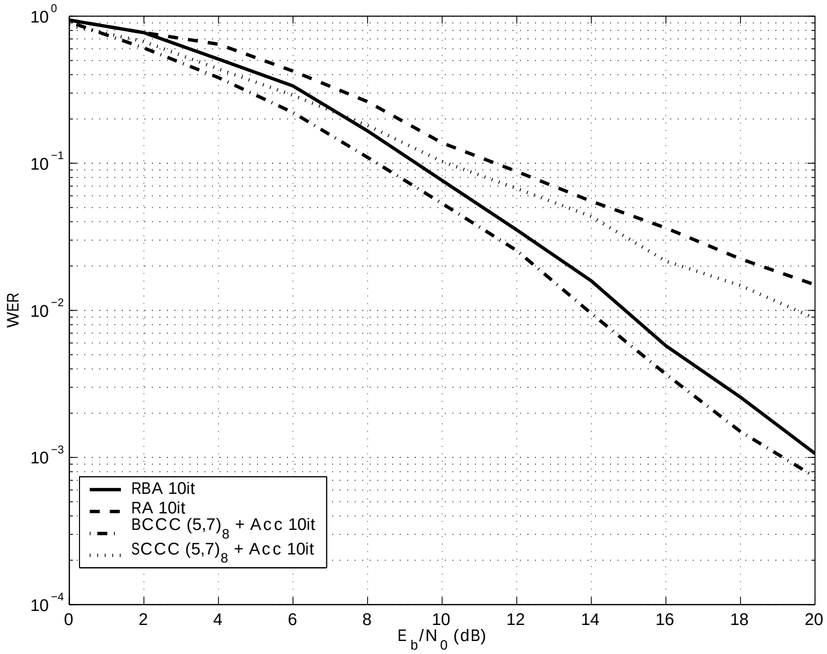


Figure 3.4: WER obtained by BP decoding (simulation with 10 iterations) of binary RBA, RA, BCCC and SCCC of rate $R = 1/2$ for $N_B = 2$ and $K = 1024$.

In order to show that we can construct MDS BCCs, Figure 3.5 illustrates the FER performance obtained by simulation with BP decoding of binary $r = 1/2$ BCCCs $(5,7)_8$ and $(25,35)_8$ both with with inner accumulators, and best known 4 and 64 states CCs [26] mapped over $N_B = 8$ fading blocks with block length of 1024 information bits. Notice that $(5,7)_8$ is not MDS, since $d_{SB} = 4$ while the Singleton bound yields $d_{SB} \leq 5$, and therefore the corresponding BCCC (and of course itself) will show some performance degradation at high SNR. Indeed, we can appreciate a steeper slope of the BCCC with $(25,35)_8$ and the 64 states CC since both are MDS codes. Moreover, we observe clear advantage of BCCCs over standard CCs at this block length (this point will be further discussed in depth in section 3.4.3).

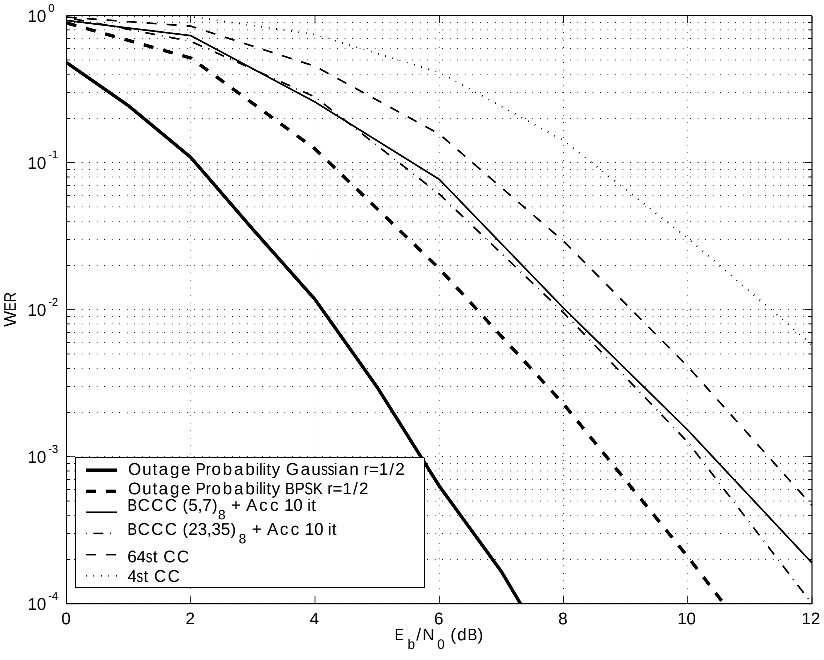


Figure 3.5: FER $r = 1/2$ BCCCs and CCs mapped over $N_B = 8$ fading blocks.

3.4.2 Upper bounds on ML decoding error probability

For the sake of simplicity we consider first codes over the QPSK with Gray labeling, or, equivalently, over BPSK. This case is particularly simple since the squared Euclidean distance between the constellation points is proportional to the Hamming distance between their binary labels. A tight upper bound on the WER of binary codes mapped over QPSK with Gray mapping and transmitted over N_B fading blocks, is given by Malkamaki and Leib (M&L) in [25], and reads

$$P_e(\rho) \leq \mathbb{E}\left[\min\left\{1, \sum_{w_1,\dots,w_{N_B}} A_{w_1,\dots,w_{N_B}} Q\left(\sqrt{\kappa\rho\sum_{b=1}^{N_B}\gamma_b w_b}\right)\right\}\right] \tag{3.13}$$

where $A_{w_1,\dots,w_{N_B}}$ is the Multivariate Weight Enumeration Function (MWEF) of $\mathcal{C}$ [50] which accounts for the number of pairwise error events with output Hamming weights per block $w_1, \dots, w_{N_B}$, $\kappa = 2$ for BPSK and $\kappa = 1$ for QPSK. Expectation in (3.13) is with respect to the fading power gains $(\gamma_1, \dots, \gamma_{N_B})$. In order to compute (3.13), we need to compute a multivariate expectation that does not break into the individual expectation of each term in the union bound because of the $\min\{1, \cdot\}$. Hence, in practice, we have to resort to Monte Carlo methods.

In [51], Byun, Park and Lee presented a simpler upper bound to (3.13) in the context of ML decoding of trellis space-time codes. Unfortunately, the bound in

[51] upperbounds (3.13) only if the sum over $w_1, \ldots, w_{N_B}$ contains a single term. Nevertheless, we shall demonstrate through several examples that this technique, referred to as the BPL *approximation*, if applied to full diversity codes (i.e., codes with blockwise Hamming distance $d_B = N_B$) yields a very good approximation of the WER, with the advantage that it is much easier to compute than the M&L bound.

Assuming $d_B = N_B$, which implies that $\min w_b \geq 0$ for all $b = 1, \ldots, N_B$, the BPL approximation takes on the form

$$P_e(\rho) \lessapprox \mathbb{E}\left[\min\left\{1, \sum_{\Delta_p} A_{\Delta_p} Q\left(\sqrt{\kappa\rho\Delta_p^{1/N_B}\sum_{b=1}^{N_B}\gamma_b}\right)\right\}\right] \tag{3.14}$$

where $\Delta_p \triangleq \prod_{b=1}^{N_B} w_b$ is the product weight and A_{Δ_p} is the Product Weight Enumeration Function (PWEF) of $\mathcal{C}$, i.e., the number of codewords of $\mathcal{C}$ with product weight Δ_p. In this way, only product weights have to be enumerated. Moreover, by noticing that

$$\gamma = \sum_{b=1}^{N_B} \gamma_b \tag{3.15}$$

is central chi-squared with $2N_B$ degrees of freedom and mean N_B, (3.14) becomes,

$$P_e(\rho) \lessapprox \int_0^{+\infty} \min\left\{1, \sum_{\Delta_p} A_{\Delta_p} Q\left(\sqrt{\kappa\rho\Delta_p^{1/N_B} z}\right)\right\} f_\gamma(z)dz \tag{3.16}$$

where

$$f_\gamma(z) = \frac{z^{N_B-1}}{(N_B-1)!}e^{-z} \tag{3.17}$$

is the pdf of γ. In this way, only product weights have to be enumerated and the computation of (3.16) requires one-dimensional integration, that is easily computed numerically.

Union bound-based techniques are known to be loose for turbo codes and other capacity-approaching code ensembles such as LDPC and RA codes over AWGN or fully-interleaved fading channels. As a matter of fact, improved bounding techniques are needed in ergodic channels order to obtain meaningful upper bounds in the SNR range between the capacity threshold and the cut-off rate threshold [39, 40, 41, 42, 43]. Among those, the tangential-sphere bound (TSB) is known to be the tightest. The TSB can be simply extended to the block-fading channel for each fixed realization of the fading vector $\mathbf{h}$ (for more details the reader is referred to Appendix D). Then, an outer Monte Carlo average over the fading is required. Since the TSB requires the optimization of certain parameters for each new fading realization, the computation of the TSB is very intensive. A slight simplification is obtained by applying the TSB technique to the PWEF, as in the BPL approximation. The resulting approximation (referred to as BPL-TSB) requires only a single variate expectation.

The following examples illustrate the bounds and the approximations described above for BPSK and QPSK with Gray mapping. The MWEF and PWEF are obtained as described in Appendix B. In particular, Figures 3.6 and 3.7 compare the simulation (with 10 iterations of the BP decoder) with the ML bounds and approximations for RBA codes of $R = 1/2$ with $K = 256$ and $K = 1024$ information bits per codeword respectively over $N_B = 2$ fading blocks. The statistical average in the M & L bound and in the TSB are computed by Monte Carlo. We observe an excellent matching between the performance of BP decoding and the bounds on ML decoding, even for such short block lengths, in contrast to the AWGN case. We also notice that the TSB is only marginally tighter than the M&L bound and, due to its computational complexity, is useless in this context. The BPL approximation predicts almost exactly the WER of the RBA code for all block lengths. Based on such examples (and on very extensive numerical computations not reported here for the sake of space limitation) we conclude that the performance of BCCs on block-fading channels can be predicted very accurately by very simple ML analysis techniques.

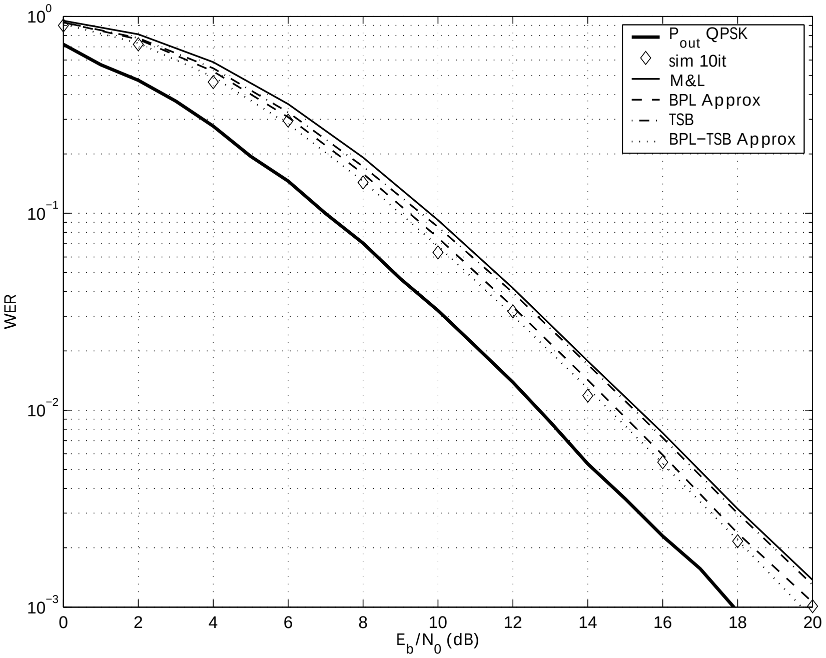


Figure 3.6: WER obtained by BP decoding simulation with 10 iterations and ML bounds and approximations for binary RBA of $R = 1/2$ and $K = 256$ over $N_B = 2$ blocks.

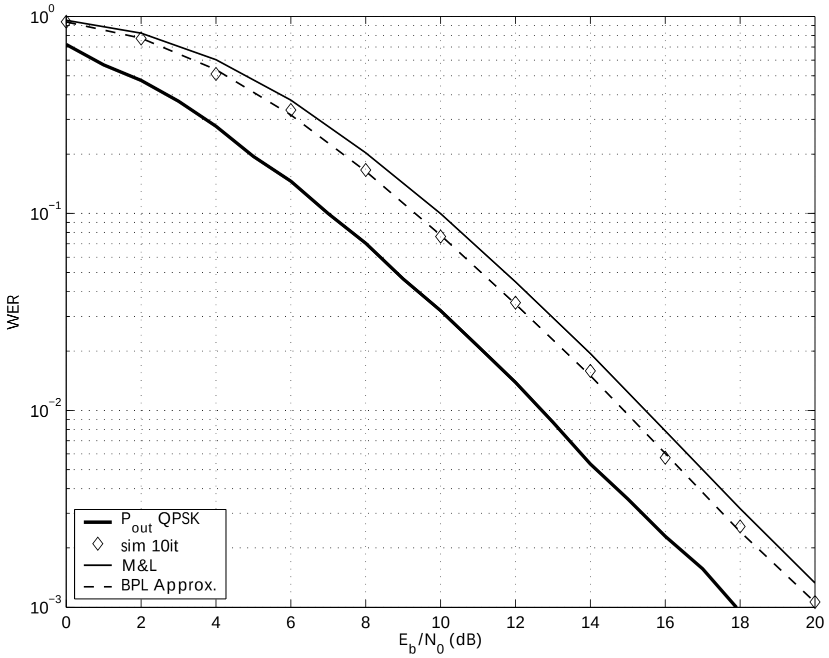


Figure 3.7: WER obtained by BP decoding simulation with 10 iterations and ML bounds for binary RBA of $R = 1/2$ and 1024 information bits per frame, over $N_B = 2$ blocks.

For general signal sets $\mathcal{X}$ and modulator mappings μ the above bounds are no longer valid since the squared Euclidean distance between signals depends, in general, on the individual labels and not only on the labels Hamming distance. Assuming bit-interleaving between the inner binary codes and the modulator mapping, we can make use of the BICM Bhattacharyya union bound developed in [36], combined with the "limit before average" approach of [25]. We then obtain

$$P_e(\rho) \leq \mathbb{E}\left[\min\left\{1, \sum_{w_1,\dots,w_{N_B}} A_{w_1,\dots,w_{N_B}} \frac{1}{2} \prod_{b=1}^{N_B} B_b(\rho,\mu,\mathcal{X})^{w_b}\right\}\right] \tag{3.18}$$

where

$$B_b(\rho,\mu,\mathcal{X}) \triangleq \frac{2^{-M}}{M} \sum_{m=1}^{M} \sum_{a=0}^{1} \sum_{x \in \mathcal{X}_a^m} \mathbb{E}\left[\sqrt{\frac{\sum_{x' \in \mathcal{X}_{\bar{a}}^m} e^{-|\sqrt{\rho\gamma_b}(x-x')+Z|^2}}{\sum_{x' \in \mathcal{X}_a^m} e^{-|\sqrt{\rho\gamma_b}(x-x')+Z|^2}}}\right] \tag{3.19}$$

is the Bhattacharyya factor of the BICM channel associated to the b-th fading block, with SNR $\gamma_b\rho$.

The bound (3.18) holds under the assumption that the mapping μ is symmetrized, as explained in [36], i.e., that a random i.i.d. scrambling sequence, known both at the transmitter and at the receiver, chooses at every symbol with probability 1/2 either the mapping μ or its complement $\bar{\mu}$, obtained by replacing all binary labels by their modulo-2 complement.[2] The factor 1/2 in front of the Bhattacharyya union bound follows from the fact that, under the symmetrized mapping assumption, the associated BICM channel with inputs $c_{b,k,m}$ and outputs $\mathcal{L}_{b,k,m}$ defined in (3.4) is binary-input output-symmetric (see [52]). The expectation in (3.19) can be efficiently computed by Gauss-Hermite quadratures.

As shown in Appendix A, the tail of the pdf of the bit-wise posterior log-probability ratio (3.4) at the output of the associated BICM channel is very close to the corresponding output of a binary-input AWGN channel with fading power gain

$$\zeta_b = -\frac{1}{\rho}\log B_b(\rho,\mu,\mathcal{X}) \tag{3.20}$$

Moreover, for given fading gain γ_b we have (see Appendix A for details)

$$\lim_{\rho\to\infty}\zeta_b = \frac{\delta_{\text{min}}^2}{4}\gamma_b \tag{3.21}$$

where δ_{min}^2 is the minimum squared Euclidean distance of $\mathcal{X}$. Under this Gaussian approximation, we obtain

$$P_e(\rho) \lessapprox \mathbb{E}\left[\min\left\{1, \sum_{w_1,\ldots,w_{N_B}} A_{w_1,\ldots,w_{N_B}} Q\left(\sqrt{2\rho\sum_{b=1}^{N_B} w_b\zeta_b}\right)\right\}\right], \tag{3.22}$$

and the corresponding BPL approximation (for full diversity codes)

$$P_e(\rho) \lessapprox \mathbb{E}\left[\min\left\{1, \sum_{\Delta_p} A_{\Delta_p} Q\left(\sqrt{2\rho\Delta_p^{1/N_B}\sum_{b=1}^{N_B}\zeta_b}\right)\right\}\right]. \tag{3.23}$$

Unfortunately, in this case $\sum_{b=1}^{N_B}\zeta_b$ is no longer chi-squared distributed (from (3.21) it follows that it is chi-squared in the limit of high SNR). Therefore, (3.23) has to be computed via a Monte Carlo average, reducing only slightly the computational burden with respect to (3.22). We will refer to (3.18) as the M&L-Bhattacharyya bound and to (3.22) as the M&L-GA.

We hasten to say that, although the proposed methods are *just* approximation, they represent so far the only alternative to extensive simulation. Indeed, they might be regarded as the analogous for the block-fading channel to the EXIT chart "analysis" commonly used for fully-interleaved fading channels and AWGN channels: they are both based on approximating a complicated binary-input output-symmetric channel by a binary-input AWGN channel, "matched" in some sense to the former.

---

[2] If the mapping μ and the constellation $\mathcal{X}$ are such that, for all label positions $m = 1,\ldots,M$, the log-probability ratio defined in (3.4) is symmetrically distributed, that is, $p_{\mathcal{L}_{b,k,m}}(z|c_{b,k,m} = a) = p_{\mathcal{L}_{b,k,m}}(-z|c_{b,k,m} = \bar{a})$, then the scrambling assumption is not needed.

In Figure 3.8 we show the WER (obtained by simulation with 10 iterations of the BP decoder) and the various upper bounds and approximations on ML decoding error probability described above, for a RBA code of rate $r = 1/2$ over $N_B = 2$ fading blocks and $K = 256$ information bits per frame, with 8-PSK and 16-QAM with Gray mapping (the corresponding spectral efficiencies are $R = 1.5$ and 2 bit/complex dimension). We also show the outage probability for 8-PSK and 16-QAM BICM for the sake of comparison. Again, we observe an excellent match between simulation with BP decoding and ML approximations, for all modulations. We also observe that the BICM Bhattacharyya bound is looser than the Gaussian Approximation (3.22).

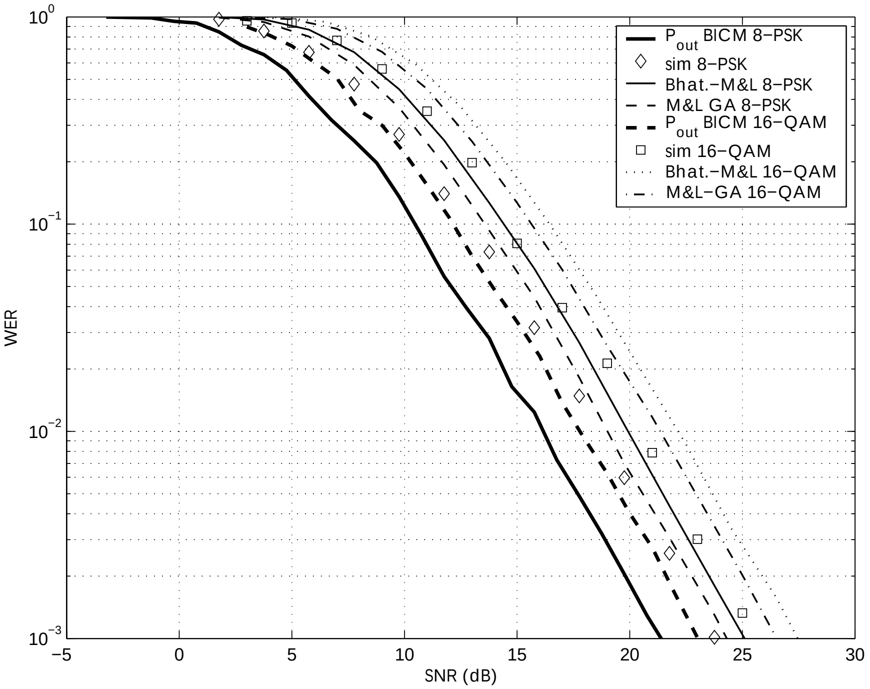


Figure 3.8: WER obtained by BP decoding simulation with 10 iterations and ML bounds and approximations for RBA with BICM of $r = 1/2$ over $N_B = 2$ blocks with 8-PSK and 16-QAM with Gray mapping.

3.4.3 Weak goodness of BCC ensembles

As introduced in Section 3.3, we say that a code ensemble over $\mathcal{X}$ is good if, for block length $L \to \infty$, its WER converges to the outage probability $P_{\text{out}}^{\mathcal{X}}(\rho, R)$. We say that a code ensemble over $\mathcal{X}$ is weakly good if, for block length $L \to \infty$, its WER shows a fixed SNR gap to outage probability, asymptotically independent of L. In this section we give an explicit sufficient condition for weak goodness in terms of the asymptotic exponential growth rate function [53] of the multivariate weight enumerator of specific ensembles.

The issue of weak goodness is non-trivial, as illustrated by the following argument. A code ensemble $\mathcal{M}(\mathcal{C}, \mu, \mathcal{X})$ such that, for all sufficiently large L, a randomly generated member in the ensemble attains the Singleton bound with probability 1 is a good candidate for weak goodness. However, this condition is neither necessary nor sufficient. For example, the ensemble $\mathcal{M}(\mathcal{C}, \mu, \mathcal{X})$ considered in Theorem 1 has a small but non-zero probability that a randomly selected member is not blockwise MDS, nevertheless it attains the optimal SNR exponent provided that L grows faster than $\log \rho$, and hence it is weakly good. On the contrary, the ensemble of random BCCs with given outer and non-trivial inner encoders and the ensemble of blockwise partitioned CCs (i.e., BCCs with convolutional outer encoder and rate-1 identity encoder considered in [24, 19, 25, 26]) that can be seen as BCCs with convolutional outer encoder and trivial (identity) inner encoder, attain the Singleton bound with probability 1 provided that the outer code is blockwise MDS. Nevertheless, simulations show that while the WER of general BCCs with recursive inner encoder is almost independent of the block length, the WER of CCs grows roughly linearly with the block length. For example, Fig.3.9 shows the WER for fixed SNR versus the information block length K, for the ensemble of $R = 1/4$ RBA codes and the standard 64-states CCs with generators $(135, 135, 147, 163)_8$ mapped over $N_B = 4$ blocks, and of $r = 1/2$ BCCs (with outer convolutional encoder $(5, 7)_8$ and inner accumulators) and the 64-states CCs mapped over $N_B = 8$ blocks optimized in [26] with generators $(103, 147)_8$ for the block-fading channel. The different behavior of the WER as a function of the block length for the two ensembles is evident.

We focus first on codes over the BPSK modulation. Therefore, in this case $L = L_B$. Let $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_{N_B}) \in [0, 1]^{N_B}$ be the vector of normalized Hamming weights per block. The asymptotic exponential growth rate function [53] of the multivariate weight enumerator is defined by

$$a(\boldsymbol{\omega}) \triangleq \lim_{\epsilon \to \infty} \lim_{L_B \to \infty} \frac{1}{L_B} \log \left| \mathcal{S}_{\epsilon}^{L_B}(\boldsymbol{\omega}) \right| \tag{3.24}$$

where $\mathcal{S}_{\epsilon}^{L_B}(\boldsymbol{\omega})$ is the set of codewords in the length-L_B ensemble with Hamming weights per block satisfying

$$|w_b / L_B - \omega_b| \leq \epsilon, \quad b = 1, \ldots, N_B \tag{3.25}$$

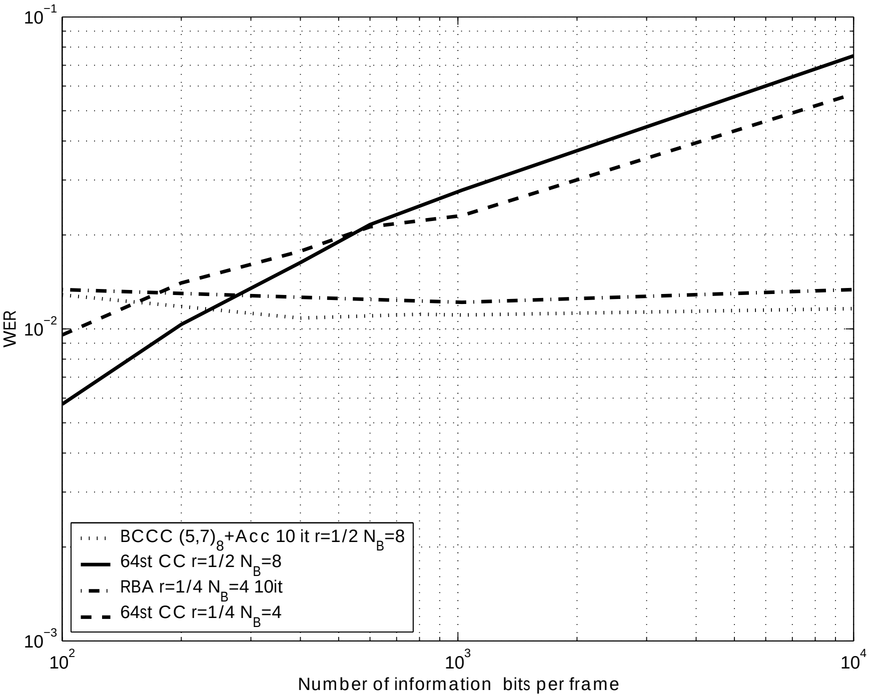


Figure 3.9: WER vs. block length (information bits) at $E_b/N_0 = 8$dB for binary BCC, RBA and trellis terminated CCs obtained by simulation (10 iterations of BP decoding for the BCCs and ML decoding for the CCs using the Viterbi algorithm).

We have the following results:

Theorem 2 *Consider an ensemble of codes* $\mathcal{M}(\mathcal{C}, \mu, \mathcal{X})$ *of rate* R*, where* $\mathcal{X}$ *is BPSK, transmitted over a block-fading channel with* N_B *blocks. Let* $a(\boldsymbol{\omega})$ *be the asymptotic exponential growth rate function of the ensemble multivariate weight enumerator. For* $1 \leq k \leq N_B$*, let* $\mathcal{W}(N_B, k) \in \mathbb{F}_2^{N_B}$ *denote the set of binary fading vectors with Hamming weight not smaller than* $N_B - k + 1$*, and define* $\widehat{s}$ *to be the infimum of all* $s \geq 0$ *such that*

$$\inf_{\mathbf{x} \in \mathcal{W}(N_B, d_{SB}(R))} \ \inf_{\boldsymbol{\omega} \in [0,1]^{N_B}} \left\{ s \sum_{b=1}^{N_B} x_b \, \omega_b - a(\boldsymbol{\omega}) \right\} > 0 \tag{3.26}$$

If $\widehat{s} < \infty$*, then the code ensemble is weakly good.*

Proof. See Appendix C.3. ☐

As far as higher order coded modulations are concerned, we have the following

Corollary 2 *Consider an ensemble of codes* $\mathcal{M}(\mathcal{C}, \mu, \mathcal{X})$ *of rate* R*, where* $\mathcal{X}$ *is a complex signal set of size* 2^M*, transmitted over a block-fading channel with* N_B *blocks, where modulation is obtained by (random) bit-interleaving and decoding by the BICM-ML decoder defined by (3.5). If the underlying ensemble of binary codes (i.e., mapping the binary symbols directly onto BPSK) is weakly good, then the ensemble* $\mathcal{M}(\mathcal{C}, \mu, \mathcal{X})$ *is weakly good.*

Proof. See Appendix C.3. ☐

The above results (and the proofs of Appendix C.3) reveal that the error probability of weakly good codes in the regime where both the block length and the SNR are large is dominated by the event that more than $d_{SB}(R)$ fading components are *small* (in the sense of the proof of Theorem 2). This is precisely the same behavior of the information outage probability for the rate R and discrete signal set $\mathcal{X}$. On the contrary, when less than $d_{SB}(R)$ fading components are small, the code projected over the significant fading components has a finite ML decoding threshold (with probability 1). Therefore, apart from some SNR gap, its error probability vanishes for all such fading realizations. It is also intuitively clear that, due to this sharp threshold behavior, hitting the SNR transition region (known as "waterfall") for which the error probability is non-vanishing even if the fading has less than $d_{SB}(R)$ small components is an event of small probability. This partially explains why BP iterative decoding performs very close to ML in block-fading channels and why more refined bounding techniques such as the TSB do not provide almost any improvement. In fact, it is well-known that BP and ML perform similarly on both the high-error probability region (below the ML decoding threshold) and in the low-error probability region (above the iterative decoding threshold). The gap between the ML and the iterative decoding thresholds seems to play a negligible role in the block-fading channel, for ensembles of weakly good codes. The sharper and sharper transition between the below-threshold and above-threshold regimes of random-like concatenated codes is referred to as *interleaving gain* in [54, 55]. We argue that weak goodness of BCCs in block-fading channels is another manifestation of interleaving gain, even if on such channel no waterfall behavior is observed.

In Appendix C.3 we show also that the ensemble of trellis terminated CCs of increasing block length considered in [24, 19, 25, 26] does not satisfy the condition of Theorem 2. Numerical verification of Theorem 2 is needed for a specific code ensemble. In particular, one has to show that

$$\sup_{\mathbf{x}\in\mathcal{W}(N_B,d_{SB}(R))} \ \sup_{\boldsymbol{\omega}\in[0,1]^{N_B}} \frac{a(\boldsymbol{\omega})}{\sum_{b=1}^{N_B} x_b\omega_b} < \infty \tag{3.27}$$

Supported by simulations and by explicit calculation of the multivariate weight enumerator for RBAs (see Appendix B), we conjecture this is true for the family of random BCCs with MDS outer code and inner recursive encoders.

As an example, in Figure 3.10 we show the asymptotic WER for the RBA ensemble of rate 1/2 with BPSK modulation, over a channel with $N_B = 2$ fading blocks. The asymptotic WER is computed via the asymptotic Bhattacharyya M&L bound given by

$$P_e(\rho) \leq \Pr\left(\max_{\boldsymbol{\omega}\in[0,1]^{N_B}} \frac{a(\boldsymbol{\omega})}{\sum_{b=1}^{N_B} \omega_b\gamma_b} \geq \rho\right) \tag{3.28}$$

and motivated in Appendix C.3. Simulations (BP iterative decoder) for block lengths $K = 100, 1000$ and 10000 information bits per codeword are shown for comparison. This figure clearly shows that the WER of these codes becomes quickly independent of the block length and shows fixed gap from the outage probability.

In order to illustrate the weak goodness of BCCs with BICM and high-order modulations, Figure 3.11 shows the asymptotic WER of an RBA code of rate $R = 2$ bit/complex dimension with 16-QAM modulation (Gray mapping) over $N_B = 2$ fading blocks. The asymptotic WER is computed via the asymptotic Bhattacharyya M&L bound given by

$$P_e(\rho) \leq \Pr\left(\max_{\boldsymbol{\omega}\in[0,1]^{N_B}} \frac{a(\boldsymbol{\omega})}{\sum_{b=1}^{N_B} \omega_b\zeta_b} \geq \rho\right) \tag{3.29}$$

and motivated in Appendix C.3, where ζ_b is defined in (3.20). Simulations (BP iterative decoder) for block lengths $K = 100, 1000$ and 10000 information bits per codeword are shown for comparison.

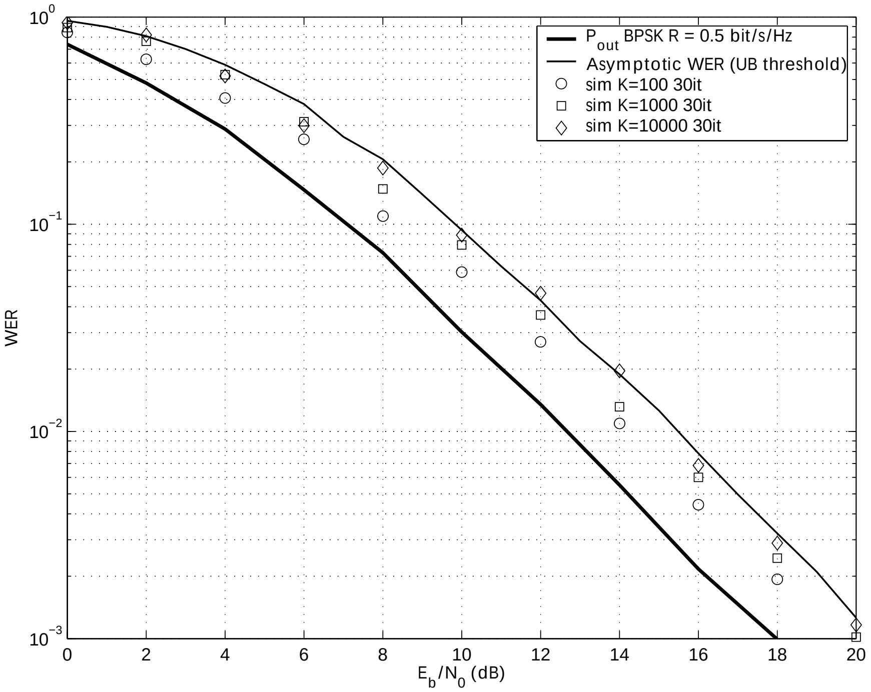


Figure 3.10: Asymptotic error probability (3.28) for a binary rate $r = 1/2$ RBA code mapped over $N_B = 2$ fading blocks and corresponding BP decoding simulation with 30 iterations and $K = 100, 1000$ and 10000 information bits per codeword.

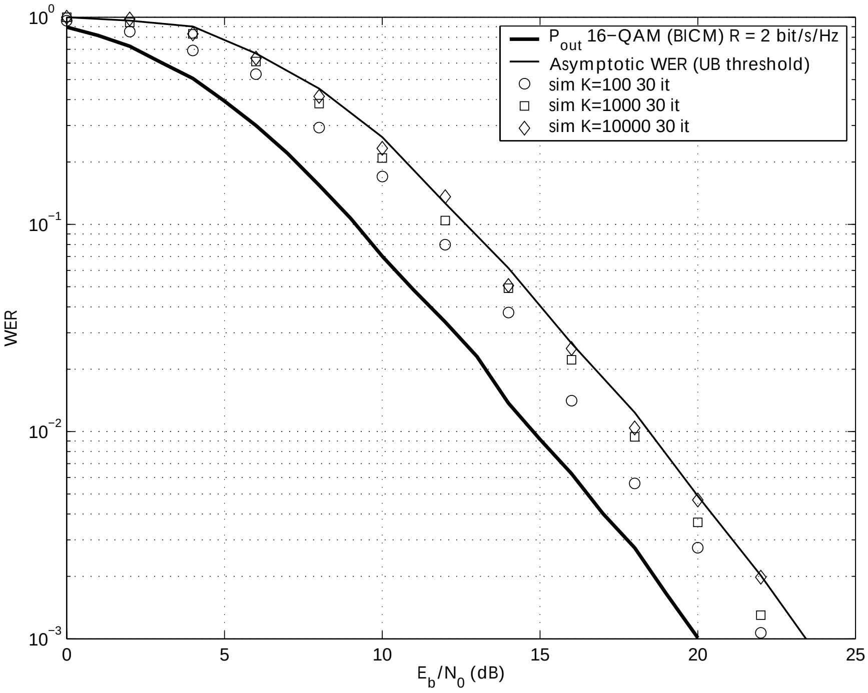


Figure 3.11: Asymptotic error probability (3.29) for a rate $R = 2$ RBA code mapped over $N_B = 2$ fading blocks with 16-QAM and Gray mapping (BICM) and corresponding BP decoding simulation with 30 iterations for $K = 100, 1000$ and 10000 information bits per codeword.

We conclude this section by pointing out an interesting fact that follows as a consequence of asymptotic weak goodness and allows the accurate WER evaluation of codes with given block length by using weight enumerators of codes in the same ensemble but with much smaller block length. In practice, computing the weight enumerator of shorter codes is in general much simpler and requires much less computational effort. This observation is illustrated by the following example. Figure 3.12 the WER and the BPL approximation for an RBA code of rate $R = 1/4$ mapped over $N_B = 4$ fading blocks with $K = 100$ information bits per codeword. We also show the simulation of BP decoding with 10 iterations. Again, we observe that the BPL approximation yields very accurate results. Moreover, notice that truncating the PWEF to maximum product weigth $\Delta_p^{\max} = 10000$, yields a too optimistic results. On the other hand, notice that with a much shorter PWEF (computed for the RBA code with $K = 20$ information bits) we are able to approximate very accurately the WER of the longer code.

As a matter of fact, the PWEF of the short code contains much more information on the code behavior than the truncated PWEF of the long code. This is clearly illustrated by the PWEFs in Figures 3.13(a) and 3.13(b), showing the growth rate of the PWEF defined as

$$F_{\overline{\Delta}_p} \triangleq \frac{1}{L_B^{N_B}} \log A_{\overline{\Delta}_p} \tag{3.30}$$

as a function of the normalized product weight $\overline{\Delta}_p = \Delta_p / L_B^{N_B}$ for the RBAs of rate $1/4$, with 20 and 100 information bits (every mark corresponds to one pairwise error event with normalized product weight $\bar{\Delta}_p$). Truncation at $\Delta_p^{\max} = 10000$ corresponds to maximum normalized product 10^{-4}, which means that only the portion for $0 \leq \overline{\Delta}_p \leq 10^{-4}$ of the distribution of Figure 3.13(b) is taken into account in the BPL approximation using the truncated enumerator. This is clearly not sufficient to describe the RBA product weight enumerator, as opposed to the PWEF of the short code.

3.4.4 On code optimization

So far we have seen that the BCC coding structure yields weakly good codes suited to the block-fading channel. However, most of the shown examples were based on the simple RBA structure. It is then natural to ask whether more general BCCs can reduce significantly the gap from outage. In this section we show some examples of other BCCs that in some case improve upon the RBA of same rate. Figures 3.14 and 3.15 show the performance of BCCCs with binary rate $r = 1/4$, attaining full diversity, with BPSK and 16-QAM BICM (with Gray mapping) respectively for $N_B = 4$ fading blocks, for $K = 1024$ information bits per codeword and 40 iterations of BP decoding. The octal generators are described in the legend. We have also considered the 4 states accumulator

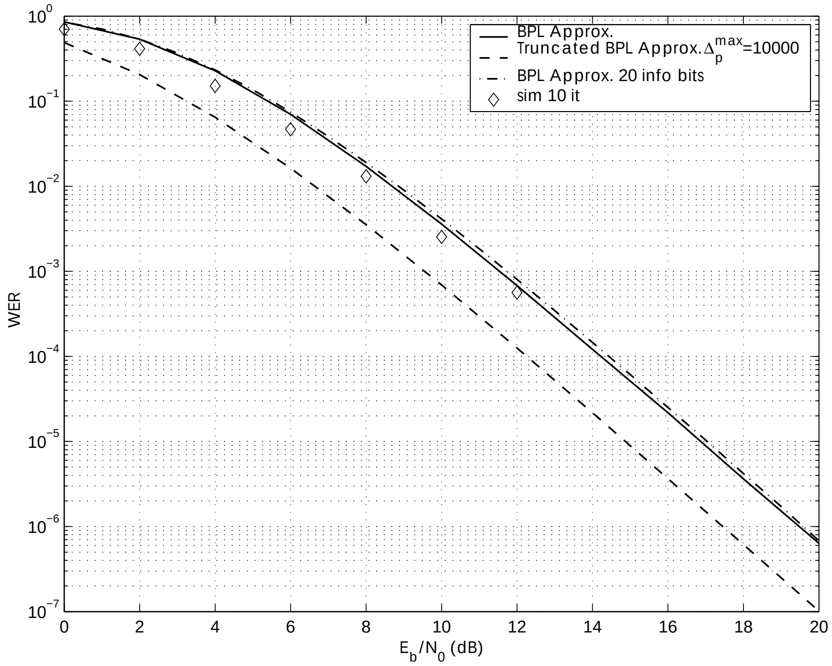
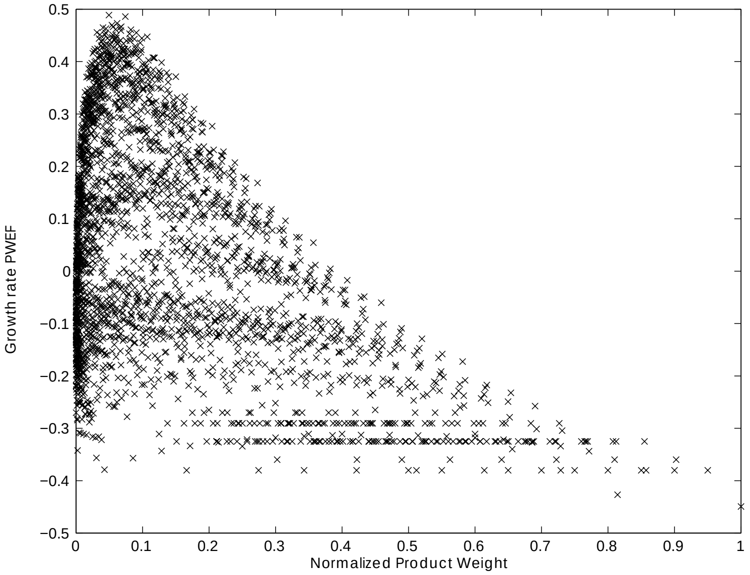


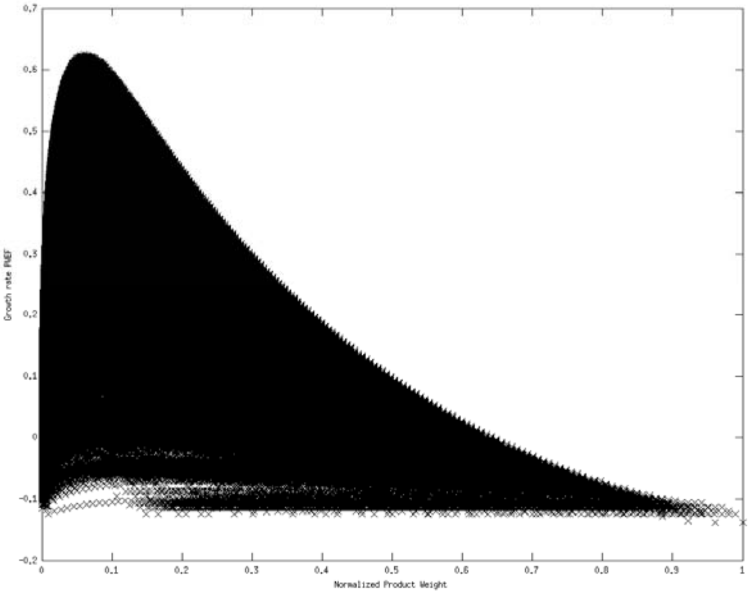
Figure 3.12: WER obtained by BP decoding simulation with 10 iterations and BPL approximations for RBA with rate $R = 1/4$ and 100 information bits per frame, over $N_B = 4$ fading blocks.

described in [56, Ch. 4] with generator $(1/7)_8$. We observe that in both cases the gap from outage is approximately of 1 dB. We notice from Figure 3.14 that using more complicated outer or inner codes does not yield any significant gain. Using the 4 states inner accumulator in an RBA scheme yields almost the same performance that the best BCCC.

From these examples, and an extensive code search not reported here for the sake of space limitation, it seems that some room is left for code optimization by searching over the component code generators. However, the improvements are not dramatic and in several cases do not justify the increasing in decoding complexity. Notice also that, in the block-fading channel, codes need to be *universal*, in the sense that they should offer very good performance for all fading realizations (i.e., a wide class of AWGN channels), and hence, a code which is best for one of such AWGN channels, may not be the best for another one. Therefore, we think that such a gap will be difficult to close, since one particular optimized code that may have very low error probability for some channel realizations, may fail to do so in other channels.



(a) RBA of rate $R = 1/4$ and $K = 20$ information bits



(b) RBA of rate $R = 1/4$ and $K = 100$ information bits

Figure 3.13: PWEF growth rate for RBA of rate $R = 1/4$ with 20 (a) and 100 (b) information bits per frame, over $N_B = 4$ blocks.

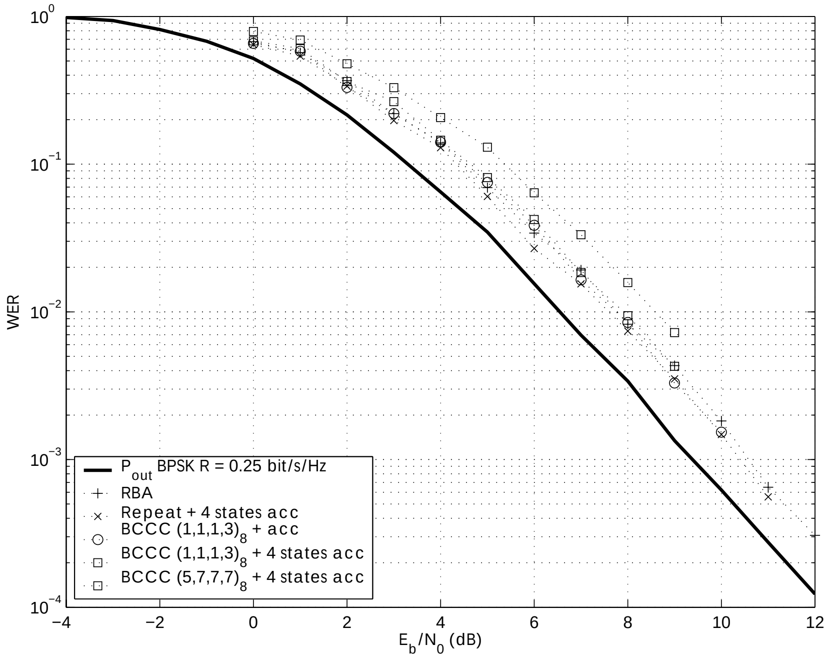


Figure 3.14: WER (simulation of BP decoding with 40 iterations) of several BCCs of rate $R = 1/4$ over BPSK, for $N_B = 4$ fading blocks with $K = 1024$ input bits per codeword.

3.5 Conclusions

In this chapter we have determined the SNR reliability function of codes over given finite signal sets over the block-fading channel. Random coding obtained by concatenating a linear binary random code to the modulator via a fixed one-to-one mapping achieve the same optimal SNR reliability function provided that the block length grows rapidly enough with SNR. Pragmatic BICM schemes under suboptimal BICM-ML decoding achieve the same random coding SNR exponent of their non-BICM counterparts (under optimal ML decoding).

Driven by these findings, we have proposed a general structure for random-like codes adapted to the block-fading channel, based on blockwise concatenation and on BICM (to attain large spectral efficiency). We provided some easily computable bounds and approximations to the WER of these codes under ML decoding and BICM-ML decoding. Remarkably, our approximations agree very well with the simulated performance of the iterative BP decoder.

We noticed that the proposed codes have WER almost independent of the block length (for large block length), that shows a fixed SNR gap from outage probability. We introduced the concept of “weak goodness” for specific ensembles of codes having this behavior for large block length (and large SNR), and we provided sufficient conditions for weak goodness of specific code ensembles in

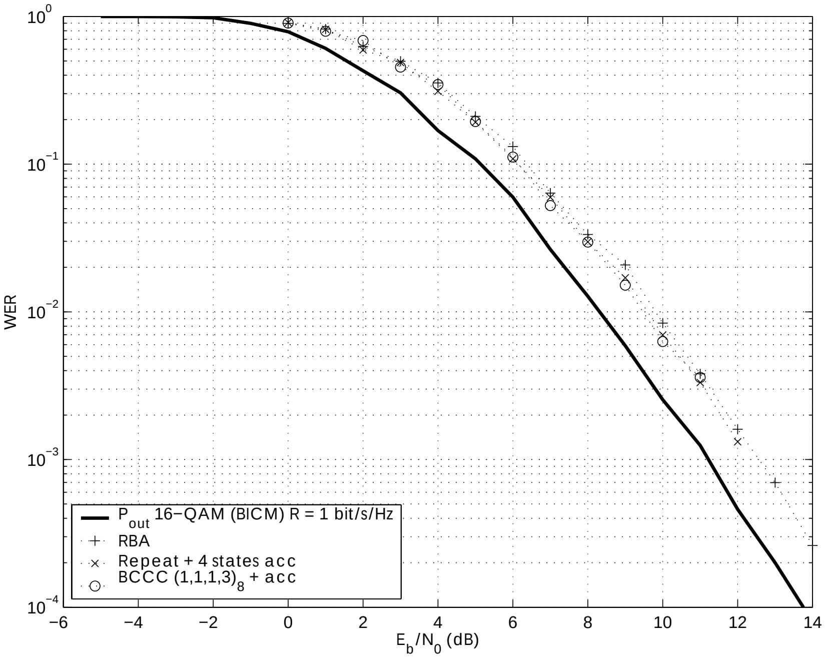


Figure 3.15: WER (simulation of BP decoding with 40 iterations) of several BCCs of rate $R = 1$ over 16-QAM (BICM with Gray mapping), for $N_B = 4$ fading blocks with $K = 1024$ input bits per codeword.

terms of their asymptotic multivariate weight enumerator growth function.

Finally, we showed through extensive computer search and simulation of the BP decoder that, while some improvement can be expected by careful optimization of the component codes, weakly good BCC ensembles have very similar behavior (Interestingly, a similar conclusion was reached in previous works on convolutional codes for the block-fading channel [24, 19, 25, 26]).

4

Coded Modulation in the MIMO Block-Fading Channel

In this chapter we study pragmatic space-time code design for MIMO channels. In particular, we will focus on the quasistatic case, i.e., $N_B = 1$. This is probably the simplest model that captures the essential features of non-ergodic MIMO channels. More general cases are briefly outlined in Section 4.2.1, and the application to OFDM systems is studied in Chapter 5. We briefly recall that the signal model can be written as

$$\mathbf{Y} = \sqrt{\rho}\,\mathbf{H}\mathbf{X} + \mathbf{Z} \tag{4.1}$$

where $\mathbf{Y} \in \mathbb{C}^{N_R \times L}$ is the received signal, $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_{N_T})^T$ is the transmitted signal drawn from the space-time code $\mathcal{S} \subseteq \mathbb{C}^{N_T \times L}$, $\mathbf{H} = (\mathbf{h}_1, \ldots, \mathbf{h}_{N_B}) \in \mathbb{C}^{N_R \times N_T}$ is the Rayleigh fading channel matrix with entries i.i.d. $\sim \mathcal{N}_{\mathbb{C}}(0,1)$, and $\mathbf{Z} \in \mathbb{C}^{N_R \times L}$ is the noise matrix, with i.i.d. samples $\sim \mathcal{N}_{\mathbb{C}}(0,1)$. We assume that the symbols transmitted from one antenna have unit energy, i.e., $\mathbb{E}[|x|^2] = 1$. Thus, ρ is the SNR per transmit antenna. Let us also recall that the pairwise error probability under ML decoding can be expressed as [30],

$$P(\mathbf{X} \to \mathbf{X}') \leq G_c \rho^{-d_r N_R} \tag{4.2}$$

where G_c is usually referred as the coding gain and

$$d_R = \min_{\mathbf{X},\mathbf{X}' \in \mathcal{S}} \operatorname{rank}(\mathbf{X} - \mathbf{X}') \tag{4.3}$$

is the rank diversity of $\mathcal{S}$. Conventional space-time code design searches for full diversity space-time codes $\mathcal{S}$, i.e., $d_r = N_T$. Then, once full diversity is ensured, traditional space-time code design searches for the best possible coding gain. As anticipated in Chapter 2, a simple upper bound on the rank diversity can be obtained by assuming a genie-aided decoder that produces observables of the transmitted symbols of one antenna, assuming that the effect of symbols from all other antennas has been artificially removed. In this case, the MIMO channel (4.1) decomposes into a set of N_T non-interfering parallel channels, i.e., a SISO block-fading channel . This upper bound is then given by the block diversity d_B studied in Chapter 3 for the SISO block-fading channel.

The main focus of this chapter is on the impact of signal constellation expansion on the achievable diversity. Signal constellation expansion can be obtained either by increasing the size of the constellation on the complex plane or by using multidimensional linear mappings. We study two different pragmatic approaches to construct full-diversity space-time codes. In particular, we first review a pragmatic construction based on bit-interleaved coded modulation (BICM) [36], which relies on the algebraic properties of the underlying binary code to achieve diversity. Secondly, we consider the concatenation of a coded modulation scheme with an inner code that is linear in the field of complex numbers (linear dispersion (LD) code [57]). It is well known that LD codes offer poor performance for large block length, and some sort of concatenation is needed. In both cases, full-diversity space time codes of any desired spectral efficiency are constructed by suitably expanding the signal constellation. In the first case, we have constellation expansion in the complex plane (Ungerböck's style expansion [58]), while in the second, we have multidimensional expansion induced by the inner code. By means of message passing decoding, we compare both approaches and we show that, in general, the concatenated LD construction is always advantageous due to its higher design flexibility.

As mentioned also in Chapter 2, the performance of some of the considered pragmatic coding strategies relies very heavily on the ability of the decoder to emulate the genie-aided decoder with the lowest possible complexity. In particular, this chapter also studies several low-complexity message passing iterative decoding approaches and we will illustrate their corresponding performance vs. complexity trade-off by means of some simple examples.

4.1 Pragmatic Space-Time Codes

We consider natural space-time codes (NSTC) coupled with BICM as a pragmatic way to construct full-diversity space-time codes (see e.g. [59, 60, 14, 61, 62, 63]). We nickname such scheme BICM NSTC. Such codes are formally defined by a binary block code $\mathcal{C} \subseteq \mathbb{F}_2^N$ of length N and rate r and a spatial modulation function $\mathcal{F} : \mathcal{C} \to \mathcal{S} \subseteq \mathcal{X}^{N_T \times L}$, such that $\mathcal{F}(\mathbf{c}) = \mathbf{X}$, where $\mathcal{X} \subseteq \mathbb{C}$ is the complex signal constellation. We study the case where $\mathcal{F}$ is obtained as the concatenation of a block/antenna parsing function $\mathcal{P} : \mathbb{Z}_+ \to \mathbb{Z}_+^2$ such that $\mathcal{P}(n) = (t, \ell), 1 \leq n \leq N, 1 \leq t \leq N_T, 1 \leq \ell \leq LM$ that partitions a codeword $\mathbf{c} \in \mathcal{C}$ into sub-blocks, and blockwise BICM, where each sub-block is independently bit-interleaved and mapped over the signal set $\mathcal{X}$ according to a labeling rule $\mu : \mathbb{F}_2^M \to \mathcal{X}$, such that $\mu(b_1, \ldots, b_M) = x$, where $M = \log_2 |\mathcal{X}|$ (see Figure 4.1). In the remainder of the chapter we shall only consider Gray labeling rules, since they are more efficient in quasistatic channels [14]. This conclusion may be reversed in fully-interleaved channels [63]. In this case, $N = N_T LM$. The transmission rate of the resulting space-time code is $R = rN_TM$ bit/s/Hz.

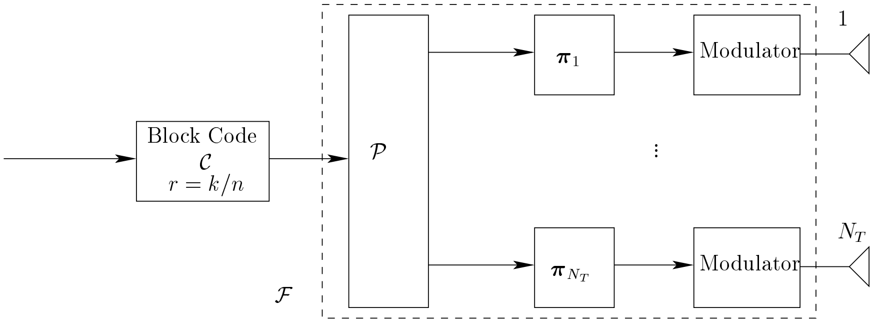


Figure 4.1: Transmission scheme of BICM NSTC.

BICM NSTCs are designed assuming that a genie aided decoder that produces observables of the transmitted symbols of one antenna, assuming that symbols from all other antennas are known[1]. In this way, the channel decomposes into an equivalent set of N_T non-interfering parallel channels (SISO block-fading channel with N_T blocks),

$$\mathbf{r}_t = \sqrt{\rho}|\mathbf{h}_t|\mathbf{x}_t + \boldsymbol{\nu}_t \ , \quad 1 \leq t \leq N_T, \tag{4.4}$$

where $\mathbf{r}_t$ is the received sequence of symbols corresponding to antenna t, $\mathbf{h}_t$ is the t-th column of $\mathbf{H}$, $\mathbf{x}_t$ is the transmitted sequence of symbols over antenna t and $\boldsymbol{\nu}_t$ is the corresponding sequence of noise samples $\sim \mathcal{N}_{\mathbb{C}}(0, 1)$.

We define the *block diversity* of a space-time code $\mathcal{S}$ as the blockwise Hamming distance,

$$d_B = \min_{\mathbf{X}, \mathbf{X}' \in \mathcal{S}} |\{t \in [1, \ldots, N_T] \ : \ \mathbf{x}_t - \mathbf{x}'_t \neq 0\}|, \tag{4.5}$$

i.e., the minimum number of nonzero rows of $\mathbf{X} - \mathbf{X}'$. Then, with a genie aided decoder, BICM NSTCs *can* achieve diversity $d_B N_R$ (see Chapters 2 and 3). Notice that applying BICM within a block preserves the block diversity of the binary code, since the binary labeling rule μ is a bijective correspondence. As we have seen in Chapter 3, the block diversity is upperbounded by the Singleton bound, with N_T blocks. In general MIMO block-fading channels with N_B temporal fading blocks (see (2.23)), the number of independent parallel channels is $N_B N_T$ [14]. Consequently, with BICM NSTCs we will search for codes maximizing d_{SB}, i.e., satisfying the Singleton bound with equality.

---

[1] The reader will notice the analogy with the case of decision feedback equalization for frequency selective channels, where correct feedback is assumed to design the equalizer filters.

On the other hand, the evaluation of the rank diversity of BICM NSTCs can be a very involved task especially for constellations with $M > 1$. For binary BICM NSTCs, it is however possible to verify through the stacking construction theorem whether a BICM NSTC has full rank diversity or not. In particular consider the following:

Theorem 3 (Stacking construction [35]) *Let* $\mathbf{G}_1, \dots, \mathbf{G}_{N_T}$ *be binary matrices* $\in \mathbb{F}_2^{K\times N}$*, and consider the binary linear code of rate* $K/(N_T N)$ *generated by*

$$\mathbf{G} = [\mathbf{G}_1, \mathbf{G}_2, \dots, \mathbf{G}_{N_T}].$$

Let the code words $\mathbf{c} = \mathbf{bG}$ *of* $\mathcal{C}$*, where* $\mathbf{b} \in \mathbb{F}_2^K$*, be parsed as*

$$\mathbf{C} = \mathcal{F}(\mathbf{c}) = \begin{bmatrix} \mathbf{bG}_1 \\ \mathbf{bG}_2 \\ \vdots \\ \mathbf{bG}_{N_T} \end{bmatrix}$$

Then, if for all $a_1, \dots, a_{N_T} \in \mathbb{F}_2$ *non all zero, the* $K \times N$ *matrix*

$$\mathbf{M} = \bigoplus_{t=1}^{N_T} a_t \mathbf{G}_t$$

($\bigoplus$ *indicates addition over the binary field* $\mathbb{F}_2$*) has rank* K*, then the binary space-time code mapped over BPSK modulation obtained from* $\mathcal{C}$ *with the above parsing has full rank-diversity* N_T *(the condition is necessary and sufficient).*

Now let $\mathbf{G}_t^{\mathcal{C}}$ be the t-th generator matrix of $\mathcal{C}$. The generator matrices $\mathbf{G}'_t$ of BICM NSTCs can be easily obtained from $\mathbf{G}_t^{\mathcal{C}}$, by simply applying the permutation $\boldsymbol{\pi}_t$ to the columns of $\mathbf{G}_t^{\mathcal{C}}$, for $t = 1, \dots, N_T$. We can now apply the stacking construction theorem with the generator matrices of the BICM NSTC $\mathbf{G}'_t$, in order to check for its a priori diversity performance under ML decoding. Notice that for BICM NSTCs, as shown in [64], we have that $d_r \leq d_B \leq N_T$.

4.2 Pragmatic Concatenated Space-Time Codes

In this section we consider the case where the codewords $\mathbf{X}$ of the space-time code $\mathcal{S}$ are obtained from the concatenation of an outer coded modulation scheme $\mathcal{C}^O \subseteq \mathcal{X}^n$ of rate r_O and length n with an inner LD code (see Figure 4.2).

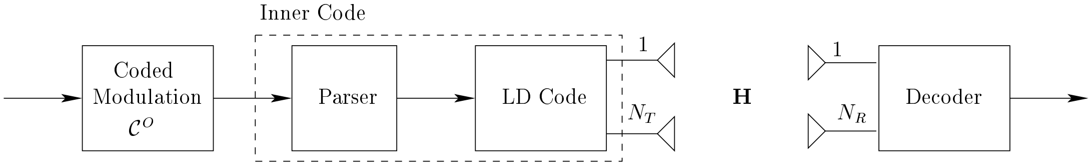


Figure 4.2: Transmission scheme of pragmatic concatenated space-time code.

The inner code is formed by a parser $\mathcal{P}$, that partitions the codewords $\mathbf{c} \in \mathcal{C}^O$ into sub-blocks $\mathbf{c}[j] = [c_1[j], \dots, c_Q[j]]$, $j = 1, \dots, J$ of length Q, with $J = n/Q$ and by a LD space-time modulation function $\mathcal{F}$ defined by,

$$\mathbf{S}[j] = \mathcal{F}(\mathbf{c}[j]) = \sum_{q=1}^{Q}(c_q[j]\mathbf{G}_q), \tag{4.6}$$

where $\mathbf{G}_q \in \mathbb{C}^{N_T \times T}$ are the LD code generator matrices. Finally, the overall space-time codeword is given by $\mathbf{X} = [\mathbf{S}[1] \dots \mathbf{S}[J]]$. Then,

$$\mathbf{Y}[j] = \mathbf{H}\mathbf{S}[j] + \mathbf{Z}[j], \ j = 1, \dots, J. \tag{4.7}$$

Equation (A.1) can be rewritten as a virtual MIMO channel with Q inputs and $N_R^v = N_R T$ outputs as,

$$\mathbf{y}[j] = \mathcal{H}\mathbf{c}[j] + \mathbf{z}[j], \ \ j = 1, \dots, J \tag{4.8}$$

where $\mathcal{H} \in \mathbb{C}^{N_R^v \times Q}$ is the equivalent channel matrix given by,

$$\mathcal{H} = [I_T \otimes \mathbf{H}]\mathbf{G} \tag{4.9}$$

where $\otimes$ is the Kronecker product, $\mathbf{G} \in \mathbb{C}^{N_T T \times Q}$ is the suitably reformatted generator matrix of the LD code, $\mathbf{y}[k] = \mathrm{vec}(\mathbf{Y}[k])$, $\mathbf{z}[k] = \mathrm{vec}(\mathbf{Z}[k])$, $N_R^v = N_R T$ is the number of virtual receive antennas, and $\mathrm{vec}(\mathbf{A}) = [\mathbf{a}_1^T \dots \mathbf{a}_l^T]^T$ for a matrix $\mathbf{A} = [\mathbf{a}_1 \dots \mathbf{a}_l]$. Will refer to Q as the number of virtual transmit antennas.

In order to obtain a dual scheme to compare with, in this work, we consider that $\mathcal{C}^O$ is obtained by BICM, i.e, a binary code $\mathcal{C} \in \mathbb{F}_2^N$ of rate r whose bit-interleaved codewords are mapped onto the signal set $\mathcal{X}$ according to the binary labeling rule $\mu : \mathbb{F}_2^M \to \mathcal{X}$ [36]. As inner LD code, we use the threaded algebraic space-time (TAST) constellations of [65]. We nickname such a transmission scheme as BICM TAST (see Figure 4.3).

The aforementioned algebraic space-time codes rely on the threaded layering first introduced in [66] for threaded space-time codes (TSTC). In TSTC, a number of component encoders (or layers) N_L and the component codewords $\mathbf{c}_\ell \in \mathcal{C}_\ell$, $\ell =$

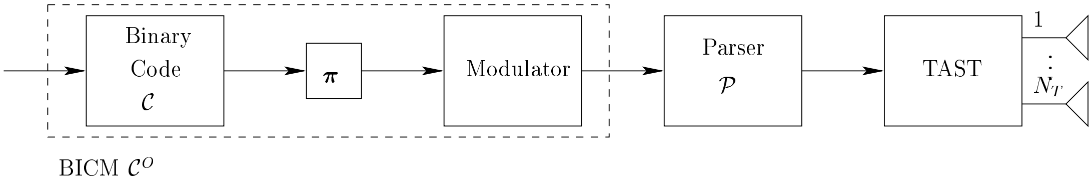


Figure 4.3: Transmission scheme of BICM TAST.

$1, \ldots, N_L$ are mapped onto the array $\mathcal{T}_{N_T,N_L,N_L}$ following the (layer, antenna, time) indexing triplet

$$\mathcal{P}(\ell, t) = (\ell, |t + \ell - 1|_{N_T}, t), \ \ell = 1, \ldots, N_L, \ 1 \leq t \leq T \tag{4.10}$$

thus having full spatial and temporal spans. Modulo-k operation is denoted by $|.|_k$. TAST threading (4.10) is shown in Figure 4.4 for $N_T = 4$, $N_L = 4$ and $T = 4$. The encoding functions corresponding to $\mathcal{C}_\ell, \ \ell = 1, \ldots, N_L$ are $\gamma_\ell = \phi_\ell \mathbf{M}$, where $\mathbf{M}$ is a rate one full-diversity linear algebraic rotation and $\phi_\ell, \ \ell = 1, \ldots, N_L$ scalar complex coefficients chosen to be Diophantine numbers that ensure that TAST codes achieve full diversity with ML decoding (see [65] for details). In particular, the scaling coefficients ϕ_ℓ ensure that *all* the difference codeword matrices have full rank. If we set all $\phi_\ell = 1, \ \ell = 1, \ldots, N_L$, most of the pairwise error events will still be full rank, but there will be some that are not, as shown in [65]. Then some difference in the WER performance is observed only at very large SNR, which is the region dominated by rank-deficient error events. Therefore, it is expected that using an inner coded modulation, this effect will be negligible, since the inner code will remove some of the rank deficient error events.

In TAST codes, $N_L = \min(N_T, N_R)$, $T = N_T$, $Q = N_L N_T$, and $\phi_\ell, \ \ell = 1, \ldots, N_L$ are chosen such that $|\phi_\ell| = 1, \ \ell = 1, \ldots, N_L$. The transmission rate of the resulting space-time code is $R = r N_L M$ bit/s/Hz. Notice that following this *divide and conquer* design, full diversity is always guaranteed by the inner code, while coding gain is left to the outer coded modulation.

4.2.1 Extension to General Block-Fading Channels

The approaches described above can also be efficiently used in a block-fading channel with N_B blocks as the one described by (2.23). In particular, for the BICM TAST, it suffices to find a rotation matrix of dimension $N_T N_B$ and parse the different threads as shown in Figure 4.5. In this case, the equivalent channel $\mathcal{H}$ will have its dimensions augmented by a factor N_B [65].

In the case of BICM NSTC, under the assumption of a genie-aided decoder, any parsing that cyclically maps the output of $\mathcal{C}$ over $N_T N_B$ fading blocks would do. However, in practice, symbols transmitted at the same time instant do interfere. Then, the cyclic interleaver should be designed in the space and block

1	4	3	2
2	1	4	3
3	2	1	4
4	3	2	1

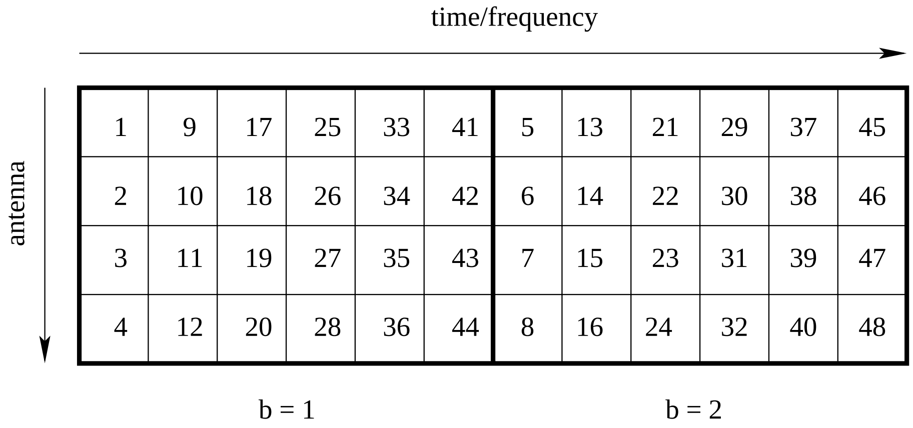
Figure 4.4: TAST threading (4.10) for $N_T = 4$, $N_L = 4$ and $T = 4$. The numbers indicate the thread index.

1	4	3	2	1	4	3	2
2	1	4	3	2	1	4	3
3	2	1	4	3	2	1	4
4	3	2	1	4	3	2	1

$b = 1$ $b = 2$

Figure 4.5: TAST threading for $N_T = 4$, $N_B = 2$, $N_L = 4$ and $T = 4$. The numbers indicate the thread index.

dimensions. One natural option is a parsing rule that first maps adjacent coded symbols in one hyper-trellis step over the antennas and then over the blocks. This parsing rule is illustrated in Figure 4.6. Notice that this parsing rule can be coupled with efficient block interleavers. This is however not the only possibility. In fact, one can perform a joint parsing and interleaver design such that binary symbols from adjacent trellis steps are not transmitted at the same time-instant [63].



1	9	17	25	33	41	5	13	21	29	37	45
2	10	18	26	34	42	6	14	22	30	38	46
3	11	19	27	35	43	7	15	23	31	39	47
4	12	20	28	36	44	8	16	24	32	40	48

Figure 4.6: Parsing rule for $N_T = 4$, $N_B = 2$, $N = 48$ and $M = 1$.

4.3 Message Passing Decoding

In this section, we shall study several low-complexity receiver architectures that, under some conditions, approach the performance of the genie-aided decoder.

Because of the pseudo-random bit interleaver present in both schemes, ML decoding of BICM NSTC or BICM TAST is of unaffordable complexity. We shall see that applying the belief propagation (BP) or sum-product algorithm [44] to the space-time code dependency graph yields to several receivers that approximate the optimal maximum a posteriori (MAP) detection rule by iteratively exchanging probability messages over the graph. Recall that for graphs with cycles, as for the codes studied here, BP only yields an approximation of MAP [44]. We denote by $\mathbf{b} = (b_1, \dots, b_K)$ the information bit vector and the proportionality symbol $\propto$ indicates that the term in the right hand side (RHS) is defined up to a multiplicative constant.

In the case of BICM NSTC[2], the optimal MAP detection rule for a given information bit b_i for a particular channel realization can be expressed as,

$$\hat{b}_i = \arg\max_{b \in \mathbb{F}_2} \mathrm{APP}_i(b) \tag{4.11}$$

---

[2] The case of BICM TAST is completely analogous, and it suffices to replace N_T by Q, L by J and $\mathbf{H}$ by $\mathcal{H}$.

where

$$\mathrm{APP}_i(b) = \sum_{\substack{\mathbf{b}\in\mathbb{F}_2^K \\ b_i=b}} P(\mathbf{b}|\mathbf{Y},\mathbf{H}) \tag{4.12}$$

denotes the a posteriori probability (APP) on the i-th information bit computed as the marginalization of the joint a posteriori pmf of the information bits $P(\mathbf{b}|\mathbf{Y},\mathbf{H})$. A general method of approximating (4.12) consists of applying the BP algorithm to the factor graph of the space-time code. A simple example of a factor graph for BICM NSTCs with a $r = 1/2$ binary code $\mathcal{C}$, $N_T = 2$, $M = 2$ (i.e., QPSK modulation), $N = 8$ and $L = 2$ is given in Figure 4.7.

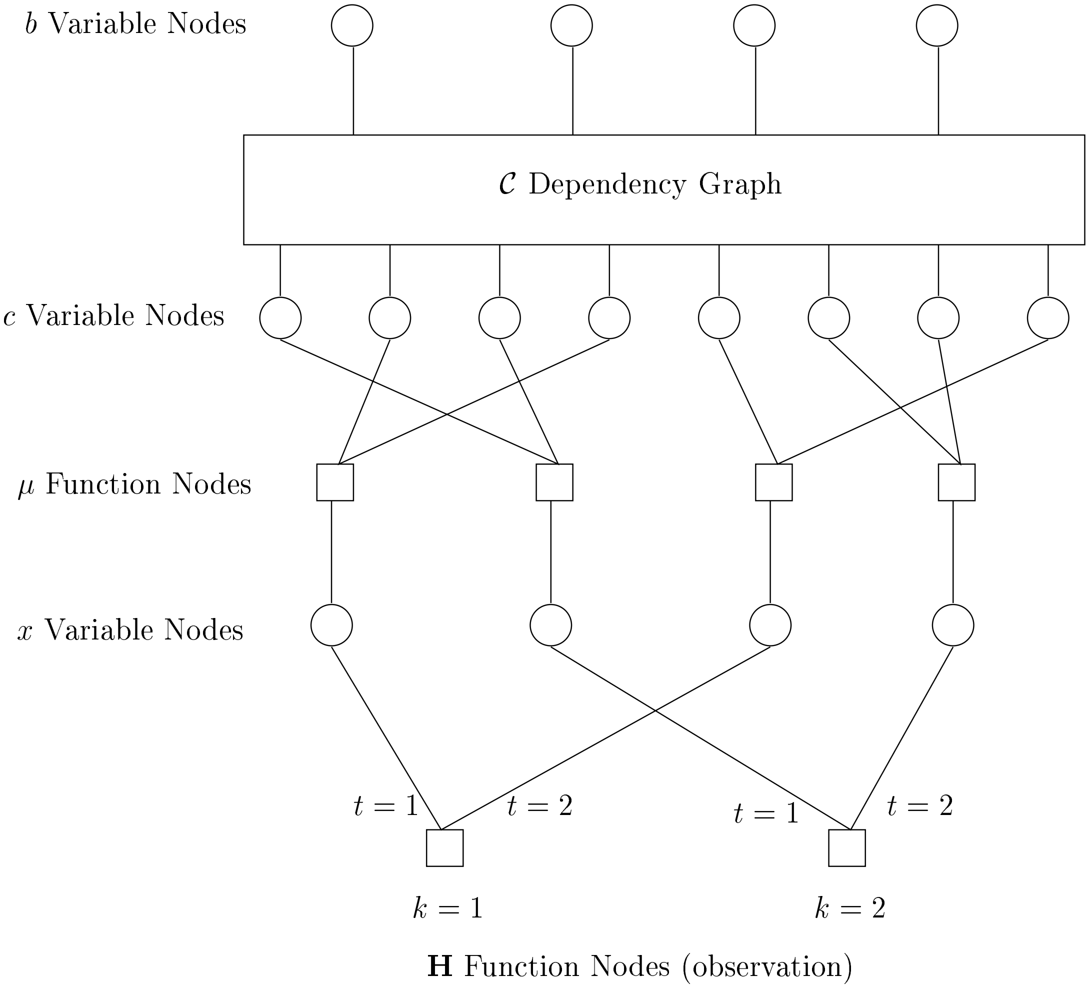

b Variable Nodes
$\mathcal{C}$ Dependency Graph
c Variable Nodes
μ Function Nodes
x Variable Nodes
$t = 1$
$t = 2$
$t = 1$
$t = 2$
$k = 1$
$k = 2$
$\mathbf{H}$ Function Nodes (observation)

Figure 4.7: Dependency graph of BICM NSTC for $r = 1/2$, $N_T = 2$, $M = 2$, $N = 8$ and $L = 2$.

4.3.1 Belief Propagation

By applying the sum-product rule of the BP algorithm to the function nodes μ, in analogy to the case of iterative multiuser receivers for CDMA [67], we obtain that the probability message that is passed from a μ node to a c node, along an edge e at time k, corresponding to the m-th coded bit of the symbol x transmitted over antenna t is given by [44, 67],

$$P^{\mu\to c}_{m,t,k}(a) = \sum_{\mathbf{x}\in\mathcal{X}^{m,t}_a} p(\mathbf{y}_k|\mathbf{x},\mathbf{H}) \prod_{t'=1}^{N_T} \prod_{\substack{m'=1\\ m'\neq m, t'=t}}^{M} P^{c\to\mu}_{m',t',k}(a_{m'}), \tag{4.13}$$

for $1 \leq m \leq M, 1 \leq t \leq N_T, 1 \leq k \leq L$, where

$$p(\mathbf{y}_k|\mathbf{x},\mathbf{H}) \propto \exp(-|\mathbf{y}_k - \sqrt{\rho}\,\mathbf{H}\mathbf{x}|^2), \tag{4.14}$$

is the conditional pdf of the received signal $\mathbf{y}_k = \mathbf{H}\mathbf{x}_k + \mathbf{z}_k$ given the input signal $\mathbf{x}$ and the channel $\mathbf{H}$, $\mathcal{X}^{m,t}_a$ is the set of all N_T-dimensional symbols with the m-th bit of the binary label corresponding to the symbol transmitted over antenna t equal to $a \in \mathbb{F}_2$, which is a dummy binary variable. The corresponding expression for the log-likelihood ratio (LLR) message is given by,

$$\mathcal{L}^{\mu\to c}_{m,t,k}(a) = \log \frac{P^{\mu\to c}_{m,t,k}(a=0)}{P^{\mu\to c}_{m,t,k}(a=1)}. \tag{4.15}$$

Performing the same computation at the code constraint function nodes of the dependency graph of $\mathcal{C}$, the probability message passed from a c node to a μ node along an edge e at time k, corresponding to the m-th coded bit of the symbol x transmitted over antenna t is given by,

$$P^{c\to\mu}_{m,t,k}(a) = \sum_{\substack{\mathbf{c}\in\mathcal{C}\\ c_{m,t,k}=a}} \prod_{j\neq k} P^{\mu\to c}_{m,t,j}(a_j). \tag{4.16}$$

Therefore, as illustrated in the block diagrams shown in Figures 4.8 and 4.9, for BICM NSTC and BICM TAST respectively, the BP algorithm reduces the receiver to a MAP soft-input soft-output bitwise demodulator and a MAP soft-input soft-output decoder of $\mathcal{C}$, that exchange extrinsic information probability messages through the iterations [68]. Messages defined by (4.16) are the extrinsic pmf of the coded bits at the output of the soft-input soft-output decoder of $\mathcal{C}$. When $\mathcal{C}$ is a trellis code, the messages (4.16) are efficiently computed by the forward-backward algorithm (BCJR) with linear complexity in N [49]. When $\mathcal{C}$ is itself a concatenated or LDPC code, further iterations are needed to approximate the MAP solution.

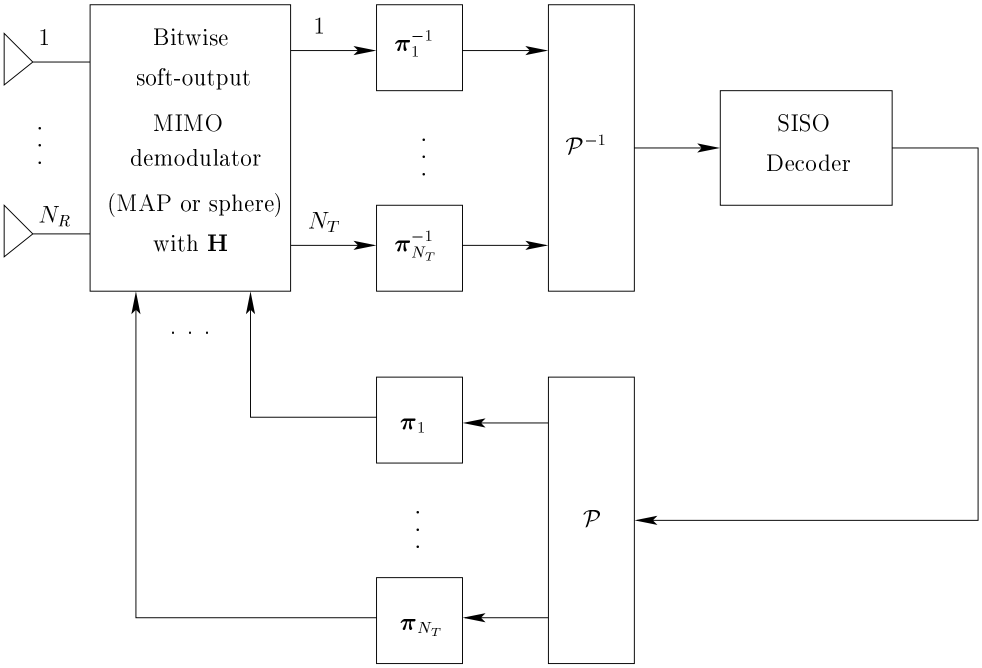

1
N_R
Bitwise soft-output MIMO demodulator (MAP or sphere) with $\mathbf{H}$
1
N_T
$\boldsymbol{\pi}_1^{-1}$
$\boldsymbol{\pi}_{N_T}^{-1}$
$\mathcal{P}^{-1}$
SISO Decoder
$\boldsymbol{\pi}_1$
$\boldsymbol{\pi}_{N_T}$
$\mathcal{P}$

Figure 4.8: Iterative decoder with the sum-product algorithm for BICM NSTC.

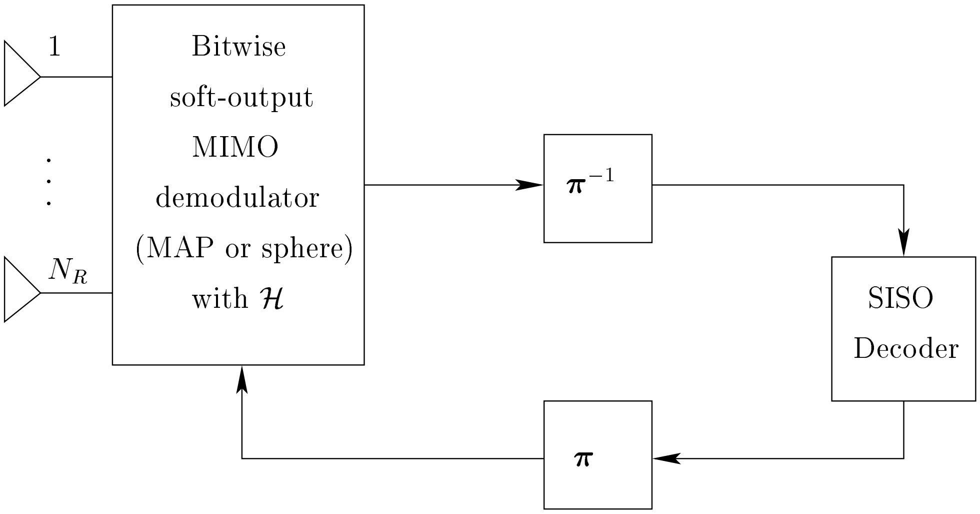

1
N_R
Bitwise soft-output MIMO demodulator (MAP or sphere) with $\mathcal{H}$
$\boldsymbol{\pi}^{-1}$
SISO Decoder
$\boldsymbol{\pi}$

Figure 4.9: Iterative decoder with the sum-product algorithm for BICM TAST.

Notice that, $|\mathcal{X}_a^{m,t}| = 2^{N_T M - 1}$, and therefore, for large N_T and/or large M, bitwise MIMO demodulation becomes of unaffordable complexity, since an exhaustive enumeration of all N_T-dimensional constellation points has to be performed in order to compute (4.13). We will use it in some cases where we can still compute it as a performance benchmark.

4.3.2 Soft-Input Soft-Output Sphere Decoding

We have seen that exact BP decoding induces exhaustive enumeration of the points in $\mathcal{X}_a^{m,t}$ in order to compute (4.13), and this yields an exponentially complex decoding algorithm in N_T and M. In order to avoid exhaustive enumeration, one can only enumerate the N_T-dimensional points that bring a relevant contribution to (4.13). This task is efficiently accomplished by the so-called soft-input soft-output sphere decoders.

Sphere decoding, classically refers to algorithms that perform low-complexity search for the closest lattice point in lattice-based code constructions [69, 70, 71, 72]. In particular, such algorithms use the lattice code structure to find the ML solution without exhaustively enumerating all the possible lattice points. In particular, they chose the ML solution from a set of lattice points that lie in a sphere of radius C centered at the received point. Here we are mostly concerned with providing decision statistics to the decoder of $\mathcal{C}$, and therefore, original sphere decoders have to be modified in order to output a list of constellation points rather that the ML solution only. In this work, we shall use the low-complexity approach described by Boutros et al. in [73]. In this work, rather than enumerating all constellation the points inside a sphere centered at the received point [74, 75], the authors propose to enumerate all constellation points inside a sphere centered at the *uncoded* ML solution. This stabilizes the list size and consequently produces a soft-output of higher quality, specially for the noise realizations that bring the received point outside the constellation.

In particular, for the quasistatic channel, the low-complexity algorithm proposed in [73] can be summarized as follows,

1. `for` each channel realization $\mathbf{H}$ `compute` $\mathbf{Q}$ and $\mathbf{R}$ such that $\mathbf{B} = \mathbf{QR}$
2. `compute` the lattice list radius R_L
3. use the Pohst method to enumerate all lattice points inside a sphere of radius R_L centered at the $\mathbf{x} = \mathbf{0}$ lattice point and put them in the list L_L
4. sort the lattice list L_L with increasing Euclidean distance to the $\mathbf{x} = \mathbf{0}$ lattice point
5. `for` each symbol period $1 \leq k \leq L$ `compute` the *uncoded* ML point $\mathbf{x}_k^{\mathrm{ML}}$ using Schnorr-Euchner enumeration

6. translate the lattice list to the $\mathbf{x}_k^{\mathrm{ML}}$ point and evaluate the constellation points that fall inside, i.e., $\mathsf{L}_{\mathcal{X}} = (\mathsf{L}_L + \mathbf{x}_k^{\mathrm{ML}}) \cap \mathcal{X}_{\mathrm{PAM}}^{2N_T}$

where, assuming transmission that $\mathcal{X}$ is a Gray-mapped QAM signal, such that real and imaginary parts are independent Gray mapped PAM constellations $\mathcal{X}_{\mathrm{PAM}}$,

$$\mathbf{B} = \begin{pmatrix} \mathrm{Re}\{\mathbf{H}\} & -\mathrm{Im}\{\mathbf{H}\} \\ \mathrm{Im}\{\mathbf{H}\} & \mathrm{Re}\{\mathbf{H}\} \end{pmatrix} \tag{4.17}$$

is the generator matrix of the *real* lattice Λ induced by the MIMO channel. We compute the lattice list radius R_L in order to have approximately N_L lattice points in the sphere. This can be accomplished by choosing R_L to be

$$\mathsf{R}_L = \left(\frac{N_L \, |\det \mathbf{B}| \, N_T!}{\pi^{N_T}} \right)^{\frac{1}{2N_T}} . \tag{4.18}$$

As recognized in [76], significant complexity reduction in finding the uncoded ML point $\mathbf{x}_k^{\mathrm{ML}}$ can be achieved if we perform the detection in a permuted order. In particular, as done in [77] we will use the method proposed in [78] since the ordering can be achieved with very low complexity by appropriately modifying the Gram-Schmidt algorithm in the QR decomposition. Remark that this is not the optimal ordering. In particular, this optimal ordering is given by the V-BLAST optimal ordering [79, 76].

4.3.3 Lower-Complexity Approximations

Motivated by the high complexity of BP and SD decoding, we study the impact of lower complexity receiver architectures on the system performance of BICM TAST and BICM NSTC. In particular, we consider interference cancellation (IC) and linear filtering, for which the message passed from a μ node to a c node, along an edge e at time k, corresponding to the m-th coded bit of the symbol x transmitted over antenna t is now given by

$$P_{m,t,k}^{\mu \to c}(a) = \sum_{x \in \mathcal{X}_a^m} p(z_{t,k} | x, \mathbf{H}) \prod_{\substack{m'=1 \\ m' \neq m}}^{M} P_{m',t,k}^{c \to \mu}(a_{m'}). \tag{4.19}$$

where $\mathcal{X}_a^m$ is now the set of complex signal constellation points for which the m-th position of its binary label is equal to a, $z_{t,k}$ is the output at symbol time k of the front-end linear filter $\mathbf{f}_t$ of antenna t,

$$z_{t,k} = \mathbf{f}_t^H \left(\mathbf{y}_k - \sqrt{\rho} \sum_{t' \neq t}^{N_T} \mathbf{h}_{t'} \hat{x}_{t',k} \right) \tag{4.20}$$

where (dropping antenna and time indexes for simplicity),

$$\hat{x} = \mathbb{E}[x \,|\, \mathrm{EXT}] = \sum_{x \in \mathcal{X}} x \prod_{m=1}^{M} P_{m,t,k}^{c \to \mu}(\mu_m^{-1}(x)), \tag{4.21}$$

is the minimum mean-square error estimate (conditional mean) of the symbol x given the extrinsic information (briefly denoted in (4.21) by EXT) relative to the bits in the label of x, and $\mu_m^{-1}(x) = a$ is a shortcut notation to denote that the m-th position of the bit label of x is equal to a. Figure 4.10 illustrates the interference cancellation and filtering steps represented by equation (4.20) for detecting the symbols of the t-th antenna.

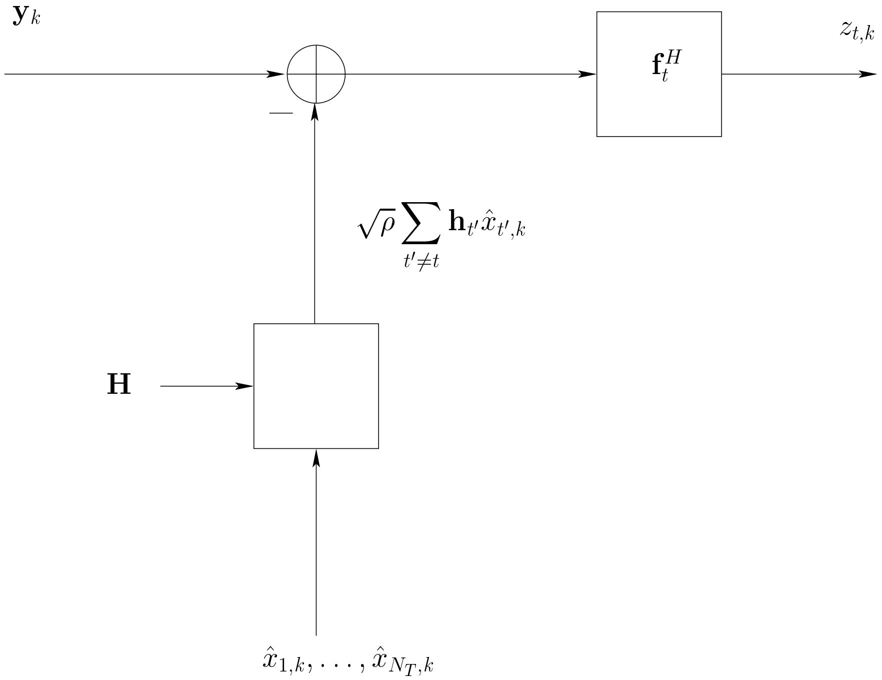

$\mathbf{y}_k$
$-$
$\mathbf{f}_t^H$
$z_{t,k}$
$\sqrt{\rho}\sum_{t' \neq t} \mathbf{h}_{t'} \hat{x}_{t',k}$
$\mathbf{H}$
$\hat{x}_{1,k}, \ldots, \hat{x}_{N_T,k}$

Figure 4.10: Interference cancellation and filtering.

Given the analogy with CDMA multiuser detection, we consider first the *Unbiased Minimum Mean Squared Error* (MMSE) filter as IC linear front-end $\mathbf{f}_t$, which minimizes the MSE $\mathbb{E}[|x_t - z_t|^2]$, and is given by

$$\mathbf{f}_t^{\mathrm{MMSE}} = \alpha_t \sqrt{\rho}\, \mathbf{R}^{-1} \mathbf{h}_t, \tag{4.22}$$

where

$$\alpha_t = (\rho\, \mathbf{h}_t^H \mathbf{R}^{-1} \mathbf{h}_t)^{-1} \tag{4.23}$$

is the normalization constant that induces unbiasedness,

$$\mathbf{R} = \mathbf{I} + \rho \sum_{t=1}^{N_T} \mathbf{h}_t \mathbf{h}_t^H v_t \tag{4.24}$$

is the covariance matrix of the input signal to the filter, and

$$v_t = E[|x_t - \hat{x}_t|^2] \tag{4.25}$$

is the variance of the residual interference at antenna t (see [67, 80] and references therein).

According to (4.19), n order to compute the messages $P^{\mu\to c}_{m,t,k}(a)$, we need $p(z_{t,k}|x, \mathbf{H})$. This corresponds to the conditional pdf of the signal at the output of the filter, once the interference has been canceled (see Figure 4.10). It is reasonable to approximate the residual interference plus noise at the filter output as Gaussian, and thus we write that

$$p(z_{t,k}|x, \mathbf{H}) \approx \frac{1}{\pi\sigma_t^2} \exp\left(-\frac{1}{\sigma_t^2}|z_{t,k} - x|^2\right) \tag{4.26}$$

where variance of the residual interference plus noise can be computed as

$$\sigma_t^2 = |\mathbf{f}_t|^2 + \rho \sum_{t' \neq t} |\mathbf{f}_t^H \mathbf{h}_{t'}|^2 v_{t'}. \tag{4.27}$$

In a practical implementation (and in the numerical computations illustrated in section 4.4) we estimate v_t as

$$v_t \approx 1 - \frac{1}{L}\sum_{k=1}^{L} |\hat{x}_{t,k}|^2. \tag{4.28}$$

Notice that $\mathbf{f}_t^{\mathrm{MMSE}}$ has to be computed once per virtual transmit antenna and iteration. The proposed algorithm differs from that proposed in [66] in that the latter has to be computed once per symbol interval, transmit antenna and iteration.

If we wish to further reduce the complexity, we can use the *Matched Filter* (MF) as the linear interface, for which [81],

$$\mathbf{f}_t^{\mathrm{MF}} = \frac{\mathbf{h}_t}{\sqrt{\rho}|\mathbf{h}_t|^2} \tag{4.29}$$

which performs maximal ratio combining (MRC) of the receive antennas and does not take into account the noise plus interference. It has to be computed only once per channel realization, since it does not depend on the iterations, which reduces significantly the computational burden. The residual interference plus noise is also approximated as Gaussian [81].

Figures 4.11 and 4.12 show the block-diagram of such IC receivers for BICM NSTC and BICM TAST respectively.

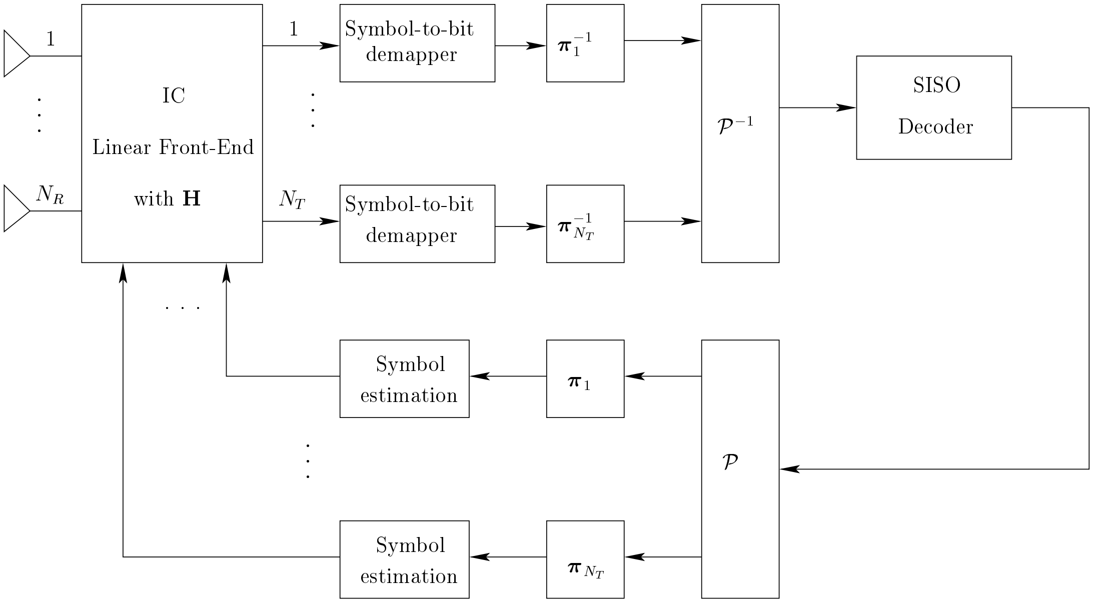

1
N_R
IC
Linear Front-End
with $\mathbf{H}$
1
N_T
Symbol-to-bit demapper
Symbol-to-bit demapper
$\boldsymbol{\pi}_1^{-1}$
$\boldsymbol{\pi}_{N_T}^{-1}$
$\mathcal{P}^{-1}$
SISO Decoder
Symbol estimation
Symbol estimation
$\boldsymbol{\pi}_1$
$\boldsymbol{\pi}_{N_T}$
$\mathcal{P}$

Figure 4.11: Iterative decoder with MMSE interference cancellation for BICM NSTC.

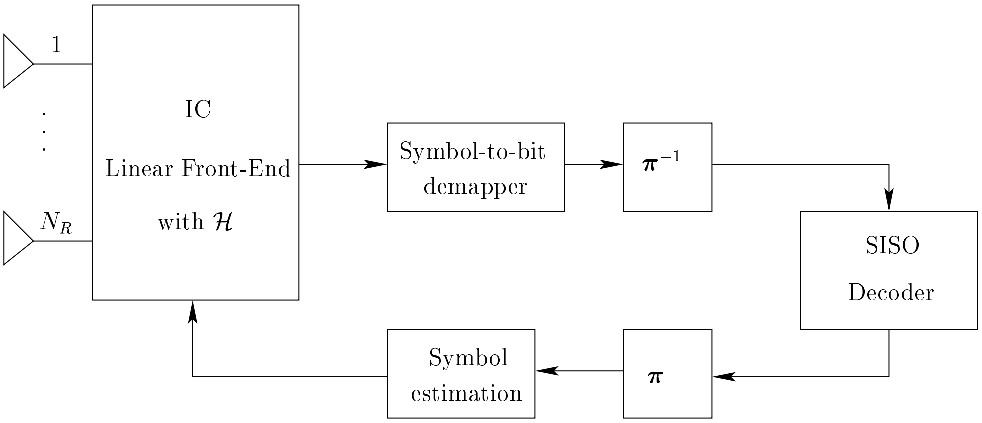

1
N_R
IC
Linear Front-End
with $\mathcal{H}$
Symbol-to-bit demapper
$\boldsymbol{\pi}^{-1}$
SISO Decoder
Symbol estimation
$\boldsymbol{\pi}$

Figure 4.12: Iterative decoder with MMSE interference cancellation for BICM TAST.

Remark 1 *A large number of front-end filters coupled with soft or hard interference cancellation can be incorporated in this framework. Such modifications import to the MIMO case well established structures for multiuser receivers in CDMA (see [82] for some recent structure using Kalman filters).*

Remark 2 *In order to further reduce the complexity for trellis codes, the BCJR module can be replaced by a weighted Viterbi decoder, in a similar way to what it is done in [81] in the context of multiuser detection for CDMA systems.*

Remark 3 *There exist several variations of the presented algorithm that estimate the interfering symbols using the a posteriori probabilities (APP) instead of the extrinsic. See e.g. [67] and references therein. It has been shown experimentally that such IC approaches, when they converge, they converge faster using the APPs rather than the EXT. These algorithms violate in principle the basic rules of belief propagation. Note however, that once the computation of the messages from μ function nodes to c variable nodes is not done according to belief propagation, but rather according to some approximation, it is not guaranteed anymore that propagating the EXT will yield better results than propagating the APPs.*

4.4 Numerical Examples

In this section we provide several numerical examples obtained by computer simulation that illustrate the relevance of the block diversity as design parameter for BICM NSTC, the effect of the constellation expansion on the achievable diversity of BICM NSTC and BICM TAST and the efficiency of the several decoders to approach the genie-aided decoder. For the sake of comparison, we include the outage probability curves with Gaussian inputs at the corresponding spectral efficiency. Unless otherwise specified we take frames of 128 information bits and 5 decoding iterations.

4.4.1 Block Diversity vs. Rank Diversity

Figure 4.13 reports the word error rate (WER) as a function of E_b/N_0 in a MIMO channel with $N_T = 4$ and $N_R = 4$, using BPSK modulation and the 4 states $(5,7,7,7)_8$ convolutional code of $r = 1/4$. The overall spectral efficiency is $R = 1$bit/s/Hz. Clearly, the block diversity of $\mathcal{C}$ is $d_B = 4$. In dashed-dotted line we show the WER for the NSTC with ML decoding. Recall that the NSTC array is constructed using identity permutations [35], and therefore ML decoding is possible using the Viterbi algorithm. Applying the stacking construction theorem yields that the NSTC code is rank deficient. We have also applied the theorem to BICM NSTC with a large number of randomly generated interleaver permutations, and none of them gave a full-rank code. However, as the curves in the figure show, in the WER region of interest full-diversity performance is achieved with the three suboptimal iterative receivers (exact BP, MMSE-IC and MF-IC) described in the previous section. This simple example serves to illustrate the key role that the block diversity plays to achieve full diversity in BICM NSTC with low-complexity receivers.

Remark again that NSTC and BICM NSTC are different space-time codes due to the presence of interleavers. Thus, it is obvious that using ML decoding of NSTCs as performance benchmark for iterative decoding of BICM NSTC is meaningless [83]. This is also well illustrated in this example.

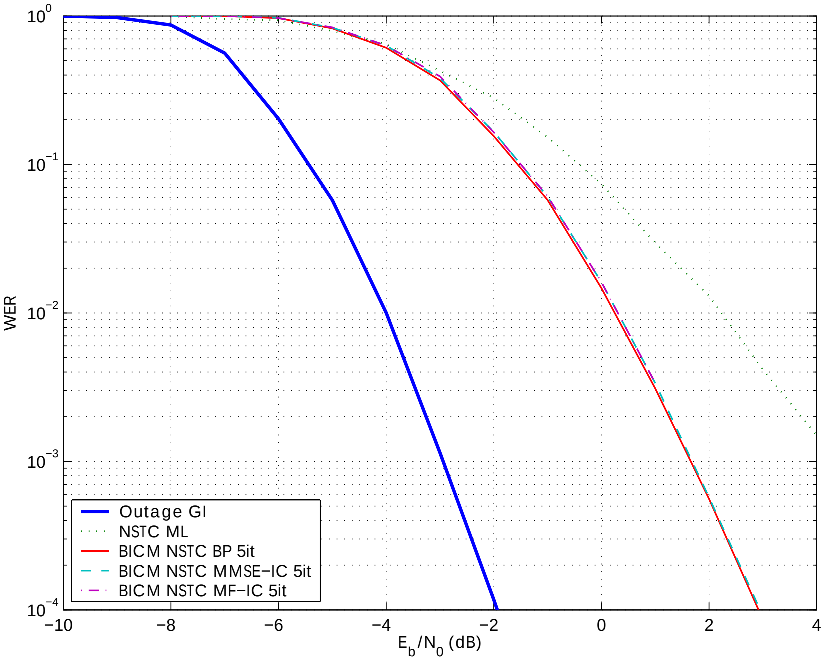


Figure 4.13: WER for $N_T = 4$, $N_R = 4$ and $R = 1$bit/s/Hz, with the $(5,7,7,7)_8$ convolutional code and BPSK modulation.

4.4.2 Impact of Signal Constellation Expansion

In this section we show some examples that compare complex-plane signal constellation expansion through BICM NSTCs with multidimensional constellation expansion through linear dispersion codes.

Figure 4.14 shows the WER performance in a $N_T = 2$ and $N_R = 2$ MIMO channel with the 4 states $(5,7)_8$ convolutional code of $r = 1/2$ with QPSK and 16-QAM modulations with Gray mapping. The spectral efficiencies are $R = 2, 4$ bit/s/Hz respectively. The block diversity of $\mathcal{C}$ is $d_B = N_T = 2$, and therefore, BICM NSTC should achieve full diversity with a *good* decoder. On the other hand, the diversity of BICM TAST is given by the TAST constellation and the coded modulation is only responsible for an horizontal shift of the error curve, i.e, the coding gain. Notice that, in such concatenated scheme, we can set $\phi_\ell = 1,\ \ell = 1, \ldots, N_L$ without any noticeable difference in performance, since the outer code removes most of the rank-deficient error events of the inner code. In this way, it suffices to find a good rotation matrix $\mathbf{M}$ in order to construct good BICM TAST codes. As we observe, under BP decoding and the same configuration, since BICM NSTC has full block diversity, all schemes perform almost identical regardless of their different nature. Notice that LD constellations induce an increased peak-to-average power ratio, which can make them impractical for applications where the power amplifier is operated close to saturation.

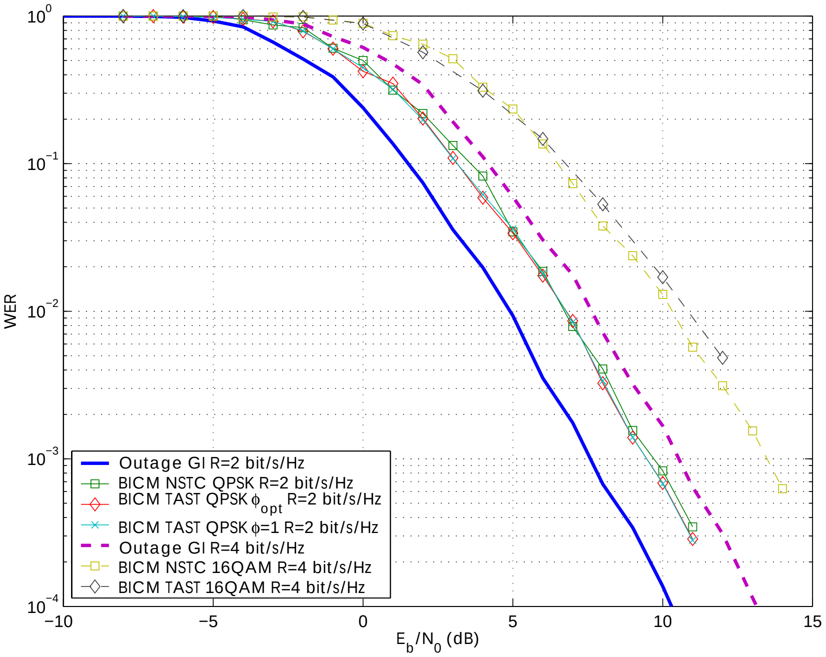


Figure 4.14: WER for $N_T = 2$ and $N_R = 2$ with $(5,7)_8$ convolutional code, QPSK and 16-QAM with Gray mapping.

In Figure 4.15 we report the WER performance in a $N_T = 2$ and $N_R = 2$ MIMO channel with the 4 states convolutional codes and QPSK and 8-PSK modulations with Gray mapping for an overall spectral efficiency of $R = 3$bit/s/Hz. In this case the frame is taken to be 132 information bits long. This figure clearly illustrates the effect of constellation expansion to achieve full diversity. In fact, in order to achieve $R = 3$bit/s/Hz with QPSK, we need the rate of $\mathcal{C}$ be $r = 3/4$. As we observe, under such configuration, BICM TAST achieves full diversity due to its inherent multidimensional constellation expansion. On the other hand, the diversity of BICM NSTC is governed by the Singleton bound (which in this case yields $d_B = 1$) and therefore under this configuration it does not achieve full-diversity. However, $R = 3$bit/s/Hz can also be achieved by using a rate $r = 1/2$

code (which has $d_B = N_T = 2$) and expanding the signal constellation in the complex plane, i.e., using 8-PSK modulation. As we observe, in this case, BICM NSTC achieves full diversity. However, it pays about a 1dB penalty in average power for the expansion with respect to the BICM TAST.

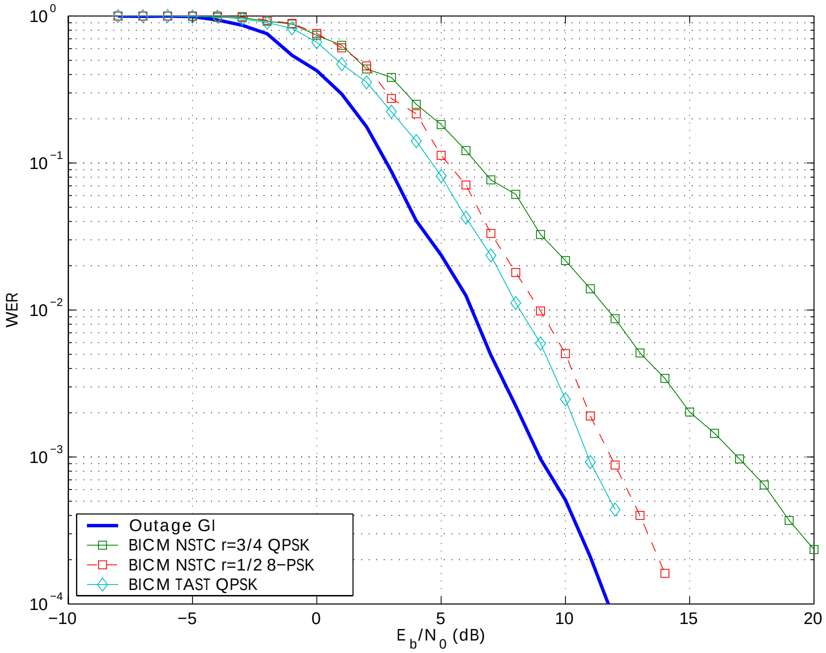


Figure 4.15: WER for $N_T = 2$, $N_R = 2$ and $R = 3$bit/s/Hz, with 4 states convolutional codes, QPSK and 8-PSK with Gray mapping.

Figure 4.16 shows the WER performance of BICM NSTC and BICM TAST in a MIMO channel with $N_T = 4$ and $N_R = 4$, with the 4 states $(5, 7, 7, 7)_8$ convolutional code of $r = 1/4$, using 16 and 64 QAM modulations with Gray mapping and MMSE-IC decoding. The corresponding spectral efficiencies are $R = 4, 6$ bit/s/Hz. In the case of 64-QAM we have considered frames of 120 information bits. We also plot the simulated matched filter bound (MFB), i.e., an ideal genie aided decoder for which the contribution of the signals other than the one it is detecting has been removed. In this example we observe that a new effect arises, namely, for too large spectral efficiency, even if the transmission schemes ensure full diversity, the MMSE-IC decoder is not able to remove the interference and achieve the correct slope. The characterization of the thresholds of the spectral efficiency for which the MMSE-IC is able to perform close to ML or BP is a very diffiult problem and at present there is no satisfactory explanation. In Appendix E we derive a semi-analytical method based on a combination of density evolution and bounding techniques, which however has complexity only sightly lower than simulation due to the outer expectation over the quasistatic fading.

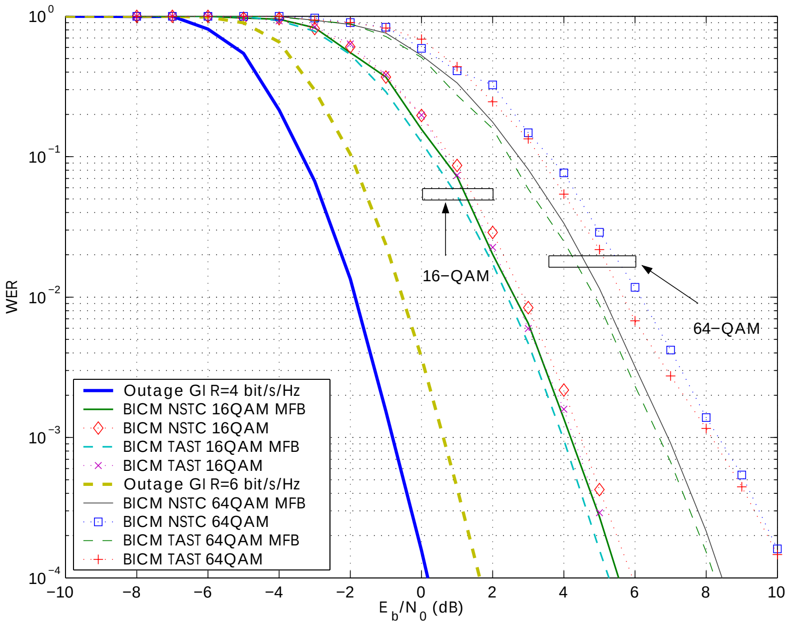


Figure 4.16: WER for $N_T = 4$, $N_R = 4$ with $(5,7,7,7)_8$ convolutional code, 16-QAM and 64-QAM with Gray mapping, with MMSE-IC decoding.

4.5 Conclusions

In this chapter we have illustrated the effect of signal constellation expansion on the achievable diversity in quasistatic MIMO channels by means of two simple pragmatic space-time code structures. In particular we have compared complex plane expansion and lattice-based expansion. We have shown that under exact belief propagation decoding and interference cancellation approximations, concatenated space-time codes with inner LD codes benefit from a higher design flexibility and show some performance advantage. However, in the same setting, both schemes perform equivalent. Thus, since LD-based methods induce an increased peak-to-average power ratio, we shall prefer the first approach in applications where the power amplifier is driven close to saturation.

5

Application to OFDM

This chapter illustrates the application of some of the proposed methods described in the previous chapters to realistic high-rate wireless multiple-antenna communication systems using OFDM modulation. In particular, we illustrate how the BICM NSTC scheme described in Chapter 4 compares with more standard approaches for space-time coding that employ orthogonal or quasi-orthogonal precoders, multidimensional linear mappings and classical V-BLAST. We perform this comparison under the framework of the ANTIPODE project of the French Telecommunications Research Council RNRT for which multiple-antenna space-time code solutions have been proposed and compared in order to increase the spectral efficiency of HIPERLAN/2. This comparative study will be done under two main channel scenarios: a typical office environment for NLOS propagation with independent antenna channels based on standard channel models, and a more realistic indoor multiple-antenna channel that includes transmit and receive antenna correlation generated with a ray tracing channel generator. We show that under both scenarios, BICM NSTCs compare favorably, and show remarkable performance advantage over the other approaches with extremely low-complexity receivers. The rest of the chapter is organized as follows. We first give a brief description of OFDM modulation and the basic models. We then review some aspects of the HIPERLAN/2 physical layer followed by a brief description of BICM NSTC design for OFDM channels. We finally give the system comparison.

5.1 OFDM with Cyclic Prefix

We consider an OFDM system with N_C subcarriers with cyclic prefix insertion in over frequency-selective fading [84]. The N_P-path discrete-time channel impulse response is given by

$$h(t) = \sum_{p=0}^{N_P-1} h_p \delta(t - pT_s) \tag{5.1}$$

where T_s is the signaling period and h_p is the amplitude gain corresponding to the p-th tap delay. This channel impulse response is obtained by sampling the *real* channel impulse response at the sampling rate. The delay intensity profile of the channel is defined as the expected path power $\mathbb{E}[|h_p|^2]$, for $0 \leq p \leq P-1$.

We consider transmission of blocks of length N_C. By adding a replica of the last N_{CP} symbols at the beginning of the block, the channel linear convolution becomes a circular convolution, which in the frequency domain, corresponds to the product of Fast Fourier Transforms (FFT). Therefore, the inter-symbol interference channel (5.1) can be modeled as a set of non-interfering parallel channels in frequency with a flat fading coefficient on each [84]. Therefore, the received signal y_c at the c-th subcarrier can be written as,

$$y_c = \sqrt{\rho} H_c x_c + z_c, \quad c = 1, ..., N_C \tag{5.2}$$

where $H_c \in \mathbb{C}$ is the flat frequency channel coefficient at subcarrier c, obtained as the c-th component of the FFT of the channel impulse response, $x_c \in \mathcal{X}$ is the transmitted constellation symbol drawn from $\mathcal{X}$ the signal constellation, which we assume to be normalized in energy, $z_c \in \mathbb{C}$ is the noise sample distributed as $\mathcal{N}_{\mathbb{C}}(0, 1)$, and ρ the SNR per sub-carrier. Notice that (5.2) shows that OFDM channels are somehow dual in the frequency domain to block-fading channel studied in the previous chapters. However, there is an important difference: the fading coefficients between adjacent frequency sub-carriers are in general correlated.

In an OFDM setting, the elements of diversity (degrees of freedom of the system) are given in the time domain. However, the equivalent transmission channel is in the frequency domain. In order to estimate the amount of diversity available in the OFDM channel, we may be interested in characterizing how different is the channel seen by a given subcarrier c from the one seen by another subcarrier c', separated by a frequency spacing ϕ. This calls for the autocorrelation of the frequency response,

$$\begin{aligned} R_H(\phi) &= \mathbb{E}\left[H(f)H^*(f-\phi)\right] && (5.3)\\ &= \mathbb{E}\left[\left(\sum_{p=0}^{N_P-1} h_p e^{-j2\pi fp}\right)\left(\sum_{q=0}^{N_P-1} h_q^* e^{j2\pi(f-\phi)q}\right)\right] && (5.4)\\ &= \sum_{p=0}^{N_P-1}\sum_{q=0}^{N_P-1} \mathbb{E}\left[h_p h_q^*\right] e^{-j2\pi f(p-q)} e^{-j2\pi\phi q} && (5.5)\\ &= \sum_{q=0}^{N_P-1} \mathbb{E}\left[|h_q|^2\right] e^{-j2\pi\phi q} && (5.6)\end{aligned}$$

which is nothing but the Fourier Transform of the delay intensity profile. Therefore, R_H will give us an indication on how the channel changes in frequency, i.e., how much frequency diversity is available in the channel.

The following example illustrates this concept. Consider the impulse responses of the typical NLOS office environment (BRAN-A [85]) rounded to the nearest path shown in Table 5.2.1 and the impulse response of an NLOS indoor channel extracted from the ray tracing program of [86]. The sampling frequency is 20 MHz. We first observe that the real indoor channel offers poor frequency diversity, due to the high correlation among the frequency sub-bands. In the best of the cases, we would have a correlation coefficient in frequency of 0.8. Moreover, it is observed that the BRAN-A model has much more diversity, due to its larger delay spread. We will use this in section 5.3, where we will design smart and greedy space-time codes based on the BICM NSTC structure introduced in the previous chapter. As we shall see, BICM NSTCs will exploit both space and frequency diversity when available based on simple designs for the standard block-fading channel.

5.2 HIPERLAN/2 Physical Layer

HIPERLAN/2 (HIgh PERformance Local Area Network), is a wireless local area network (WLAN) system standard specified by the European Telecommunications Standards Institute (ETSI) [5]. HIPERLAN/2 operates at a frequency of 5GHz, and there are 19 available channels of 20MHz each in Europe, with transmission rates up to 54 Mbit/s. HIPERLAN/2 physical layer has been harmonized with the US standard IEEE802.11a and the Japanese standard ARIB HiSWAN.

The block diagram of the basic transmission system of HIPERLAN/2 is described in Figure 5.2. HIPERLAN/2 uses BICM with several signal constellations and puncturing for multi-mode transmission. The data bits are first encoded with the rate $r = 1/2$ 64 states convolutional encoder with octal generators $(133, 171)_8$. Then, the encoded bits are interleaved with the bit-interleaver $\boldsymbol{\pi}$ and mapped onto

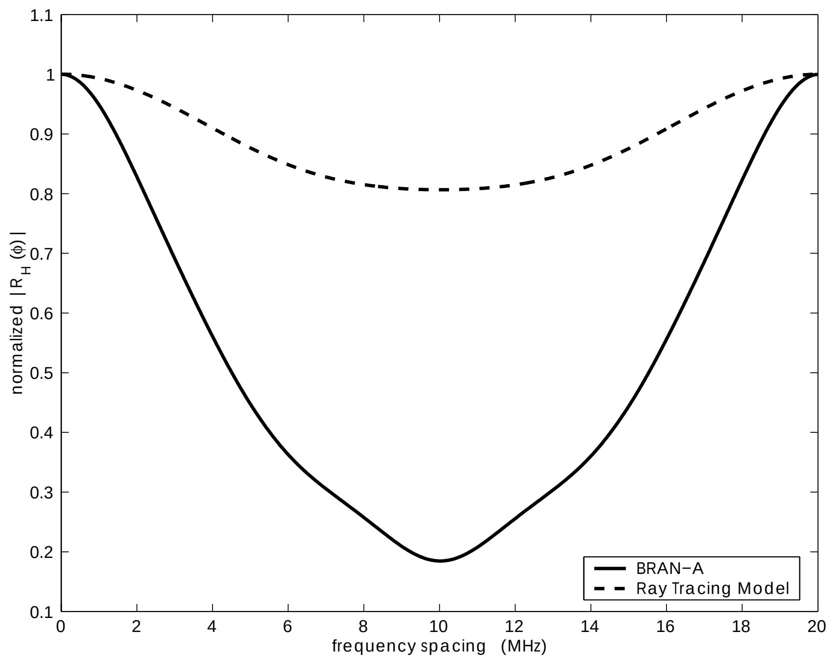


Figure 5.1: Normalized autocorrelation of the OFDM channel of BRAN-A and Ray Tracing impulse responses.

a signal constellation $\mathcal{X}$, which in this case is either BPSK, QPSK, 16-QAM or 64-QAM. The binary labeling rule μ is specified to be Gray. Then, the resulting signal is OFDM modulated with a modulator that performs an IFFT with $N_C = 64$ sub-carriers. Different transmission modes can be accommodated by appropriately puncturing the output of $\mathcal{C}$ and by using different signal sets. In particular, the transmission modes specified for HIPERLAN/2 are given in Table 5.2. The last transmission mode at 54 Mbit/s is left optional.

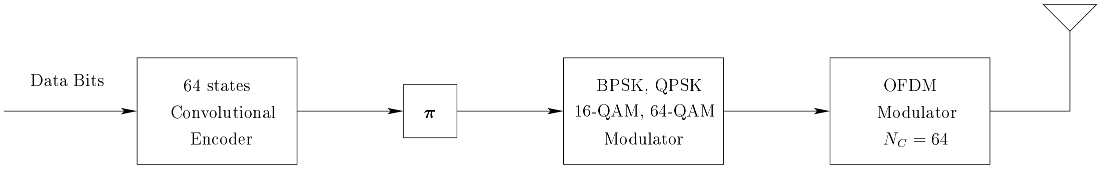


Figure 5.2: Basic transmission scheme of HIPERLAN/2.

The OFDM modulator of HIPERLAN/2 uses a sampling frequency of 20 MHz, with $N_C = 64$ sub-carriers (size of the IFFT), 48 of which are used for data transmission, 4 are used for phase tracking, while the rest are set to zero. The cyclic prefix length is $N_{CP} = 16$, so 80 symbols are effectively transmitted per OFDM symbol.

Mode	Transmission Rate	Rate r of $\mathcal{C}$	Signal Set $\mathcal{X}$
1	6 Mbit/s	1/2	BPSK
2	9 Mbit/s	3/4	BPSK
3	12 Mbit/s	1/2	QPSK
4	18 Mbit/s	3/4	QPSK
5	27 Mbit/s	9/16	16-QAM
6	36 Mbit/s	3/4	16-QAM
7	54 Mbit/s	3/4	64-QAM

Table 5.1: Transmission modes for HIPERLAN/2.

5.2.1 Environment Description

In this section we describe the channel environment and related parameters we used in order to compare the different space-time coding approaches. In particular we consider a MIMO OFDM channel with N_T transmit, N_R receive and N_C subcarriers per OFDM symbol, such that the received signal $\mathbf{Y}_c$ at sub-carrier c is given by

$$\mathbf{Y}_c = \sqrt{\rho}\,\mathbf{H}_c\mathbf{X}_c + \mathbf{Z}_c, \quad c = 1, \ldots, N_C \tag{5.7}$$

where $\mathbf{H}_c \in \mathbb{C}^{N_R \times N_T}$ is the *flat* channel matrix corresponding to the c-th sub-carrier, $\mathbf{X}_c$ is the corresponding transmitted signal at the c-th sub-carrier, $\mathbf{Z}_c$ is the noise matrix of samples i.i.d. $\sim \mathcal{N}_{\mathbb{C}}(0,1)$.

We shall consider the BRAN-A channel model specified by ETSI [85], for which the delay intensity profile is given in Table 5.2.1. The BRAN-A channel models a typical NLOS office propagation environment. We consider zero-order hold transmission filter, for which, the channel taps corresponding to the nearest multiple of 50 ns are rounded. This filter is shown to provide very similar results to other more realistic filters. Under the assumption of this zero-order hold filter, the resulting normalized (we assume that $\sum_p \mathbb{E}[|h_p|^2] = 1$) delay intensity profile of the BRAN-A channel is given in Table 5.2.1.

HIPERLAN/2 is based on ATM cells of length 54 bytes, and therefore, the number of symbols composing a packet varies depending on the selected transmission mode. Here we consider a number of information bits per packet equal to 384 bits, and we shall focus on the word error rate. We also fix a target WER of 10^{-2}. We will assume perfect channel knowledge at the receiver and no channel knowledge at the transmitter. We consider no Doppler effects and we will assume that the channel is quasistatic in time, i.e., the channel impulse response does not change during the transmission of a codeword and changes independently from one codeword to the next. We will also consider a more realistic indoor channel as the one provided by the ray-tracing simulator of France Telecom [86].

Delay (ns)	Path power (dB)
0	0.0
10	−0.9
20	−1.7
30	−2.6
40	−3.5
50	−4.3
60	−5.2
70	−6.1
80	−6.9
90	−7.8
110	−4.7
140	−7.3
170	−9.9
200	−12.5
240	−13.7
290	−18.0
340	−22.4
390	−26.7

Table 5.2: Delay intensity profile of the ETSI BRAN-A channel model.

5.3 BICM NSTC for OFDM

In this section we give a brief description on how the BICM NSTC approach described in the previous chapter can be adapted to efficiently work in OFDM channels. In particular, we propose to extend the technique proposed in Section 4.2.1 to the OFDM case and we show that the proposed method yields smart and greedy space-time codes.

Notice first, that applying the parsing rule illustrated in section 4.2.1 (Figure 4.6) assuming that $N_B = N_C$, i.e., treating the N_C carriers as independent, clearly yields to inefficient design, since N_C can be relatively large and the length of the interleavers would be constrained to a very short length. Another intuitive and more realistic option would be to take $N_B = N_P$, i.e., the number of resolvable paths of the channel impulse response. This approach, even if better than the previous, will also incur some inefficiency and diversity overestimation in the case where there is a very strong path and the rest have very low average power. We propose to base our design on the autocorrelation of the frequency response $R_H(\phi)$ from which we can estimate a rough number of decorrelated components, or *virtual* fading blocks. We denote the number of such blocks by N_{VB}. We can then apply the parsing rule described in section 4.2.1 by simply replacing N_B

Delay (multiples of 50 ns)	Normalized path power (dB)
1	-3.4630
2	-4.6006
3	-8.9151
4	-12.8223
5	-19.9222
6	-21.1202
7	-25.4329
8	-29.7891
9	-34.1993

Table 5.3: Delay intensity profile of the normalized BRAN-A channel model assuming a zero-hold order transmission filter.

by N_{VB}. Notice that standard approaches perform random interleaving across frequencies and treat the OFDM as a quasistatic channel, i.e., $N_{VB} = 1$ [60]. In the case where $N_{VB} > 1$, interleavers have to be carefully design in order to always maintain the same frequency spacing between the carriers assigned to different blocks. Notice that, using the same interleaver permutation in all the virtual blocks of the same transmit antenna would do. In this way, we ensure that the frequency diversity can be efficiently exploited.

The next example illustrates the benefits of the proposed design in MIMO OFDM channels. Figure 5.3 illustrates the WER performance for $N_T = N_R = 2$, $(5,7)_8$ convolutional code with BPSK modulation on the BRAN-A and Ray Tracing OFDM channels, with $N_C = 64$, for several values of N_{VB} with the MAP bitwise demodulator with 5 decoding iterations. We consider 128 information bits per codeword. We observe significant diversity gain with $N_{VB} > 1$ virtual blocks per OFDM symbol in both channels with respect to the design in [60], supporting the arguments given before. As mentioned before, [60] treats the OFDM channel as quasistatic, since $\mathcal{P}$ is used with $N_{VB} = 1$. As illustrated in Figure 5.1, there is a relevant diversity difference between the two channel models, and therefore, the relative performance gain with virtual blocks is accordingly related. We remark that when we have a large diversity order (in space, frequency or both), we will only observe this effect at low WER. Also, using more powerful codes will induce such differences at low WER.

5.4 Performance Comparison with Other Schemes

In this section we compare the BICM NSTC approach described in the previous section with other space-time coding schemes for HIPERLAN/2 . We will

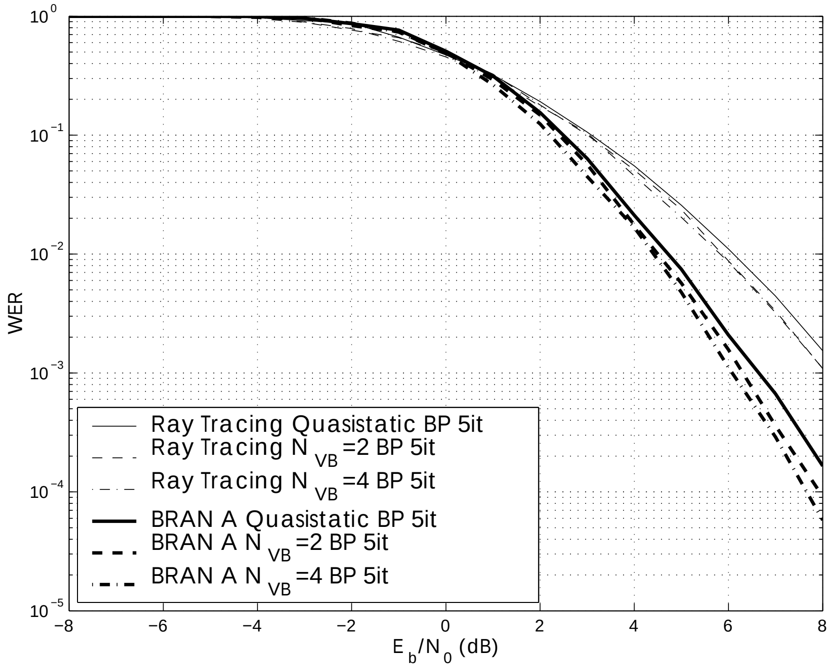


Figure 5.3: WER for $N_T = N_R = 2$, $(5,7)_8$ code with BPSK, OFDM BRAN-A and Ray Tracing models, for several N_{VB}.

see, that, under both, the BRAN-A channel with independent antenna channels and the realistic channel with correlated antennas, BICM NSTCs significantly outperform the other approaches with extremely low-complexity receivers.

We consider a variety of space-time codes proposed for the specific application of HIPERLAN/2. A detailed description of the schemes can be found in [15]. In the following we list the approaches described in [15] and we outline their main features.

Concatenation with Orthogonal Precoders This approach considers the HIPERLAN/2 BICM described in section 5.2 followed by a space-time orthogonal precoder that maps the modulated symbols onto the transmit antennas. In the case of $N_T = 2$ this approach uses a classical Alamouti scheme [31] that at every two BICM symbols $x_1, x_2 \in \mathcal{X}$ transmits the orthogonal rate 1 matrix (rows correspond the antenna dimension while columns correspond to different time instants)

$$\begin{bmatrix} x_1 & x_2 \\ -x_2^* & x_1^* \end{bmatrix} \tag{5.8}$$

over the channel. As indicated in [87], there are no space-time orthogonal codes for $N_T > 2$ of rate 1 symbol per channel use. Then, in order to avoid such rate loss, the ABBA code [88] will be used instead for the case of $N_T = 4$ transmit antennas. We recall that the ABBA code is simply a rearrangement of Alamouti matrices, i.e., every 4 BICM symbols $x_1, x_2, x_3, x_4 \in \mathcal{X}$ it transmits,

$$\begin{bmatrix} x_1 & x_2 & x_3 & x_4 \\ -x_2^* & x_1^* & -x_4^* & x_3^* \\ x_3 & x_4 & x_1 & x_2 \\ -x_4^* & x_3^* & -x_2^* & x_1^* \end{bmatrix}. \tag{5.9}$$

This matrix is no longer orthogonal and in order to cope with the residual interference and keep low-complexity decoding, a zero-forcing (ZF) strategy is used [88, 15].

Concatenation with Space-Frequency Linear Precoders This scheme is essentially based on using threaded layering combined with a linear multidimensional rotation. A detailed exposition of this approach can also be found in [89]. Essentially it can be seen as the OFDM counterpart of BICM TAST, even though there are several differences with the scheme presented in Chapter 4. This approach considers data streams that are independently encoded, bit-interleaved, mapped over the signal constellation and rotated by a multidimensional linear mapping. Then, a greedy antenna-tone assignment algorithm is performed in order to exploit both space and frequency diversity. At the receiver side, MMSE-IC decoding is performed. We shall refer to this approach as the linear precoding (LP) scheme.

Coded V-BLAST As in the orthogonal precoders case, this scheme treats again the BICM described in Section 5.2 essentially as uncoded data, and cyclically distributes the modulation symbols across the transmit antennas. At the decoder side, the classical MMSE V-BLAST decoding algorithm is used [90, 79], that performs a combination of linear filtering and successive interference cancellation in the optimal detection ordering proposed in [79].

5.4.1 Comparison in the Independent BRAN-A Channel

In this section we present some numerical examples that illustrate the performances of BICM NSTC and the aforementioned schemes in the BRAN-A channel model, with 384 information bits per codeword and $N_C = 64$. We shall compare all schemes for equal spectral efficiency and same trellis complexity, i.e., they will all use 64 states convolutional codes. We model each separable path p of the impulse response as an i.i.d. Gaussian random variable with zero mean and variance $\mathbb{E}[|h_p|^2]$ and we assume no correlation among antennas. We also assume that BICM NSTC uses the simplest possible iterative decoder, namely, the MF-IC described in Chapter 4. We will use 5 decoding iterations. From extensive results not shown here, we have observed that the number of virtual blocks shows its effect only at very low WER, and thus, in these examples, we only report the corresponding curves for $N_{VB} = 1$. As we shall see, even in this case, BICM NSTCs show remarkable performance gain.

In Figures 5.4, 5.5, 5.6 and 5.7 we show the WER performance of all studied schemes in the BRAN-A channel with $N_T = N_R = 2$ with $R = 1, 2$ bit/s/Hz and $N_T = N_R = 4$ with $R = 2, 4$ bit/s/Hz respectively. As we observe, BICM NSTC outperforms in general the other schemes. In particular, in the $N_T = N_R = 4$ the performance advantage is remarkable. Notice also that the complexity of the receiver is extremely low: once the channel parameters are available to the receivers, it suffices to perform a matched filter operation and perform several IC iterations.

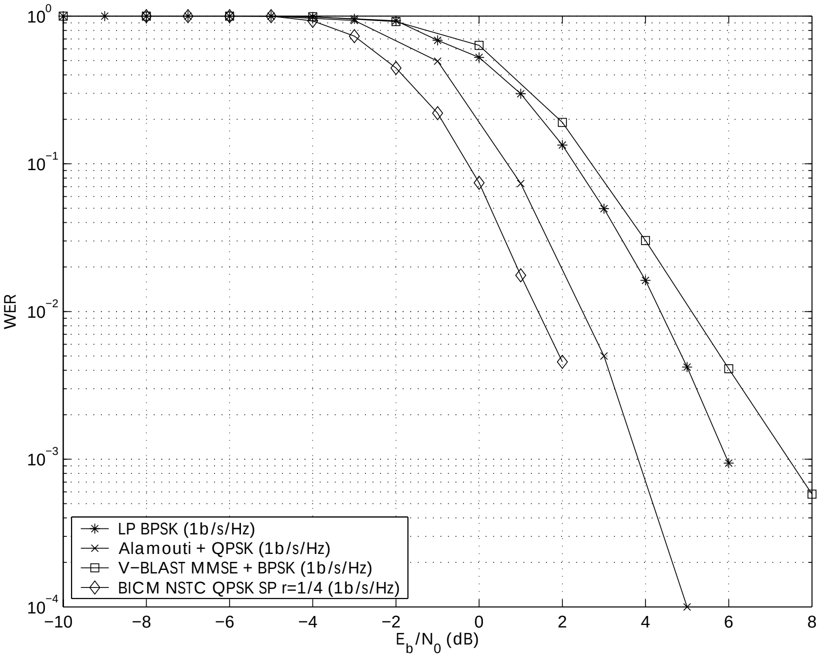


Figure 5.4: WER for $N_T = N_R = 2$, $R = 1$ bit/s/Hz with OFDM BRAN-A channel.

5.4.2 Comparison in the Realistic NLOS Indoor Channel

In this section we show the results corresponding to the same comparison but under the realistic NLOS indoor channel output of the ray-tracing program provided by France Telecom [16, 86]. This is a ray-tracing program that computes the channel impulse responses of an indoor NLOS MIMO channel assuming a 5GHz carrier frequency, sampling frequency of 20MHz, with the antenna correlation pattern defined in [16].

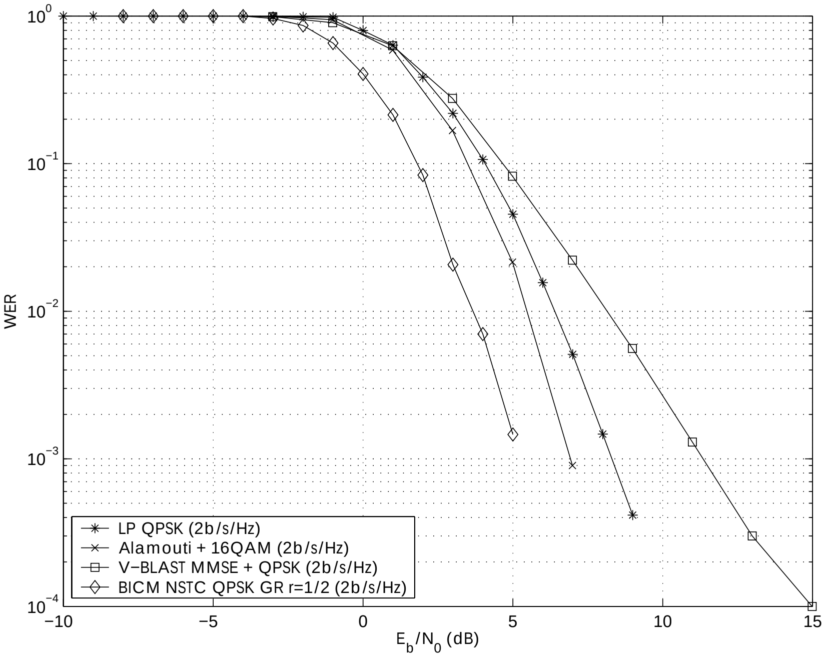


Figure 5.5: WER for $N_T = N_R = 2$, $R = 2$ bit/s/Hz with OFDM BRAN-A channel.

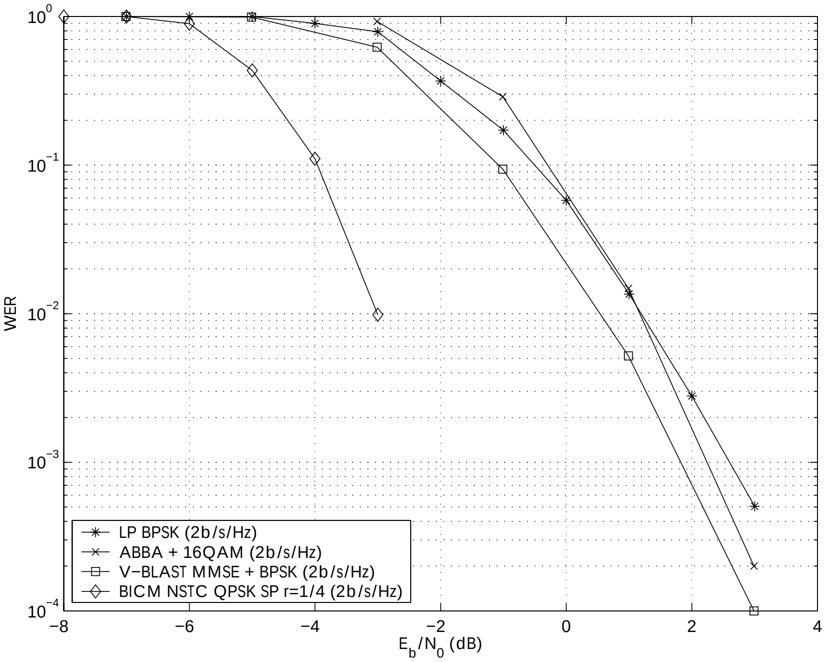


Figure 5.6: WER for $N_T = N_R = 4$, $R = 2$ bit/s/Hz with OFDM BRAN-A channel.

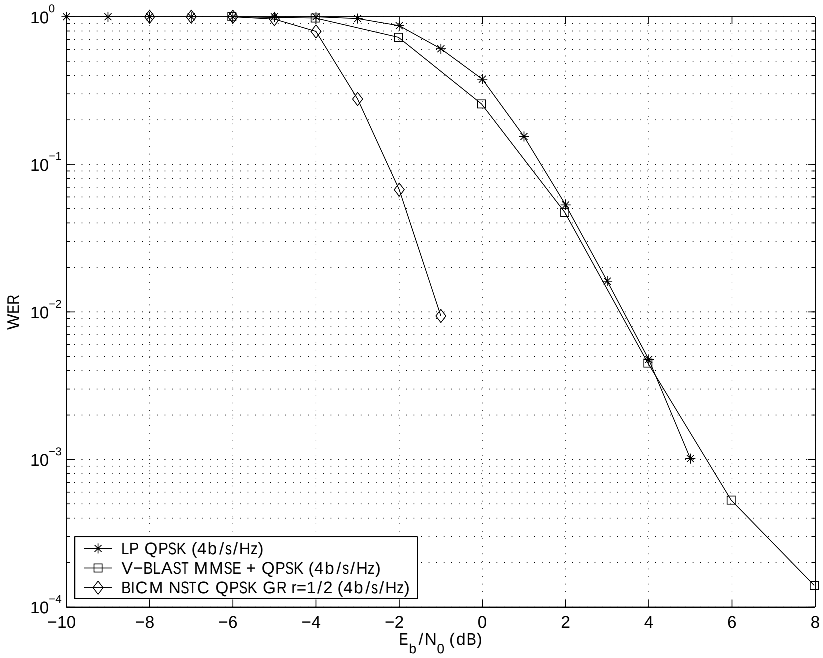


Figure 5.7: WER for $N_T = N_R = 4$, $R = 4$ bit/s/Hz with OFDM BRAN-A channel.

As we observe from Figures 5.8, 5.9, 5.10 and 5.11, where we replicate the BRAN-A examples over the realistic channel, the relative performance between the different approaches is essentially the same, even if, as we clearly see, the performance curves are less steep than those of the previous case, due to the lower diversity order of the realistic channel. Notice also that in Figure 5.11 BICM NSTC use the MMSE-IC decoder, since, the MF-IC exhibits poor performance. Again, as anticipated in Chapter 4, it is very difficult to analytically state the reasons for which a given IC decoder does not converge in average to the MFB.

5.5 Conclusions

In this Chapter we have seen how BICM NSTCs can be adapted to the OFDM channel by suitably designing the parsing rule and the bit-interleavers. We have performed a system comparison in the framework of the HIPERLAN/2 system with other space-time coding schemes that use orthogonal or quasi-orthogonal precoders, multidimensional linear precoders and standard coded V-BLAST. We have performed such comparison in two different channel environments, namely, a typical office environment with independent pairwise antenna channels, and a more realistic environment generated with a ray tracing simulation program that takes into account the antenna correlation pattern. We have shown, by means

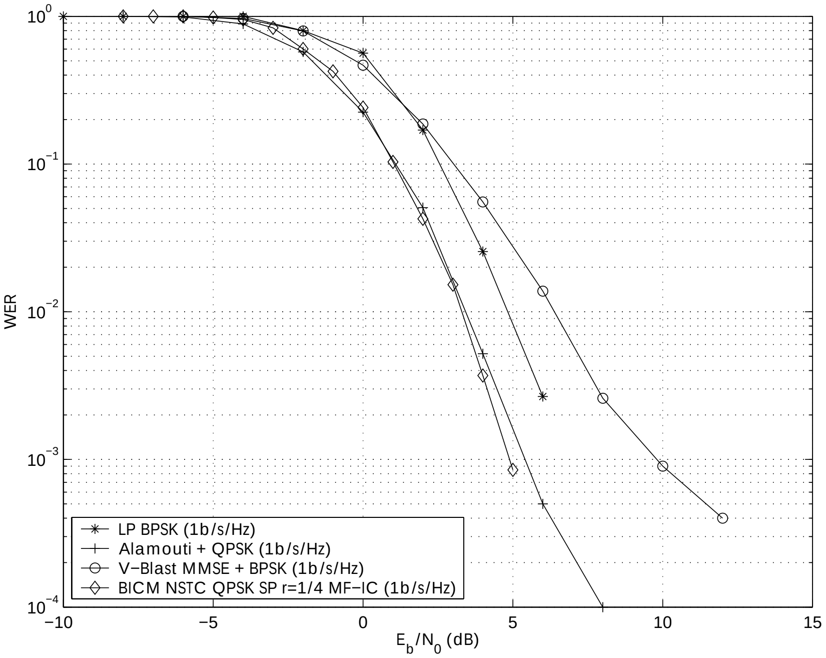


Figure 5.8: WER for $N_T = N_R = 2$, $R = 1$ bit/s/Hz with the realistic NLOS indoor OFDM channel generated with the ray tracing program simulator.

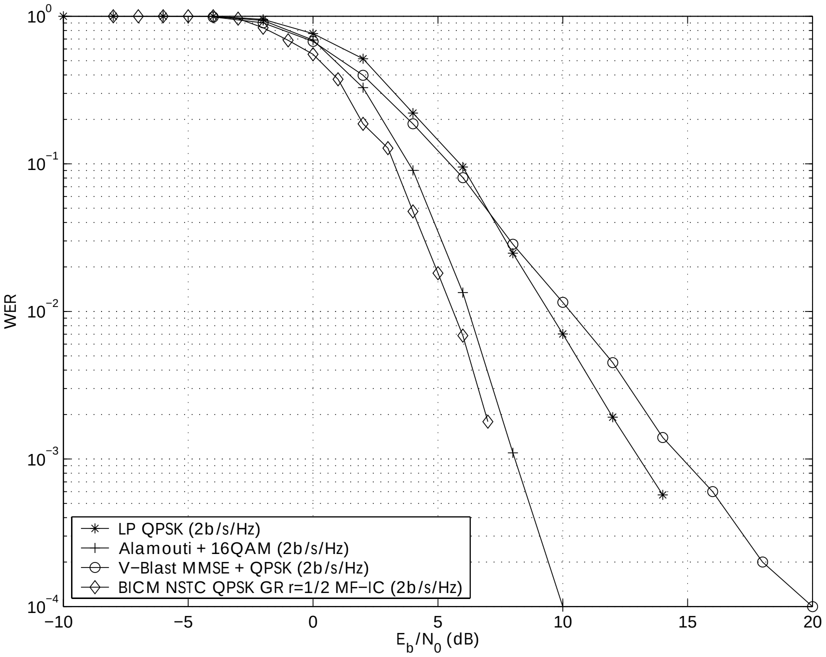


Figure 5.9: WER for $N_T = N_R = 2$, $R = 2$ bit/s/Hz with the realistic NLOS indoor OFDM channel generated with the ray tracing program simulator.

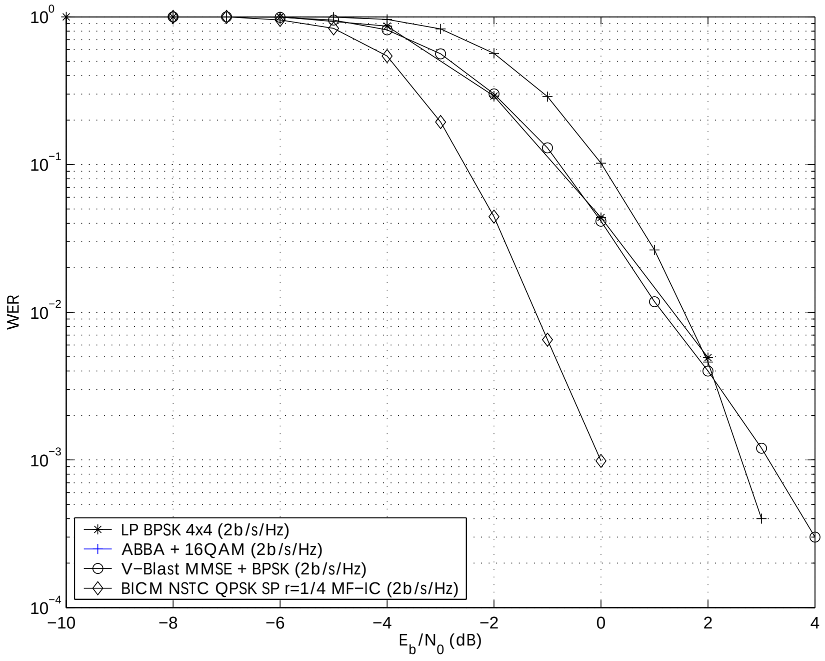


Figure 5.10: WER for $N_T = N_R = 4$, $R = 2$ bit/s/Hz with the realistic NLOS indoor OFDM channel generated with the ray tracing program simulator.

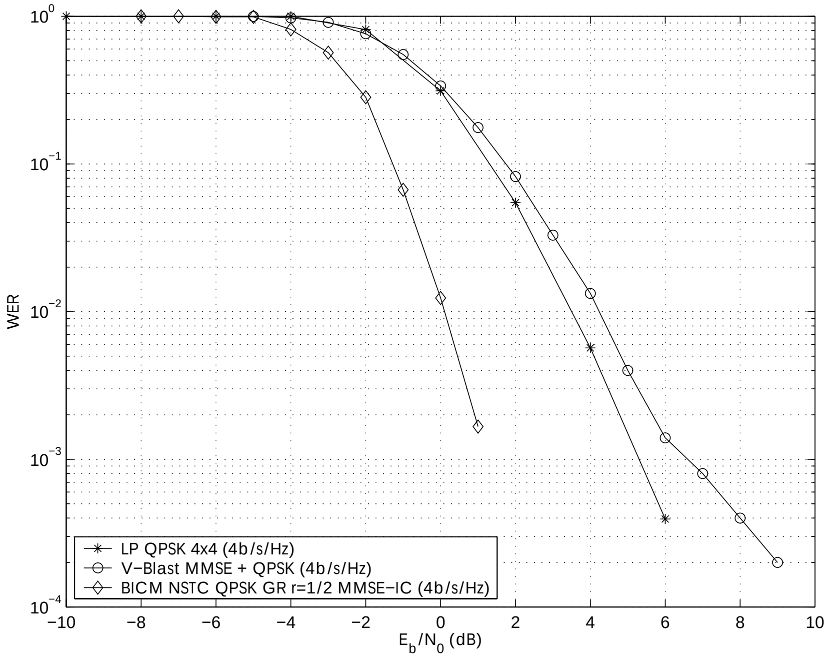


Figure 5.11: WER for $N_T = N_R = 4$, $R = 4$ bit/s/Hz with the realistic NLOS indoor OFDM channel generated with the ray tracing program simulator.

of numerical examples, that BICM NSTCs remarkably outperform the other schemes with very low-complexity iterative decoding, thus becoming a strong candidate for real implementation due to its design simplicity and flexibility.

6

Conclusions and Future Work

In this thesis, we have focused on designing *efficient* coded modulation schemes for block-fading channels. In such non-ergodic fading channels, there is a nonzero probability that the transmitted message is decoded in error, i.e., the outage probability. When both transmitter and receiver operate with one antenna, by studying the asymptotic behavior of the outage probability, we have determined the SNR reliability function of coded modulation schemes. In particular, we have rigorously proved that coded modulation schemes constructed over a discrete signal constellation achieve the optimal trade-off provided that the block length grows sufficiently fast with SNR. We have also shown that pragmatic schemes with high spectral efficiency obtained by concatenating a binary code and a discrete signal constellation through BICM coupled with the suboptimal BICM-ML decoder achieve the same optimal trade-off that their non-BICM counterparts with optimal ML decoding. This represents an important step towards the construction of practical coding schemes that transmit at large spectral efficiencies and achieve the fundamental limits of this channel.

Motivated by this fact, we have proposed a new randomlike coded modulation family based on the combination of a blockwise concatenation of binary codes and BICM. We analyzed the optimal performance of such code ensembles by using easily computable upper bounds and tight approximations on the ML and BICM-ML decoding error probability. As opposed to standard ergodic channels, we have illustrated that the proposed bounds and approximations accurately match the performance obtained by simulation of the iterative BP decoder. We have shown that the proposed codes have an error probability almost independent of the block length, that implies a fixed SNR gap (independent of the block length) from the outage probability. This stands in sharp contrast to the case of standard block codes obtained from trellis-terminated convolutional codes specifically designed for the block-fading channel, which show a gap from the outage probability limit that increases with the block length. In fact, in the limit for large block length, the error probability of such concatenated codes is given by the distribution of their decoding threshold seen as a random variable function of the channel state. This is the dual behavior of classical capacity achieving codes in ergodic channels.

In a second part of the thesis, we have investigated coded modulation schemes for the multiple-antenna block-fading channel. Traditional design criteria are difficult to satisfy and usually lead to coding schemes that induce large decoding complexity. Here we have taken a step back and we have designed pragmatic space-time codes based on the optimistic assumption of a genie-aided decoder. Under this assumption, code design is then reduced to classical techniques for single-antenna block-fading channels. We have emphasized that achieving full diversity will depend on the ability of the real decoder to emulate the genie-aided performance. We have primarily focused the role of signal constellation expansion to achieve full diversity and we have compared two different means of expanding the signal constellation, namely, complex-plane and multidimensional expansion. We have shown that under average power constraint and message passing decoding, multidimensional expansion is in general advantageous due to its inherent design flexibility. On the other hand, this is no longer true under stringent peak power constraints, namely, in systems that operate power amplifiers close to saturation, since multidimensional signal expansion induces a significant increase in the peak-to-average power ratio.

We have also reported a comparative study of several space-time coding techniques in the MIMO quasistatic OFDM channel. This channel arises in important applications such as future generation indoor wireless LANs, for which higher data rates than those achieved by the IEEE802.11a current standard should be obtained by appropriately using the multiple antennas. We have seen that under both ideal and realistic channel environments, a very simple scheme based on a design for the SISO block-fading channel remarkably outperforms the classical Alamouti and V-BLAST approaches as well as multidimensional rotated schemes.

6.1 Areas for Further Research

Further research is needed to establish the performance of blockwise concatenated codes under iterative decoding. This is of particular interest in order to establish the *real* performance of such codes as well as to validate how close iterative and ML decoding perform in the block-fading channel. In order to do this, one should have at hand a fast density evolution algorithm, since we are interested in the distribution of the threshold seen as a random variable function of the channel state. This will in turn have an impact on establishing the rate regions where iterative IC multiple-antenna receivers work properly. First sparks in this line can be found in [91], where the authors have proposed a very fast density evolution algorithm for standard LDPC codes.

We have not studied how LDPC codes behave in the block-fading channel. Some simulation results can be found in [92], where the authors compare LDPC codes with convolutional codes both with short block length. The final conclusion was that they showed marginal improvement over convolutional codes. From the results shown in this thesis, we believe this conclusion can be reversed, but further research is needed in this line. For example, it is not clear the way LDPC codes approach the optimal trade-off given by the Singleton bound, and whether there is a specific graph design that would yield MDS LDPC codes. As done for standard LDPC codes in ergodic channels, it is probably worth analyzing first the performance of LDPC codes in the block-erasure channel.

We have completely ignored the impact of channel estimation in the robustness of the proposed algorithms. This is indeed a crucial aspect that must be assessed in order to fairly compare different coding strategies.

As we have seen in Chapter 4, there are some situations in which in average, simple iterative IC decoders are not able to remove the interference from the signals corresponding to other antennas. This major bottleneck can be overcome by using low-complexity implementations of soft-input soft-output sphere (or list) decoders. Preliminary results seem to indicate that further improvements have to be done when using the shifted list decoder proposed in [73]. In this case, the list is only based on the uncoded ML point and the properties of the induced channel lattice. In some cases, this decision can be extremely unreliable and thus, the effect of the feedback from the decoder of the underlying binary code should also be taken into account in order to improve the detection list quality through the iterations. In [76] (see also [93]) the authors recognize the importance of combining MMSE filtering with sphere decoding. We believe that MMSE prefiltering can also play a significant role in soft-input soft-output sphere decoders resulting in a decoding list of higher quality. In this way, one can perform system comparisons for much larger spectral efficiencies, which are of real interest in high data rate systems.

A related broad area that we have not considered is the design of efficient coding schemes for the case when channel state information is available at the transmitter. It is well known that channel state information at the transmitter can yield remarkable gain. In particular, it is possible to control or remove outage events by using power control or variable-rate transmission techniques. The optimal power control rule is given in [20] for Gaussian inputs, which can be computed in closed form and in most of the cases leads to zero-outage error probability. However, to our knowledge, there are no practical coding schemes able to approach these limits.

A

Error Probability of BICM with the Gaussian Approximation

This appendix presents a very accurate and simple to compute approximation to the performance of BICM systems in AWGN channels. We employed this method to compute tight error probability estimations of BCC codes with BICM in Chapter 3. The proposed method is based on approximating the binary-input continuous-output equivalent BICM channel by a binary-input AWGN (BI-AWGN) channel with scaled SNR. The scaling factor can be easily computed numerically and depends on the actual channel SNR and on the modulation signal set and binary labeling. The key is that very good approximation results when the Bhattacharyya parameter is used to estimate the variance of the underlying Gaussian channel. Under such approximation, we can use all bounding techniques known for binary codes in Gaussian channels. In particular, we use the union and the tangential-sphere bounds and we apply such results to both convolutional and turbo-like codes. The proposed method represents a simple yet powerful tool for estimating the error probability of finite-length turbo-like codes with BICM.

A.1 Introduction and Motivation

Bit-interleaved coded modulation (BICM) was introduced in [94] and further generalized and elaborated in [36] as a means of coding for spectrally efficient modulations. In essence, it states that nearly optimal performance can be achieved by concatenating a powerful binary code with a non-binary modulator, by the simple addition of a bit-interleaver between these two components. An additional advantage offered by BICM is its inherent flexibility, as a single mother code may be used for several modulations, with no additional adaptations. This is an appealing feature for future communication systems where a large set of spectral efficiencies is needed.

The original works on BICM [94, 36], consider a decoder that for every symbol produces soft statistics for the bits of its binary label, and feeds these values to a ML decoder of the mother code, as if they were outputs of a *virtual* binary-input continuous-output channel. We shall refer to this decoder as BICM-ML decoder and the virtual channel as the equivalent BICM channel.

Several works have also considered iterative decoding of convolutionally encoded BICM with optimized mappings, showing remarkable performance increase with respect to Gray mapping (see e.g. [38, 95]). Simple iterative decoding analysis of such system has been provided in [38] based on an approximation of density evolution techniques [37] for infinite block length. However, no such satisfactory results have been observed when using capacity approaching codes such as turbo-like or LDPC codes, where iterative decoding analysis is very complicated (see [96] for recent results on the subject). Therefore it is common practice to couple turbo-like or LDPC codes and BICM with Gray mapping, since it offers the best performance-complexity trade-off.

Error probability bounds of finite-length BICM under BICM-ML decoding have been derived in [36]. The metric model assumed by the BICM-ML decoder is nearly optimal for Gray mapping and assumes no demapping iterations. A simple union bound based on a bitwise Bhattacharyya factor was found to be quite loose. Several refined techniques, also derived in [36], provided more accurate results, but are much more complex to compute. In this work, we provide a very simple method that allows for the computation of very accurate approximations on the error probability of BICM with BICM-ML decoding, which is mainly based on a Gaussian approximation (GA) of the binary-input BICM equivalent channel. We verify the validity of the approximation and we apply the results to compute tight union and tangential-sphere bounds for both convolutional and turbo-like codes. We also illustrate how the proposed approximation can be used to compute BICM-ML thresholds. We show that the proposed method is a simple and powerful tool, since it yields very accurate error probability estimates with little computational effort.

A.2 System model

We consider a classical additive white Gaussian noise (AWGN) channel, for which the received signal at time k, $y_k \in \mathbb{C}$ is given by,

$$y_k = \sqrt{\rho} x_k + z_k \ \ , \ k = 1, \dots, L \tag{A.1}$$

where $x_k \in \mathbb{C}$ is the transmitted signal at time k with L the codeword length, $\rho = \frac{E_s}{N_0}$ is the the signal-to-noise ratio (SNR) and $z_k \in \mathbb{C}$ is the complex noise sample at time k i.i.d. $\sim \mathcal{N}_{\mathbb{C}}(0,1)$. We denote by $\mathcal{X} \in \mathbb{C}$ the complex signal constellation (i.e. PSK, QAM). Without loss of generality we study unit energy constellations, i.e., $\mathbb{E}[|x|^2] = 1$.

The codewords $\mathbf{x} = (x_1, \ldots, x_L) \in \mathcal{X}^L$ are BICM codewords obtained by bit-interleaving the codewords $\mathbf{c}$ of the binary code $\mathcal{C} \in \mathbb{F}_2^N$ of rate $r = K/N$, and mapping with the mapping rule μ, that maps binary labels of length $M = \log_2 |\mathcal{X}|$ over to the points of $\mathcal{X}$ [94, 36]. The resulting BICM codeword length is L modulation symbols, with $L = N/M$, and the spectral efficiency of such system is $R = rM$ bit/s/Hz.

Due to the presence of the bit-interleaver, ML decoding of BICM is only possible by exhaustive search. Several suboptimal strategies have been used, all derived from the belief-propagation iterative algorithm [44], for which, the bit metrics of a given bit b being in the m-th label position of a given symbol are given by,

$$p(y|b,m) \propto \sum_{z \in \mathcal{X}_b^m} p(y|z) P(z) \tag{A.2}$$

where $\mathcal{X}_b^m$ is the set of all signal constellation symbols with bit b in the label position m, $p(y|z) = \frac{1}{\pi}\exp(-|y - \sqrt{\rho}z|^2)$ is the channel transition probability density function (pdf), and $P(z)$ denotes the *a priori* probability of the symbol z. When iterative demapping is performed, $P(z)$ are given by the decoder of $\mathcal{C}$. In this work, we restrict our attention to the case of equally likely symbols where no demapping iterations are performed, i.e. $P(z) = \frac{1}{|\mathcal{X}|}$. The suboptimal BICM-ML decoder is known to perform near optimal for signal constellations with Gray mapping [36, 38]. We will refer to the channel between a given binary codeword $\mathbf{c} \in \mathcal{C}$ and its corresponding bit-metrics, as the equivalent binary-input BICM channel. In particular, for a coded binary symbol mapped to the m-th label position of the k-th modulation symbol, the *bit-wise* posterior log-probability likelihood ratio (LLR) is given by

$$\mathcal{L}_{k,m} = \log \frac{\displaystyle\sum_{x \in \mathcal{X}_0^m} \exp\left(-|y_k - \sqrt{\rho}x|^2\right)}{\displaystyle\sum_{x \in \mathcal{X}_1^m} \exp\left(-|y_k - \sqrt{\rho}x|^2\right)}. \tag{A.3}$$

We also report the capacity with BICM under suboptimal BICM-ML decoding, which is given by [36],

$$C = M - \frac{1}{2^M} \sum_{m=1}^{M} \sum_{b=0}^{1} \sum_{x \in \mathcal{X}_b^m} \mathbb{E}_n \left[\log_2 \frac{\displaystyle\sum_{z \in \mathcal{X}} e^{-|\sqrt{\rho}(x-z)+n|^2}}{\displaystyle\sum_{z \in \mathcal{X}_b^m} e^{-|\sqrt{\rho}(x-z)+n|^2}} \right], \tag{A.4}$$

which is known to be maximized for Gray or quasi-Gray binary labeling rules. For the sake of future reference, we show in Figure A.1 the capacity with 16-QAM inputs and the BICM capacity with 16-QAM and Gray mapping.

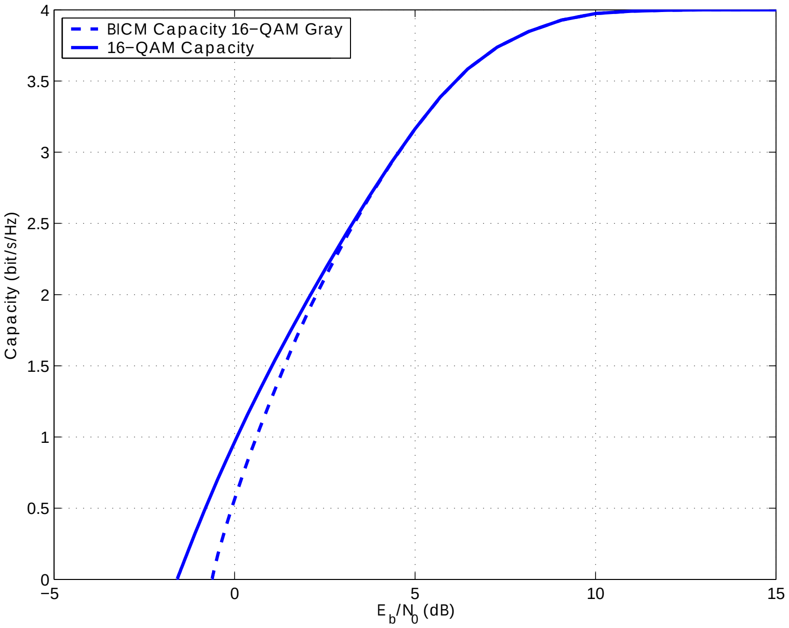


Figure A.1: Capacity for 16-QAM inputs and BICM capacity with Gray mapping.

A.3 The Bhattacharyya Union Bound

The Bhattacharyya Union Bound for BICM was first proposed in [36] as a simple approach for upper bounding the error probability of BICM under BICM-ML, i.e., ML decoding of $\mathcal{C}$ using the bit-metrics (A.2) with no demapper iterations. For the frame error probability it is given by,

$$P_e \leq \sum_w A_w B(\rho)^w \tag{A.5}$$

where $A_w = |\mathcal{S}_w|$ is the weight enumeration function (WEF) of $\mathcal{C}$ and accounts for the number of pairwise error events of $\mathcal{C}$ at Hamming distance w, with $\mathcal{S}_w = \{\mathbf{c} \in \mathcal{C} : w_H(\mathbf{c}) = w\}$ denoting the set of codewords with Hamming weight w, and $B(\rho)$ is the Bhattacharyya factor which is given by,

$$
\begin{aligned}
B(\rho) &= \mathbb{E}_{y,m,b}\left[\sqrt{\frac{p(y|\bar{b},m)}{p(y|b,m)}}\right] \\
&= \frac{1}{M2^M}\sum_{m=1}^{M}\sum_{b=0}^{1}\sum_{x\in\mathcal{X}_b^m}\mathbb{E}_n\left[\sqrt{\frac{\sum_{z\in\mathcal{X}_{\bar{b}}^m} e^{-|\sqrt{\rho}(x-z)+n|^2}}{\sum_{z\in\mathcal{X}_b^m} e^{-|\sqrt{\rho}(x-z)+n|^2}}}\right]
\end{aligned}
\qquad (A.6)
$$

where $p(y|b,m)$ are given in (A.2), the expectation is over the joint distribution of the received signal y, the labeling position m and the bit b as value of the m-th label position and $\bar{b}$ denotes the binary complement. Notice that $B(\rho)$ depends actually on the signal constellation and the binary mapping. Also note that this expectation can be easily evaluated using the Gauss-Hermite quadrature rules which are tabulated in [45], since $n \sim \mathcal{N}_{\mathbb{C}}(0,1)$. The results in [36] show that (A.5) can be very loose, and therefore, not very useful as analytical tool to describe the error probability of BICM. Reference [36] elaborates more refined bounds on the performance of the BICM-ML decoder, which are, however, much harder to compute.

A.4 The Gaussian Approximation

In this section we describe how the Gaussian approximation can be used to describe the binary-input equivalent channel of BICM. Notice that the Gaussian approximation of the binary-input BICM channel is commonly employed in convergence analysis of iterative decoding based in density evolution techniques [38, 37] for infinite block length. Consider for a moment that the binary code $\mathcal{C}$ mapped over a BPSK signal constellation (i.e., $\mathcal{X} = \{-1,+1\}$) is transmitted across a binary-input AWGN channel with SNR γ. Then, the standard union and Bhattacharyya bounds can be written as,

$$
P_e \le \sum_{w} A_w Q\left(\sqrt{2w\gamma}\right) \le \sum_{w} A_w B_2(\gamma)^w \qquad (A.7)
$$

the Bhattacharyya factor for binary inputs $B_2(\gamma)$ is given by [97],

$$
B_2(\gamma) = \mathbb{E}_y\left[\sqrt{\frac{p(y|-1)}{p(y|+1)}}\right] = e^{-\gamma}. \qquad (A.8)
$$

It is also well-known that for such binary-input AWGN channel (BI-AWGN), the log-likelihood ratio (LLR) defined as

$$
\mathcal{L} = \log\frac{p(y|-1)}{p(y|+1)} \qquad (A.9)
$$

is $\sim \mathcal{N}(4\gamma, 8\gamma)$.

Then, by comparing (A.6) with (A.8), we can approximate the equivalent binary input BICM channel as a BI-AWGN with SNR γ. Therefore, we can write,

$$B(\rho) = e^{-\gamma} \tag{A.10}$$

from where we obtain that the signal-to-noise ratio of the equivalent binary-input BICM channel is given by,

$$\gamma = \rho\alpha, \tag{A.11}$$

where

$$\alpha = -\frac{1}{\rho} \log B(\rho) \tag{A.12}$$

is the scaling factor with respect to the nominal SNR of the channel ρ.

It is not difficult to show that for $\rho \to \infty$, by simply taking a single dominant term in the numerator and denominator of the term under $\sqrt{\cdot}$ in (A.6) we have,

$$\alpha \approx -\frac{1}{\rho} \log \left(\frac{1}{M 2^M} \sum_{m=1}^{M} \sum_{b=0}^{1} \sum_{x \in \mathcal{X}_b^m} \exp\left(-\frac{1}{4} \rho d^2(x, x') \right) \right) \tag{A.13}$$

where for each $x \in \mathcal{X}_b^m$, the symbol x' is the point at minimum squared Euclidean distance $d^2(x, x')$ from x in the complement subset $\mathcal{X}_{\bar{b}}^m$. By letting $\rho \to \infty$ and using Varadhan integral lemma, we keep only the dominant term in the above sum and we get

$$\lim_{\rho \to \infty} \alpha = \frac{d_{\min}^2}{4} = \left(\frac{d_{\min}}{2} \right)^2 \tag{A.14}$$

where $\delta_{\min}^2$ is the minimum squared Euclidean distance of the constellation. Notice that this asymptotic value does not depend on the binary labeling rule. Also note that the term $\frac{d_{\min}}{2}$ represents the distance from one constellation point to the decision threshold corresponding to its nearest neighbor. Figure A.2 shows the SNR scaling α for BICM with 16-QAM with Gray and Set-Partitioning binary labelings as a function of ρ. As predicted by the analysis above, both mappings approach the asymptotic value $\alpha = \frac{d_{\min}^2}{4} = 0.1$. Notice however that Gray mapping shows a smaller scaling thus implying that the equivalent BICM channel is less noisy. This conclusion was already observed directly from the upper bounds with BICM-ML decoding in [36].

We suggest to replace the BICM equivalent channel by a BI-AWGN with scaled SNR $\gamma = \alpha\rho$. Therefore, any suitable bounding technique for binary codes over the BI-AWGN channel can be applied verbatim on the binary code underlying the BICM scheme. The resulting error probability bound will only depend on the SNR of the actual channel ρ, on the scaling factor α,which incorporates the effects of the signal constellation $\mathcal{X}$ and the binary labeling μ, and the weight distribution of the underlying binary code.

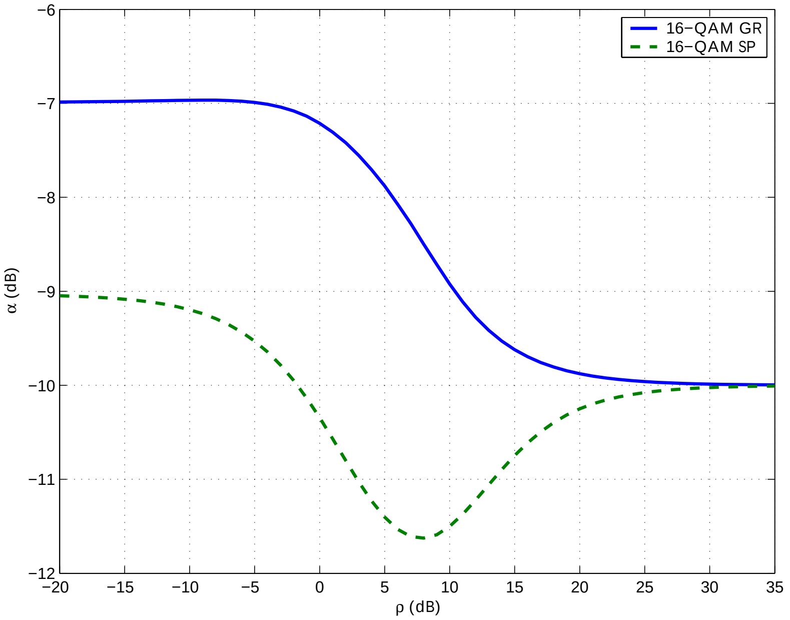


Figure A.2: Signal-to-noise ratio scaling of the equivalent BI-AWGN channel for BICM with 16-QAM and Gray and Set-Partitioning mapping.

In order to verify the validity of the proposed Gaussian approximation for the computation of error probabilities, in the following we report several numerical examples. We plot in Figures A.3, A.4 and A.5 the simulated pdfs of the log-likelihood ratio (A.3) given that a 0 was transmitted[1], denoted by LLR_0, for BICM with 16-QAM and Gray mapping at $\rho = 10$dB and $\rho = 20$dB. We also plot the corresponding Gaussian approximation of the equivalent BI-AWGN channel, i.e., a Gaussian distribution $\mathcal{N}(4\gamma, 8\gamma)$. In the case of $\rho = 10$ dB, $\gamma = 1.07$dB while when $\rho = 20$ dB, $\gamma = 10.16$dB. We observe that in both cases, the error probability behavior, i.e., $\Pr(\text{LLR}_0 < 0)$ is approximately the same for the BICM as for the corresponding Gaussian case, since tails of both distributions are almost identical.

A.5 Approximations on BICM-ML Error Probability

From the results in the previous section, we here recall some BICM-ML decoding error probability upper bounds for BICM based on the Gaussian approximation of the binary-input BICM equivalent channel. Consider the equivalent binary-input BICM AWGN channel with signal-to-noise ratio γ described in the previous

---

[1] Notice that the LLR given that a 1 was transmitted is completely symmetric, since the binary-input BICM channel is output symmetric (BIOS) according to the assumptions made in [36].

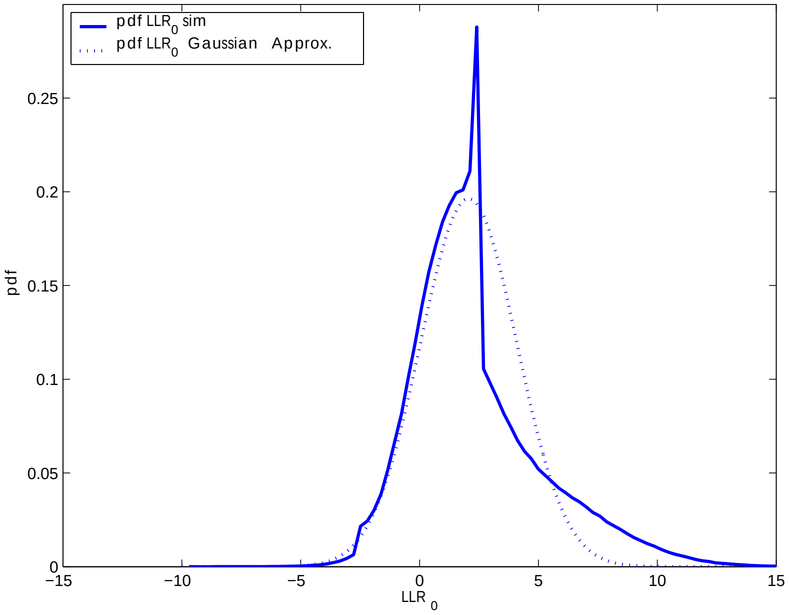


Figure A.3: Simulated pdf of the LLR given that a 0 was transmitted for BICM with 16-QAM and Gray mapping at $\rho = 5$dB.

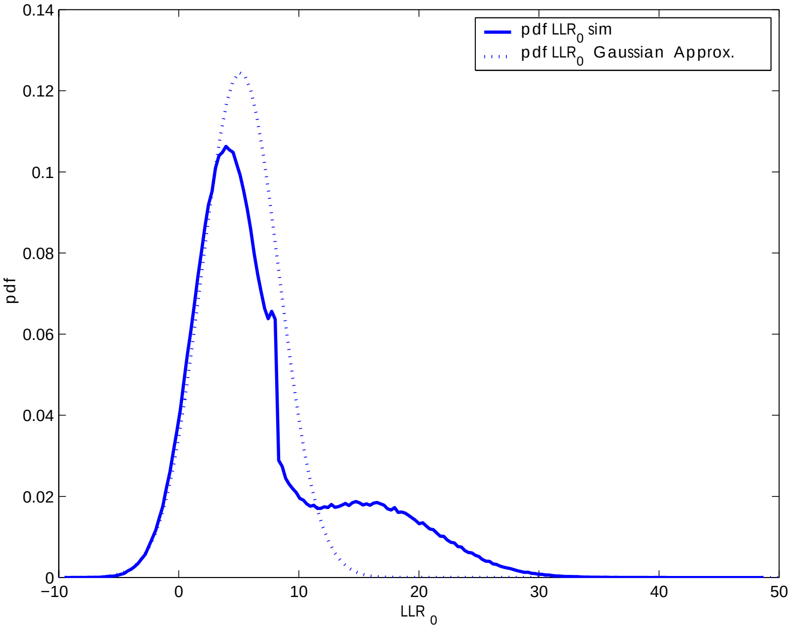


Figure A.4: Simulated pdf of the LLR given that a 0 was transmitted for BICM with 16-QAM and Gray mapping at $\rho = 10$dB.

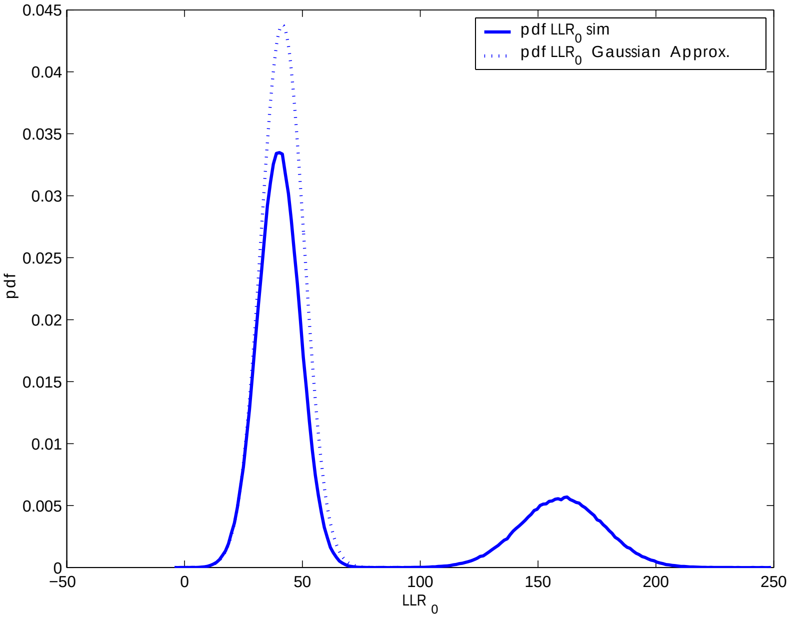


Figure A.5: Simulated pdf of the LLR given that a 0 was transmitted for BICM with 16-QAM and Gray mapping at $\rho = 20$dB.

section. Then, the union bound on the frame error probability is given by,

$$P_e \lessapprox \sum_w A_w Q\left(\sqrt{2w\rho\alpha}\right). \tag{A.15}$$

Notice that in order to compute the bit error probability, we should replace A_d by

$$B_w = \sum_i \frac{i}{K} A_{i,w},$$

where $A_{i,w}$ is the input-output WEF (IOWEF) of $\mathcal{C}$ [97].

Union bound-based techniques are known not to provide good estimates of the error probability of capacity-approaching codes over AWGN channels. On the contrary, improved bounding techniques such as the tangential-sphere bound (TSB) [98, 39] have been shown to provide very accurate results. In our case, the tangential-sphere bound is given by[2],

$$
\begin{aligned}
P_e \lessapprox \int_{-\infty}^{+\infty} \frac{dz_1}{\sqrt{2\pi\sigma^2}} e^{-z_1^2/2\sigma^2} \Bigg\{ 1 - \bar{\Gamma}\left(\frac{N-1}{2}, \frac{r_{z_1}}{2\sigma^2}\right) + \\
+ \sum_{w:\,\delta/2<\alpha_\delta} A_w \, \bar{\Gamma}\left(\frac{N-2}{2}, \frac{r_{z_1}^2 - \beta_\delta(z_1)^2}{2\sigma^2}\right) \cdot \qquad (A.16) \\
\cdot \left[Q\left(\frac{\beta_\delta(z_1)}{\sigma}\right) - Q\left(\frac{r_{z_1}}{\sigma}\right) \right] \Bigg\}
\end{aligned}
$$

where $\delta = 2\sqrt{w}$ is the Euclidean distance corresponding to a pairwise error event at Hamming distance w with unit energy, $\mathsf{R}^2 = N$ is the squared sphere radius,

$$
\sigma^2 = (2\rho\alpha)^{-1} \qquad (A.17)
$$

is the variance Gaussian noise corresponding to the equivalent binary-input BICM channel,

$$
\bar{\Gamma}(a, x) = \frac{1}{\Gamma(a)} \int_0^x t^{a-1} e^{-t} dt
$$

is the normalized incomplete gamma function and

$$
\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt
$$

is the gamma function,

$$
r_{z_1} = r\left(1 - \frac{z_1}{\mathsf{R}}\right),
$$

$$
\beta_\delta(z_1) = \frac{r_{z_1}}{\sqrt{1 - \frac{\delta^2}{4\mathsf{R}^2}}} \frac{\delta}{2\mathsf{r}},
$$

$$
\alpha_\delta = \mathsf{r}\sqrt{1 - \frac{\delta^2}{4\mathsf{R}^2}}
$$

and r, the cone radius, is the solution of

$$
\sum_{w:\,\delta/2<\alpha_\delta} A_w \int_0^{\theta_k} \sin^{N-3}\phi \, d\phi = \frac{\sqrt{\pi}\,\Gamma\left(\frac{N-2}{2}\right)}{\Gamma\left(\frac{N-1}{2}\right)} \qquad (A.18)
$$

with

$$
\theta_k = \cos^{-1}\left(\frac{\delta}{2\mathsf{r}} \frac{1}{\sqrt{1 - \frac{\delta^2}{4\mathsf{R}^2}}}\right). \qquad (A.19)
$$

---

[2] The tangential-sphere bound can be further improved for computing the bit-error probability using [99].

Again, notice that the integral in (A.16) can be efficiently computed using the Gauss-Hermite quadratures.

In Figure A.6 we illustrate the BICM-ML error probability approximations presented above for the 64 states rate 1/2 convolutional code with 16-QAM with Gray mapping, with a frame of $K = 128$ information bits. The overall spectral efficiency is $R = 2$ bit/s/Hz. We have used in all bounds the *truncated* bit-error distance spectrum of the code, i.e., we have considered all error events for which $w \leq 256$. We show the bounds for the bit-error probability and we compare them with the bit error rate simulation. We observe that the Bhattacharyya union bound is quite loose [36], while the union bound with the Gaussian approximation denoted by UB-GA is much tighter. Moreover, as expected, the TSB with the Gaussian approximation, denoted by TSB-GA, is the tightest and offers a better estimate in the low-SNR regime. This suggests that there is no loss in tightness if we use the Bhattacharyya bound, provided that we use it as a means of estimating the variance of an underlying Gaussian channel, and then use the standard bounding functions to estimate the error probability of the channel.

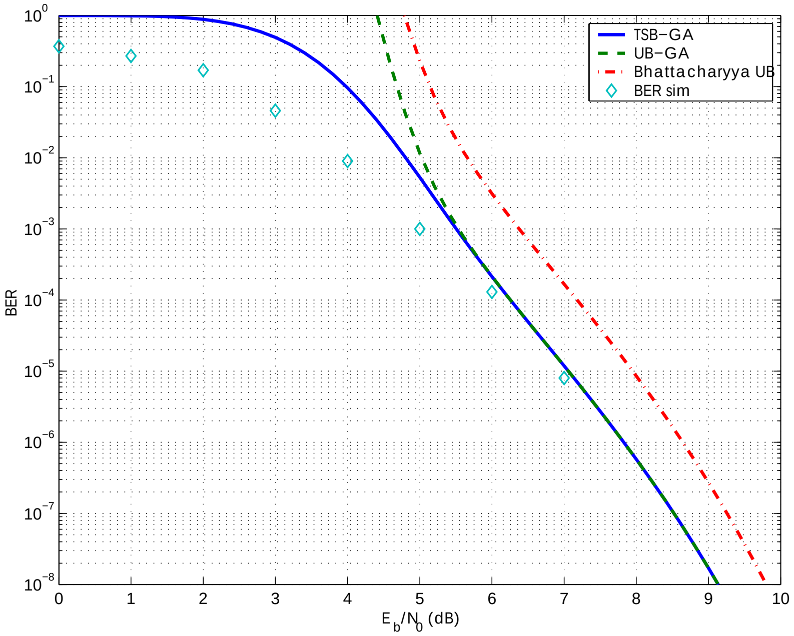


Figure A.6: BER ML BICM Bounds for the 64 states convolutional code of rate 1/2 with 16-QAM and Gray mapping.

Figures A.7 and A.8 illustrate the same BICM-ML bit error probability approximations and simulation for a repeat-accumulate (RA) code [48] of rate 1/4 with $K = 512$ and $K = 1024$ information bits respectively with 16-QAM with Gray mapping. The overall spectral efficiency is $R = 1$ bit/s/Hz. Notice that for such code ensembles, the weight enumerator can be computed explicitly [48, 53]. For the sake of comparison, in Figure A.7 we plot also the corresponding (true ML) bounds and simulation for the BPSK case. We observe the same behavior of the binary case, i.e., the TSB-GA yields a good estimate of the waterfall region while the UB-GA and the Bhattacharyya are only valid for estimating the error floor. Similar comments apply to Figure A.9, where we show the BER performance for the quasi-repeat and accumulate (QRA) code ensemble[3] with $K = 1024$ and 16-QAM with Gray mapping. The capacity at $R = 1$ bit/s/Hz is also shown. This constitutes a simple and yet accurate finite length analysis for turbo-like codes with BICM, since convergence analysis of iterative decoding can be very complicated task [96].

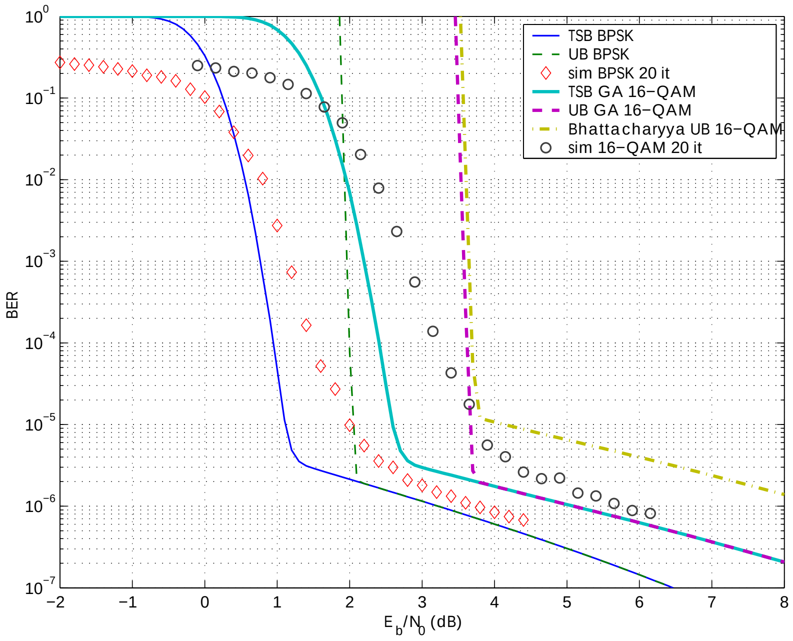


Figure A.7: BER ML BICM Bounds for the a repeat-and-accumulate code of rate 1/4 and $K = 512$ with 16-QAM with Gray mapping.

A.6 BICM-ML Thresholds for Turbo-coded BICM

In [41], the author proposed a tight upper bound on the ML decoding signal-to-noise ratio threshold γ_{th} for binary codes. For signal-to-noise ratios $\gamma > \gamma_{\text{th}}$

---

[3] We denote the quasi-repeat and accumulate (QRA) ensemble as the serially concatenated convolutional code ensemble of rate $r = 1/q$ that has as outer code generators (in octal form) $(\underbrace{1,\ldots,1}_{q-1},3)_8$ and inner accumulator.

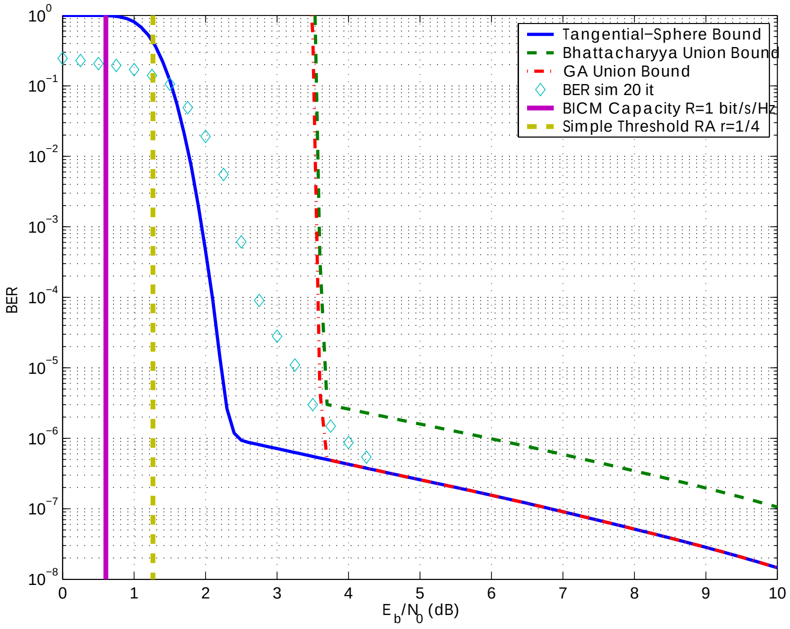


Figure A.8: BER ML BICM Bounds for the a RA code of rate $1/4$ and $K = 1024$ with 16-QAM with Gray mapping.

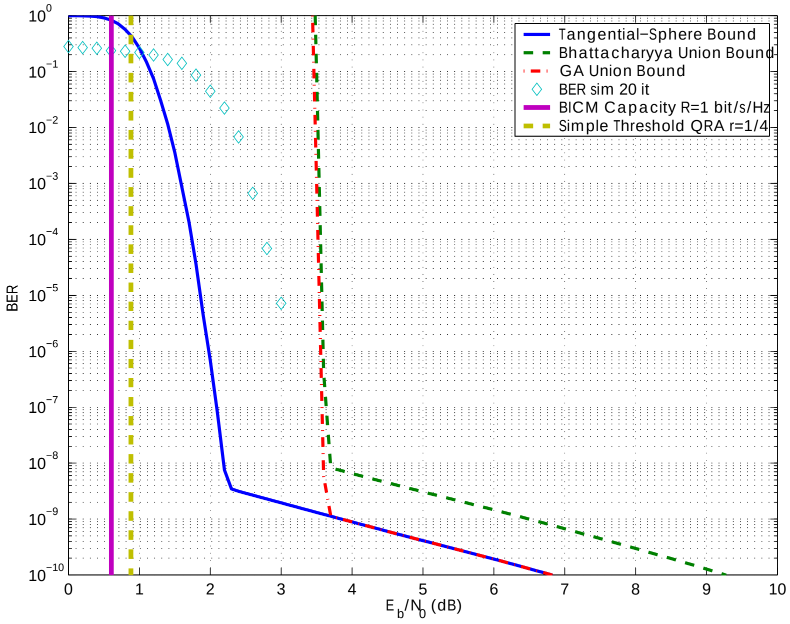


Figure A.9: BER ML BICM Bounds for the a QRA code of rate $1/4$ and $K = 1024$ with 16-QAM with Gray mapping.

the exponent of the simple bound of Divsalar is positive, and therefore, for large block length $P_e \to 0$ [41]. Following the footsteps of the previous sections, we can easily extend this result to the binary input BICM channel through the Gaussian approximation, which yields that,

$$\left|\frac{E_b}{N_0}\right|_{\mathrm{th}} \leq \frac{1}{\alpha R} \max_{0\leq\omega\leq 1} \left[\frac{(1-e^{-2a(\omega)})(1-\omega)}{2\omega}\right] \tag{A.20}$$

where $\omega = w/N$ is the normalized output Hamming weight, and

$$a(\omega) = \lim_{N\to\infty} \frac{1}{N} \log A_{\omega N}$$

is the asymptotic growth rate of the normalized weight enumerator A_ω of $\mathcal{C}$ [53]. For the binary RA and QRA code ensembles, table I in [53] reports the upper bounds on the ML $\left|\frac{E_b}{N_0}\right|_{\mathrm{th}}$ thresholds for BPSK modulation. As remarked in [53] the QRA ensemble significantly outperforms the RA ensemble as far as ML thresholds are concerned.

Based on the results obtained in the previous sections, we can easily establish the thresholds for BICM-ML decoding of turbo-coded BICM using the Gaussian approximation. For example, by performing simple computations, when we use BICM with 16-QAM with Gray mapping with a RA code of rate $r = 1/4$, as done in Figures A.7 and A.8, the BICM-ML with the Gaussian approximation threshold is $\left|\frac{E_b}{N_0}\right|_{\mathrm{th}} = 1.2648$ dB, while the BICM capacity with BICM-ML decoding (A.4) for 16-QAM with Gray mapping for $R = 1$ bit/s/Hz is at 0.6050 dB (see Figure A.1). For the sake of comparison, we show the BICM capacity limit and the simple threshold in Figures A.8 and A.9. Table A.6 summarizes the BICM-ML decoding simple bound thresholds (A.20) using the Gaussian approximation for the RA and QRA ensembles with 16-QAM with Gray mapping with corresponding spectral efficiencies of $R = 1, 2$ bit/s/Hz.

Rate R	Capacity (A.4)	RA	QRA
1 bit/s/Hz	0.6050 dB	1.2648 dB	0.8820 dB
2 bit/s/Hz	2.2671 dB	6.1512 dB	3.1409 dB

Table A.1: Upper bounds on the BICM-ML decoding $\left|\frac{E_b}{N_0}\right|_{\mathrm{th}}$ thresholds (A.20) using the Gaussian approximation for 16-QAM with Gray mapping compared to the BICM capacity limit (A.4).

Reference [100] provides some ML thresholds based on the union bound. For example, for the RA code ensemble of rate $r = 1/4$ with 16-QAM modulation with Gray mapping, the $\left|\frac{E_b}{N_0}\right|_{\mathrm{th}} = 5.91$ dB. As we can observe, the proposed thresholds are much tighter than those proposed in [100], due to the improved threshold based on the simple bound of [41] and to the accuracy of the Gaussian approximation of the binary-input BICM channel.

A.7 Conclusions

We have presented a very accurate and simple to compute approximation to the error probability of BICM under BICM-ML decoding using the Gaussian approximation of the binary-input BICM equivalent channel. We have verified the validity of the approximation to compute error probabilities and we have applied it to compute simple and accurate estimates of the error probability with BICM based on union and tangential-sphere bounds for both convolutional and turbo-like codes. We have also found accurate estimates of the BICM-ML decoding threshold. The key result is that very good approximation is given when the Bhattacharyya bound is used to estimate the variance of the underlying binary-input BICM equivalent channel. The proposed method constitutes a simple and very powerful tool for finite-length ML analysis of capacity approaching codes with BICM. Moreover, in [18] we have shown that the Gaussian approximation is a particular case within the general framework of saddle-point approximations of error probabilities in binary-input output-symmetric channels.

B

Multivariate Weight Enumerators

Throughout this thesis we have used extensively the multivariate weight enumerating function (MWEF) and the product distance weight enumerating function (PWEF) of binary codes partitioned into N_B blocks. In this Appendix we focus on their calculation.

B.1 MWEF of BCCs

To compute the MWEF, we assume that blockwise concatenation is performed through a set of N_B uniform interleavers [54, 55]. Then, we can compute the average multivariate weight enumeration function according to the following

Proposition 2 *Let $\mathcal{C}^{BCC}$ be a blockwise concatenated code mapped over N_B fading blocks constructed by concatenating an outer code $\mathcal{C}^O$ mapped over N_B blocks with input multivariate-output weight enumeration function $A^O_{i,w_1,\ldots,w_{N_B}}$, and N_B inner codes $\mathcal{C}^I$ with input-output weight enumeration functions $A^I_{i,w}$, through N_B uniform interleavers of length L_π. Then, the average input multivariate-output weight enumeration function of $\mathcal{C}^{BCC}$, $A^{BCC}_{i,w_1,\ldots,w_{N_B}}$, is given by,*

$$A^{BCC}_{i,w_1,\ldots,w_{N_B}} = \sum_{\ell_1,\ldots,\ell_{N_B}} \frac{A^O_{i,\ell_1,\ldots,\ell_{N_B}} \prod_{b=1}^{N_B} A^I_{\ell_b,w_b}}{\prod_{b=1}^{N_B} \binom{L_\pi}{\ell_b}}. \tag{B.1}$$

Proof. Define the Cartesian product code $\mathcal{C}_P = \mathcal{C}^I \times \ldots \times \mathcal{C}^I = (\mathcal{C}^I)^{N_B}$ with multivariate-input multivariate-output weight enumeration function $A^P_{i_1,\ldots,i_{N_B},w_1,\ldots,w_{N_B}}$. Then,

$$\begin{aligned} A^{BCC}_{i,w_1,\ldots,w_{N_B}} &= \mathbb{E}\left[A^P_{\ell_1,\ldots,\ell_{N_B},w_1,\ldots,w_{N_B}}\middle| i\right] \\ &= \sum_{\ell_1,\ldots,\ell_{N_B}} P(\ell_1,\ldots,\ell_{N_B}|i) A^P_{\ell_1,\ldots,\ell_{N_B},w_1,\ldots,w_{N_B}} \end{aligned} \tag{B.2}$$

where $\mathbb{E}[.]$ denotes expectation over all uniform interleavers and $P(\ell_1,\ldots,\ell_{N_B}|i)$ is the probability of having an error event of weights per block $\ell_1,\ldots,\ell_{N_B}$ at the input of $\mathcal{C}^P$ given the input weight i of $\mathcal{C}^O$ and it is given by,

$$P(\ell_1,\ldots,\ell_{N_B}|i) = \frac{A^O_{i,\ell_1,\ldots,\ell_{N_B}}}{\prod_{b=1}^{N_B}\binom{L_\pi}{\ell_b}}. \tag{B.3}$$

By construction,

$$A^P_{\ell_1,\ldots,\ell_{N_B},w_1,\ldots,w_{N_B}} = \prod_{b=1}^{N_B} A^I_{\ell_b,w_b} \tag{B.4}$$

and thus we obtain (B.1). □

Remarkably, in the case of RBAs the function $A^{RBA}_{i,w_1,\ldots,w_{N_B}}$ can be written in closed form in a way similar to [48] for RA codes in the AWGN channel. We obtain

$$A^{RBA}_{i,w_1,\ldots,w_{N_B}} = \frac{\prod_{b=1}^{N_B}\binom{L_\pi - w_b}{\lfloor i/2\rfloor}\binom{w_b-1}{\lceil i/2\rceil - 1}}{\binom{L_\pi}{i}^{N_B-1}} \tag{B.5}$$

The MWEF of the BCC is eventually obtained as

$$A_{w_1,\ldots,w_{N_B}} = \sum_i A_{i,w_1,\ldots,w_{N_B}}.$$

B.2 Asymptotic MWEF of BCCs

Following in the footsteps of [53], we can derive the asymptotic exponential growth rate of the MWEF $A^{BCC}_{i,w_1,\ldots,w_{N_B}}$. Let

$$F^O_{\iota,\delta_1,\ldots,\delta_{N_B}} \triangleq \frac{1}{Lr_I}\log A^O_{\iota K,\delta_1 L_\pi,\ldots,\delta_{N_B}L_\pi} \tag{B.6}$$

$$F^I_{\delta_b,\omega_b} \triangleq \frac{1}{L}\log A^I_{\delta_b L,\omega_b L}, \qquad b = 1,\ldots,N_B \tag{B.7}$$

$$F^{BCC}_{\iota,\omega_1,\ldots,\omega_{N_B}} \triangleq \frac{1}{L}\log A^{BCC}_{\iota K,\omega_1 L,\ldots,\omega_{N_B}L} \tag{B.8}$$

be the *normalized* weight enumerators of the outer, inner and BCC codes as a function of its corresponding exponential growth rates, where $r_I = L_\pi/L$ is the rate of the inner code, $\iota = i/K$ is the normalized input weight, $\delta_b = \ell_b/L_\pi$ for $b = 1, \ldots, N_B$ are the normalized output weights per block of the outer code and $\omega_b = w_b/L$ for $b = 1, \ldots, N_B$ are the normalized output weights per block of the BCC. Moreover,

$$\lim_{L_\pi \to \infty} \binom{L_\pi}{L_\pi \delta_b} = e^{L_\pi h(\delta)} = e^{L r_I h(\delta)} \tag{B.9}$$

where $h(p) = -p\log(p) - (1-p)\log(1-p)$ is the binary entropy function. Therefore, from (B.1) we find that for $L \to \infty$ the asymptotic exponential growth rate of the MWEF of the BCC is given by,

$$F^{BCC}_{\iota,\omega_1,\ldots,\omega_{N_B}} = \max_{\substack{0 \le \delta_b \le 1 \\ b=1,\ldots,N_B}} \left\{ r_I F^{O}_{\iota,\delta_1,\ldots,\delta_{N_B}} + \sum_{b=1}^{N_B} F^{I}_{\delta_b,\omega_b} - r_I \sum_{b=1}^{N_B} h(\delta_b) \right\} \tag{B.10}$$

Also, the output growth rate is given by,

$$a_{\omega_1,\ldots,\omega_{N_B}} = \max_{0 \le \iota \le 1} \left\{ F^{BCC}_{\iota,\omega_1,\ldots,\omega_{N_B}} \right\}. \tag{B.11}$$

Again, for the case of RBA codes, we can write explicitly the asymptotic input-output growth rate as,

$$F^{RBA}_{\iota,\omega_1,\ldots,\omega_{N_B}} = \sum_{b=1}^{N_B} \left[(1-\omega_b) h\left(\frac{\iota}{2(1-\omega_b)} \right) + \omega_b h\left(\frac{\iota}{2\omega_b} \right) \right] - (N_B - 1) h(\iota) \tag{B.12}$$

and thus, the asymptotic growth rate of the MWEF is given by,

$$a^{RBA}_{\omega_1,\ldots,\omega_{N_B}} = \max_{0 \le \iota \le \wp} \left\{ F^{RBA}_{\iota,\omega_1,\ldots,\omega_{N_B}} \right\} \tag{B.13}$$

where

$$\wp = \min_{b=1,\ldots,N_B} \{ 2(1-\omega_b), 2\omega_b \}.$$

As an example, in Fig. B.1 we plot the growth rate of the MWEF for a rate $r = 1/2$ RBA code of length 100 information bits mapped over $N_B = 2$ fading blocks, computed using (3.24) and (B.5). In Fig. B.2 we plot the asymptotic growth rate of the multivariate weight distribution for a rate $r = 1/2$ RBA code mapped over $N_B = 2$ fading blocks. Notice that, already for block length as shows as 100 information bits, the finite-length growth rate and its corresponding asymptotic counterpart are indeed very similar.

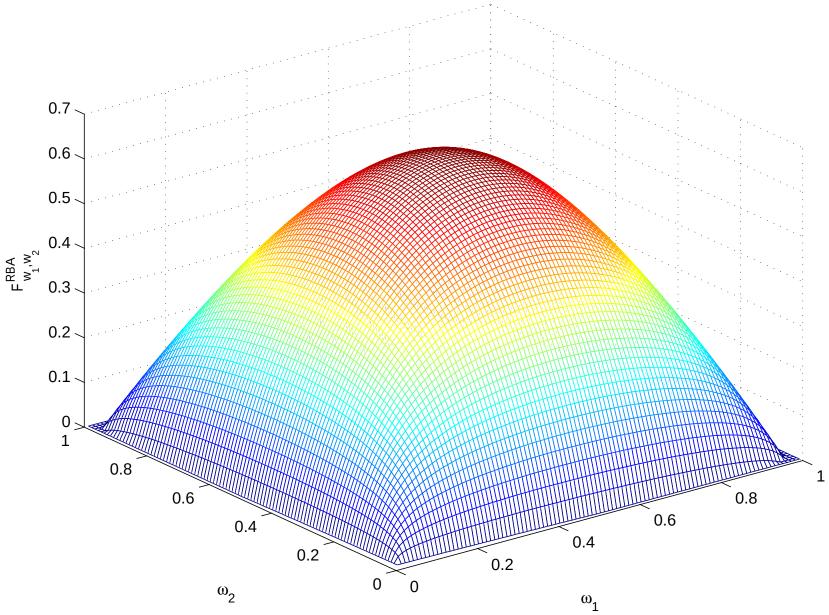


Figure B.1: Growth rate of the MWEF, $F^{RBA}_{\omega_1,\ldots,\omega_{N_B}}$ for a rate $r = 1/2$ RBA code of blocklength 100 information bits per frame mapped over $N_B = 2$ fading blocks.

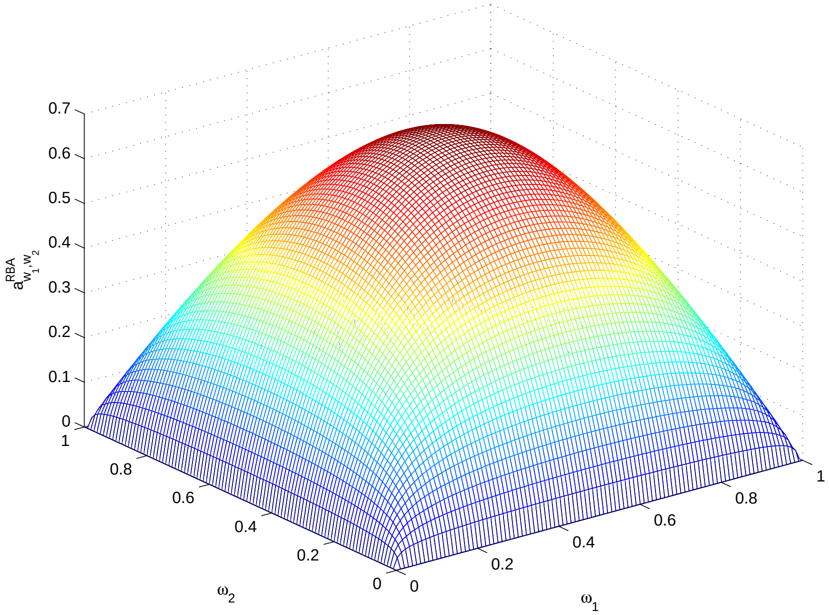


Figure B.2: Asymptotic growth rate $a^{RBA}_{\omega_1,\ldots,\omega_{N_B}}$ for a rate $r = 1/2$ RBA code mapped over $N_B = 2$ fading blocks.

B.3 PWEF of BCCs

In order to compute the approximated bounds based on [51], we need to compute the PWEF. In the following, we propose a general algorithm to compute the PWEF for BCCs.

```
for $\Delta_p = 1, \ldots, \Delta_p^{max}$
    PWEF[$\Delta_p$] = 0
end for

for $w_1 \leq \ldots \leq w_{N_B} \leq L$
    compute $\Delta_p = \prod_{b=1}^{N_B} w_b$
    if $\Delta_p \leq \Delta_p^{max}$
        compute $\mathcal{A} = \sum_{n=1}^{\mathcal{N}} A_{w_1(n),\ldots,w_{N_B}(n)}$
        PWEF[$\Delta_p$] = PWEF[$\Delta_p$] + $\mathcal{A}$
    end if
end for
```

where the multinomial coefficient

$$\mathcal{N} = \frac{N_B!}{n_1!n_2!\ldots n_p!} \tag{B.14}$$

is the number of permutations of N_B elements (the components of the weight vector $\mathbf{w} = (w_1, \ldots, w_{N_B})$), of which n_1 are alike, n_2 are alike,..., and n_p are alike, and $A_{w_1(n),\ldots,w_{N_B}(n)}$ denotes the MWEF corresponding to permutation n. For example, the number $\mathcal{N}$ corresponding to a weight vector $\mathbf{w} = (20, 22, 45, 45, 50, 50, 50, 50)$ with $N_B = 8$ is $\mathcal{N} = \frac{8!}{1!1!2!4!} = 840$.

It is clear from the above algorithm that the PWEF vector will have some positions equal to 0. In particular, the positions indexed with the prime numbers between L and Δ_p^{max} will store a 0. The number of prime numbers between 1 and n is $O\left(\frac{n}{\log(n)}\right)$ (see i.e. [101]). Therefore, in the limit for large n, the density of prime numbers in a vector of size n goes to zero. This implies that in our case, storing the PWEF vector as is, i.e., indexed by Δ_p is more practical than storing the PWEF and Δ_p in separate vectors thus doubling the memory and requiring more complicated indexing.

In the case of RBAs, the computation of the PWEF is greatly simplified due to the symmetry of its MWEF, since $A_{w_1(n),\ldots,w_{N_B}(n)} = A_{w_1,\ldots,w_{N_B}}$ for all permutations $n = 1, \ldots, \mathcal{N}$. In particular, steps 7 and 8 can be replaced by,

```
7        compute $A_{w_1,\ldots,w_{N_B}}$
8        PWEF[$\Delta_p$] = PWEF[$\Delta_p$] + $\mathcal{N}$ $A_{w_1,\ldots,w_{N_B}}$
```

Notice that with this algorithm we can easily truncate the PWEF if we set $\Delta_p^{max} < L^{N_B}$. The need for truncation is evident for large L and/or N_B.

C

Proofs

C.1 Proof of Lemma 1

Consider a family of codes for the block-fading channel (3.2) of given block length L, indexed by their operating SNR ρ, such that the ρ-th code has rate $R(\rho) = r \log \rho$ (in nat), where $r \in [0, 1]$, and WER (averaged over the channel fading) $P_e(\rho)$. Using Fano inequality [22] it is easy to show that $P_{\rm out}^G(\rho, R(\rho))$ yields an upper bound on the best possible SNR exponent. Recall that the fading power gains are defined as $\gamma_b = |h_b|^2$, for $b = 1, \ldots, N_B$, and are i.i.d. exponentially distributed. Following in the footsteps of [22] we define the normalized log-fading gains as $\alpha_b = -\log \gamma_b / \log \rho$. Hence, the joint distribution of the vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_{N_B})$ is given by

$$p(\boldsymbol{\alpha}) = (\log \rho)^{N_B} \exp\left(-\sum_{b=1}^{N_B} \rho^{-\alpha_b}\right) \rho^{-\sum_{b=1}^{N_B} \alpha_b} \tag{C.1}$$

The information outage event under Gaussian inputs is given by $\{\sum_{b=1}^{N_B} \log(1 + \rho\gamma_b) \leq N_B r \log \rho\}$. By noticing that $(1 + \rho\gamma_b) \doteq \rho^{[1-\alpha_b]_+}$, we can write the outage event as

$$\mathcal{A} = \left\{\boldsymbol{\alpha} \in \mathbb{R}^{N_B} \ : \ \sum_{b=1}^{N_B} [1 - \alpha_b]_+ \leq r N_B \right\} \tag{C.2}$$

The probability of outage is easily seen to satisfy the exponential equality

$$P_{\rm out}^G(\rho, R(\rho)) \doteq \int_{\mathcal{A} \cap \mathbb{R}_+^{N_B}} \rho^{-\sum_{b=1}^{N_B} \alpha_b} d\boldsymbol{\alpha} \tag{C.3}$$

Therefore, the SNR exponent of outage probability is given by the following limit

$$d_{\text{out}}(r) = -\lim_{\rho\to\infty} \frac{1}{\log(\rho)} \log \int_{\mathcal{A}\cap\mathbb{R}_+^{N_B}} \exp\left(-\log(\rho)\sum_{b=1}^{N_B}\alpha_b\right) d\boldsymbol{\alpha} \tag{C.4}$$

We apply Varadhan's integral lemma [102] and we obtain

$$d_{\text{out}}(r) = \inf_{\boldsymbol{\alpha}\in\mathcal{A}\cap\mathbb{R}_+^{N_B}} \left\{\sum_{b=1}^{N_B}\alpha_b\right\} \tag{C.5}$$

The constraint set is the intersection of the region defined by $\sum_{b=1}^{N_B}\alpha_b \geq N_B(1-r)$ and the region defined by $\alpha_b \in [0,1]$ for all $b = 1,\dots,N_B$. For all $r \in [0,1]$, the infimum in (C.5) is given by

$$d_{\text{out}}(r) = N_B(1-r) \tag{C.6}$$

In order to show that this exponent is actually the SNR reliability function for any $L \geq 1$, we have to prove achievability. We examine the average WER of a Gaussian random coding ensemble. Fix $L \geq 1$ and, for any SNR ρ, consider the ensemble generated with i.i.d. components with input probability $P_X = \mathcal{N}_{\mathbb{C}}(0,1)$ and rate $R(\rho) = r\log\rho$. The pairwise error probability under ML decoding, for two codewords $\mathbf{X}(0)$ and $\mathbf{X}(1)$ for given fading coefficients is upperbounded by the Chernoff bound

$$P(\mathbf{X}(0)\to\mathbf{X}(1)|\mathbf{H}) \leq \exp\left(-\frac{\rho}{4}\left\|\mathbf{H}(\mathbf{X}(0)-\mathbf{X}(1))\right\|_F^2\right) \tag{C.7}$$

By averaging over the random coding ensemble and using the fact that the entries of the matrix difference $\mathbf{X}(0)-\mathbf{X}(1)$ are i.i.d. $\sim \mathcal{N}_{\mathbb{C}}(0,2)$, we obtain

$$\overline{P(\mathbf{X}(0)\to\mathbf{X}(1)|\mathbf{H})} \leq \prod_{b=1}^{N_B}\left[1+\frac{1}{2}\rho\gamma_b\right]^{-L} \doteq \rho^{-L\sum_{b=1}^{N_B}[1-\alpha_b]_+} \tag{C.8}$$

(in general, the bar denotes quantities averaged over the code ensemble). By summing over all $\rho^{rN_BL}-1$ messages $w \neq 0$, we obtain the ensemble average union bound

$$\overline{P_e(\rho|\mathbf{H})} \;\dot{\leq}\; \rho^{-L\sum_{b=1}^{N_B}[1-\alpha_b]_+ + LN_Br} \tag{C.9}$$

Next, we upperbound the ensemble average WER by separating the outage event from the non-outage event (the complement set denoted by $\mathcal{A}^c$) as follows:

$$\overline{P_e(\rho)} \leq \Pr(\mathcal{A}) + \overline{\Pr(\text{error},\mathcal{A}^c)} \tag{C.10}$$

Achievability is proved if we can show that $\overline{\Pr(\text{error},\mathcal{A}^c)} \;\dot{\leq}\; \rho^{-N_B(1-r)}$ for all $L \geq 1$ and $r \in [0,1]$. We have

$$\overline{\Pr(\text{error},\mathcal{A}^c)} \;\dot{\leq}\; \int_{\mathcal{A}^c\cap\mathbb{R}_+^{N_B}} \exp\left(-\log(\rho)\left(\sum_{b=1}^{N_B}\alpha_b + L\left(\sum_{b=1}^{N_B}[1-\alpha_b]_+ - rN_B\right)\right)\right) \tag{C.11}$$

By using again Varadhan integral lemma we obtain the lower bound on the random coding exponent

$$d_B^{(r)}(r) \geq \inf_{\boldsymbol{\alpha} \in \mathcal{A}^c \cap \mathbb{R}_+^{N_B}} \left\{ \sum_{b=1}^{N_B} \alpha_b + L \left(\sum_{b=1}^{N_B} [1-\alpha_b]_+ - rN_B \right) \right\} \tag{C.12}$$

where $\mathcal{A}^c$ is defined explicitly by

$$\sum_{b=1}^{N_B} [1-\alpha_b]_+ \geq rN_B$$

It is easily seen that for all $L \geq 1$ and $r \in [0,1]$ the infimum is obtained[1] by

$$\begin{aligned} \alpha_b &= 1, \quad \text{for } b = 1, \ldots, k-1 \\ \alpha_k &= 1 + \lfloor rN_B \rfloor - rN_B \\ \alpha_b &= 0, \quad \text{for } b = k+1, \ldots, N_B \end{aligned} \tag{C.13}$$

where $k = N_B - \lfloor rN_B \rfloor$, and yields $d_B^{(r)}(r) \geq N_B(1-r)$. Since this lower bound coincides with the outage probability upper bound (C.6), we obtain that $d_B^\star(r) = N_B(1-r)$ and it is achieved by Gaussian codes for any $L \geq 1$. Any fixed coding rate R corresponds to the case $r = 0$, from which the statement of Lemma 1 follows.

---

[1] This solution is not unique. Any configuration of the variables α_b having $k-1$ variables equal to 1, $N_B - k - 1$ variables equal to 0 and one variable equal to $1 + \lfloor rN_B \rfloor - rN_B$ yields the same value of the infimum. Moreover, for $L = 1$ also the solution $\alpha_b = 0$ for all b yields the same value.

C.2 Proof of Theorem 1

We fix the number of fading blocks N_B, the coding rate R and the (unit energy) modulation signal set $\mathcal{X}$. Using Fano inequality [22] it is easy to show that $P_{\text{out}}^{\mathcal{X}}(\rho, R)$ yields an upper bound on the best possible SNR exponent attained by coded modulations with signal set $\mathcal{X}$. The outage probability with discrete inputs is lower bounded by

$$\begin{aligned}
P_{\text{out}}^{\mathcal{X}}(\rho, R) &\overset{\text{a}}{\geq} \Pr\left(\frac{1}{N_B}\sum_{b=1}^{N_B} J_{\mathcal{X}}(\gamma_b\rho) < R\right) \\
&= \Pr\left(\frac{1}{N_B}\sum_{b=1}^{N_B}\left(M - \frac{1}{2^M}\sum_{x\in\mathcal{X}}\mathbb{E}\left[\log_2\sum_{x'\in\mathcal{X}} e^{-|\sqrt{\rho\gamma_b}(x-x')+Z|^2+|Z|^2}\right]\right) < R\right) \\
&= \Pr\left(\frac{1}{N_B}\sum_{b=1}^{N_B}\left(M - \frac{1}{2^M}\sum_{x\in\mathcal{X}}\mathbb{E}\left[\log_2\sum_{x'\in\mathcal{X}} e^{-\left|\sqrt{\rho^{1-\alpha_b}}(x-x')+Z\right|^2+|Z|^2}\right]\right) < R\right)
\end{aligned}$$

where $Z \sim \mathcal{N}_{\mathbb{C}}(0,1)$, and the last equality follows just from the definition of the normalized log-fading gains $\alpha_b = -\log\gamma_b/\log\rho$. The inequality (a) is due to the fact that we have a strict inequality in the probability on the right. Since the term,

$$0 \leq \log_2\left(\sum_{x'\in\mathcal{X}} e^{-\left|\sqrt{\rho^{1-\alpha_b}}(x-x')+z\right|^2+|z|^2}\right) \leq \log_2\left(|\mathcal{X}|e^{|z|^2}\right) = \frac{\log|\mathcal{X}| + |z|^2}{\log 2} \tag{C.14}$$

and

$$\mathbb{E}\left[\frac{\log|\mathcal{X}| + |z|^2}{\log 2}\right] < \infty \tag{C.15}$$

we can apply the dominated convergence theorem [103], for which,

$$\lim_{\rho\to\infty}\mathbb{E}\left[\log_2\left(\sum_{x'\in\mathcal{X}} e^{-\left|\sqrt{\rho^{1-\alpha_b}}(x-x')+Z\right|^2+|Z|^2}\right)\right] = \tag{C.16}$$

$$= \mathbb{E}\left[\lim_{\rho\to\infty}\log_2\left(\sum_{x'\in\mathcal{X}} e^{-\left|\sqrt{\rho^{1-\alpha_b}}(x-x')+Z\right|^2+|Z|^2}\right)\right]. \tag{C.17}$$

Since for all $z \in \mathbb{C}$ with $|z| \leq \infty$ and $s \neq 0$ we have that

$$\lim_{\rho\to\infty} e^{-\left|\sqrt{\rho^{1-\alpha_b}}s+z\right|^2+|z|^2} = \begin{cases} 0 & \text{for} \quad \alpha_b < 1 \\ 1 & \text{for} \quad \alpha_b > 1 \end{cases}$$

we have that, for large ρ and $\alpha_b < 1$, $J_{\mathcal{X}}(\rho^{1-\alpha_b}) \to M$ and for $\alpha_b > 1$, $J_{\mathcal{X}}(\rho^{1-\alpha_b}) \to 0$. Hence, for every $\epsilon > 0$, we have the lower bound

$$\begin{aligned} P_{\text{out}}(\rho, R) &\dot{\geq} \Pr\left(\frac{1}{N_B}\sum_{b=1}^{N_B} \mathbb{1}\{\alpha_b \leq 1+\epsilon\} < \frac{R}{M}\right) \\ &\doteq \int_{\mathcal{A}_\epsilon \cap \mathbb{R}_+^{N_B}} \rho^{-\sum_{b=1}^{N_B}\alpha_b} d\boldsymbol{\alpha} \end{aligned} \tag{C.18}$$

where we define the event

$$\mathcal{A}_\epsilon = \left\{\boldsymbol{\alpha} \in \mathbb{R}^{N_B} \ : \ \frac{1}{N_B}\sum_{b=1}^{N_B} \mathbb{1}\{\alpha_b \leq 1+\epsilon\} < \frac{R}{M}\right\} \tag{C.19}$$

Using Varadhan's lemma, we get the upperbound to the SNR reliability function as

$$d_B^\star(R) \leq \inf_{\epsilon > 0} \inf_{\boldsymbol{\alpha} \in \mathcal{A}_\epsilon \cap \mathbb{R}_+^{N_B}} \left\{\sum_{b=1}^{N_B} \alpha_b\right\} \tag{C.20}$$

It is not difficult to show that the $\boldsymbol{\alpha}$ achieving the inner infimum in (C.20) is given by

$$\begin{aligned} \alpha_b &= 1+\epsilon, \quad \text{for } b = 1, \ldots, N_B - k \\ \alpha_b &= 0, \quad \text{for } b = N_B - k + 1, \ldots, N_B \end{aligned} \tag{C.21}$$

where $k = 0, \ldots, N_B - 1$ is the unique integer satisfying $\frac{k}{N_B} < \frac{R}{M} \leq \frac{k+1}{N_B}$. Since this holds for each $\epsilon > 0$, we can make the bound as tight as possible by letting $\epsilon \downarrow 0$. We obtain $d_{SB} = N_B - k$ which is equivalent to $d_{SB}(R)$ defined in (3.10).

In order to show that this exponent is actually the SNR reliability function, we have to prove achievability. We examine the average WER of the coded modulation ensemble obtained by concatenating a random binary linear code $\mathcal{C}$ with the signal set $\mathcal{X}$ via an arbitrary fixed one-to-one mapping $\mu : \mathbb{F}_2^M \to \mathcal{X}$. The random binary linear code and the one-to-one mapping induce a uniform i.i.d. input distribution over $\mathcal{X}$. The pairwise error probability under ML decoding, for two codewords $\mathbf{X}(0)$ and $\mathbf{X}(1)$ for given fading coefficients is again upperbounded by the Chernoff bound (C.7). By averaging over the random coding ensemble and using the fact that the entries of each codeword $\mathbf{X}(0)$ and $\mathbf{X}(1)$ are i.i.d. uniformly distributed over $\mathcal{X}$, we obtain

$$\overline{P(\mathbf{X}(0) \to \mathbf{X}(1)|\mathbf{H})} \leq \prod_{b=1}^{N_B} B_b^L \tag{C.22}$$

where we define the Bhattacharyya coefficient [97]

$$B_b = 2^{-2M} \sum_{x \in \mathcal{X}} \sum_{x' \in \mathcal{X}} \exp\left(-\frac{\rho}{4}\gamma_b |x - x'|^2\right) \tag{C.23}$$

By summing over all $2^{N_B LR} - 1$ messages $w \neq 0$, we obtain the union bound

$$\overline{P_e(\rho|\mathbf{H})} \leq \exp\left(-N_B LM \log(2) G(\rho, \boldsymbol{\alpha})\right) \tag{C.24}$$

where

$$G(\rho, \boldsymbol{\alpha}) \triangleq 1 - \frac{R}{M} - \frac{1}{N_B M} \sum_{b=1}^{N_B} \log_2 \left(1 + 2^{-M} \sum_{x \neq x'} e^{-\frac{1}{4}|x-x'|^2 \rho^{1-\alpha_b}} \right) \tag{C.25}$$

We notice a fundamental difference between the above union bound for random coding with discrete inputs and the corresponding union bound for random coding with Gaussian inputs given in (C.9). With Gaussian inputs, we have that the union bound vanishes for any finite block length L as $\rho \to \infty$. On the contrary, with discrete inputs the union bound (C.24) is bounded away from zero for any finite L as $\rho \to \infty$. This is because with discrete inputs the probability that two codewords coincide (i.e., have zero Euclidean distance) is positive for any finite L. Hence, it is clear that in order to obtain a non-zero random coding SNR exponent we have to consider a code ensemble with block length $L = L(\rho)$, increasing with ρ. To this purpose, we define

$$\beta \triangleq \lim_{\rho \to \infty} \frac{L(\rho)}{\log \rho}. \tag{C.26}$$

By averaging over the fading and using the fact that error probability cannot be larger than 1, we obtain

$$\overline{P_e(\rho)} \dot{\leq} \int_{\mathbb{R}_+^{N_B}} \rho^{-\sum_{b=1}^{N_B} \alpha_b} \min\left\{1, \exp\left(-N_B L(\rho) M \log(2) G(\rho, \boldsymbol{\alpha})\right)\right\} \, d\boldsymbol{\alpha} \tag{C.27}$$

We notice that

$$\lim_{\rho \to \infty} \log_2 \left(1 + 2^{-M} \sum_{x \neq x'} e^{-\frac{1}{4}|x-x'|^2 \rho^{1-\alpha_b}} \right) = \left\{ \begin{array}{lll} 0 & \text{for} & \alpha_b < 1 \\ M & \text{for} & \alpha_b > 1 \end{array} \right.$$

Hence, for $\epsilon > 0$, a lower bound on the random coding SNR exponent can be obtained by replacing $G(\rho, \boldsymbol{\alpha})$ by

$$\widetilde{G}_\epsilon(\boldsymbol{\alpha}) = 1 - \frac{R}{M} - \frac{1}{N_B} \sum_{b=1}^{N_B} \mathbb{1}\{\alpha_b \geq 1 - \epsilon\} \tag{C.28}$$

We define the event

$$\mathcal{B}_\epsilon = \left\{ \boldsymbol{\alpha} : \widetilde{G}_\epsilon(\boldsymbol{\alpha}) \leq 0 \right\} \tag{C.29}$$

Hence, from (C.27) and what said above we obtain

$$\begin{aligned}\overline{P_e(\rho)} \quad &\dot{\le} \int_{\mathcal{B}_\epsilon\cap\mathbb{R}_+^{N_B}} \rho^{-\sum_{b=1}^{N_B}\alpha_b}\, d\boldsymbol{\alpha} + \\ + \quad & \int_{\mathcal{B}_\epsilon^c\cap\mathbb{R}_+^{N_B}} \exp\left(-\log(\rho)\left[\sum_{b=1}^{N_B}\alpha_b + N_B\beta M\log(2)\widetilde{G}_\epsilon(\boldsymbol{\alpha})\right]\right) d\boldsymbol{\alpha}\end{aligned} \tag{C.30}$$

By applying again Varadhan's lemma we obtain a lower bound to the random coding SNR exponent given by

$$d_B^{(r)}(R) \ge \sup_{\epsilon>0}\min\left\{\inf_{\boldsymbol{\alpha}\in\mathcal{B}_\epsilon\cap\mathbb{R}_+^{N_B}}\left\{\sum_{b=1}^{N_B}\alpha_b\right\}, \inf_{\boldsymbol{\alpha}\in\mathcal{B}_\epsilon^c\cap\mathbb{R}_+^{N_B}}\left\{\sum_{b=1}^{N_B}\alpha_b + N_B\beta M\log(2)\widetilde{G}_\epsilon(\boldsymbol{\alpha})\right\}\right\} \tag{C.31}$$

It is not difficult to show that the $\boldsymbol{\alpha}$ achieving the first infimum in (C.31) is given by

$$\begin{aligned}\alpha_b &= 1-\epsilon, \quad \text{for } b = 1,\ldots,N_B-k \\ \alpha_b &= 0, \quad \text{for } b = N_B-k+1,\ldots,N_B\end{aligned} \tag{C.32}$$

where $k = 0,\ldots,N_B-1$ is the unique integer satisfying $\frac{k}{N_B} \le \frac{R}{M} < \frac{k+1}{N_B}$.

For the second infimum in (C.31) we can rewrite the argument in the form

$$N_B\beta M\log(2)\left(1-\frac{R}{M}\right) + \sum_{b=1}^{N_B}\left[\alpha_b - \beta M\log(2)\mathbb{1}\left\{\alpha_b \ge 1-\epsilon\right\}\right] \tag{C.33}$$

We distinguish two cases. If $0 \le \beta M\log(2) < 1$, then

$$\alpha_b - \beta M\log(2)\mathbb{1}\left\{\alpha_b \ge 1-\epsilon\right\} \tag{C.34}$$

attains its minimum at $\alpha_b = 0$. Hence, we obtain that the infimum is given by $N_B\beta M\log(2)\left(1-\frac{R}{M}\right)$. If $\beta M\log(2) \ge 1$, then (C.34) attains its absolute minimum at $\alpha_b = 1-\epsilon$, and its second smallest minimum at $\alpha_b = 0$. The number of terms α_b that can be made equal to $1-\epsilon$ subject to the constraint $\boldsymbol{\alpha} \in \mathcal{B}_\epsilon^c\cap\mathbb{R}_+^{N_B}$ is given by $\lfloor N_B(1-R/M)\rfloor$. Hence, the infimum is given by

$$N_B\beta M\log(2)\left(1-\frac{R}{M}\right) + (1-\epsilon-\beta M\log(2))\left\lfloor N_B\left(1-\frac{R}{M}\right)\right\rfloor \tag{C.35}$$

Both the first and the second infima are simultaneously maximized by letting $\epsilon \downarrow 0$. By collecting all results together, we obtain that the random coding SNR exponent is lower bounded by (3.11).

The random coding SNR exponent of the associated BICM channel is immediately obtained by using again the Bhattacharyya union bound [36]. In particular, for two randomly generated codewords $\mathbf{C}(0), \mathbf{C}(1) \in \mathcal{C}$

$$\overline{P(\mathbf{C}(0) \to \mathbf{C}(1)|\mathbf{H})} \leq 2^{-LMN_B} \prod_{b=1}^{N_B} (1 + B_b(\rho, \mu, \mathcal{X}))^{LM} \tag{C.36}$$

where $B_b(\rho, \mu, \mathcal{X})$ is defined in (3.19). The error probability averaged over the random BICM ensemble can be upperbounded by

$$\begin{aligned} \overline{P_e(\rho|\mathbf{H})} &\leq \exp\left(-N_B LM \log(2) \left[1 - \frac{R}{M} - \frac{1}{N_B} \sum_{b=1}^{N_B} \log_2\left(1 + B_b(\rho, \mu, \mathcal{X})\right)\right]\right) \\ &= \exp\left(-N_B LM \log(2) G'(\rho, \boldsymbol{\alpha})\right) \end{aligned} \tag{C.37}$$

It is not difficult to see that

$$\lim_{\rho \to \infty} \log_2\left(1 + B_b(\rho, \mu, \mathcal{X})\right) = \begin{cases} 0 & \text{for} \quad \alpha_b < 1 \\ 1 & \text{for} \quad \alpha_b > 1 \end{cases}$$

Hence, for $\epsilon > 0$, a lower bound on the random coding SNR exponent can be obtained by replacing $G'(\rho, \boldsymbol{\alpha})$ by $\widetilde{G}_\epsilon(\boldsymbol{\alpha})$ defined in (C.28). It then follows that the random coding SNR exponent of the associated BICM channel satisfies the same lower bound of the original block-fading channel.

C.3 Proofs for section 3.4.3

Proof of Theorem 2. Let $F(\boldsymbol{\omega})$ and $a(\boldsymbol{\omega})$ denote the exponential growth rate and the asymptotic exponential growth rate of the considered code ensemble over the BPSK modulation. For each length L_B, the conditional error probability given the fading vector $\mathbf{h}$ can be upperbounded by the Bhattacharyya union bound as

$$\begin{aligned} P_e(\rho|\mathbf{H}) &\leq \sum_{\boldsymbol{\omega}\in[0,1]^{N_B}} \exp\left(-L_B\left(\rho\sum_{b=1}^{N_B}\omega_b\gamma_b - F(\boldsymbol{\omega})\right)\right) \\ &= \exp\left(-L_B G_{L_B}(\rho,\boldsymbol{\gamma})\right) \end{aligned} \tag{C.38}$$

where we define the fading power gain vector $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_{N_B})$ and let

$$G_{L_B}(\rho,\boldsymbol{\gamma}) \triangleq -\frac{1}{L_B}\log\left\{\sum_{\boldsymbol{\omega}\in[0,1]^{N_B}} \exp\left(-L_B\left(\rho\sum_{b=1}^{N_B}\omega_b\gamma_b - F(\boldsymbol{\omega})\right)\right)\right\}$$

The limit of the above exponent for $L_B \to \infty$ is given by

$$G_\infty(\rho,\boldsymbol{\gamma}) = \inf_{\boldsymbol{\omega}\in[0,1]^{N_B}} \left\{\rho\sum_{b=1}^{N_B}\omega_b\gamma_b - a(\boldsymbol{\omega})\right\} \tag{C.39}$$

Let $\gamma_0 > 0$ be an arbitrary small quantity. We define the set

$$\Gamma_{\gamma_0,d_{SB}(R)} = \left\{\boldsymbol{\gamma}\in\mathbb{R}_+^{N_B} \quad : \quad \sum_{b=1}^{N_B}\mathbb{1}\{\gamma_b \leq \gamma_0\} \geq d_{SB}(R)\right\} \tag{C.40}$$

The set $\Gamma_{\gamma_0,d_{SB}(R)}$ contains the "bad fading vectors", i.e., the fading vectors having $d_{SB}(R)$ or more small components (smaller than γ_0).

Using fading statistical independence, the property of the fading cdf $\Pr(\gamma_b \leq \gamma_0) \triangleq p_0 = \gamma_0 + o(\gamma_0)$ (e.g., valid for Rayleigh fading) and standard bounds on the tail of the binomial distribution, we can write

$$\begin{aligned} \Pr(\Gamma_{\gamma_0,d_{SB}(R)}) &= \sum_{k=d_{SB}(R)}^{N_B}\binom{N_B}{k}p_0^k(1-p_0)^{N_B-k} \\ &\leq \mathcal{K}_1\gamma_0^{d_{SB}(R)} + o(\gamma_0^{d_{SB}(R)}) \end{aligned} \tag{C.41}$$

for a suitable constant $\mathcal{K}_1 > 0$.

Using (C.38), (C.39), (C.40) and (C.41), indicating by $f_{\boldsymbol{\gamma}}(\mathbf{z})$ the pdf of $\boldsymbol{\gamma}$ we can write, for sufficiently large ρ and L_B,

$$\begin{aligned} P_e(\rho) &\leq \mathbb{E}[\min\{1, \exp(-L_B G_{L_B}(\rho,\boldsymbol{\gamma}))\}] \\ &\leq \int_{\Gamma_{\gamma_0,d_{SB}(R)}} f_{\boldsymbol{\gamma}}(\mathbf{z})d\mathbf{z} + \int_{\Gamma^c_{\gamma_0,d_{SB}(R)}} e^{-L_B G_{L_B}(\rho,\boldsymbol{\gamma})} f_{\boldsymbol{\gamma}}(\mathbf{z})d\mathbf{z} \\ &\leq \mathcal{K}_1\gamma_0^{d_{SB}(R)} + \int_{\Gamma^c_{\gamma_0,d_{SB}(R)}} e^{-L_B G_\infty(\rho,\boldsymbol{\gamma})} f_{\boldsymbol{\gamma}}(\mathbf{z})d\mathbf{z} + \epsilon_{L_B} + o(\gamma_0^{d_{SB}(R)}) \end{aligned} \tag{C.42}$$

where $\epsilon_{L_B} \to 0$ as $L_B \to \infty$.

Recall that $\mathcal{W}(N_B, k) \subset \{0,1\}^{N_B}$ is the set of binary vectors of length N_B with Hamming weight not smaller than $N_B - k + 1$. For any $\boldsymbol{\gamma} \in \Gamma^c_{\gamma_0, d_{SB}(R)}$ there exists $\mathbf{x} \in \mathcal{W}(N_B, d_{SB}(R))$ such that $\gamma_0 \mathbf{x} \leq \boldsymbol{\gamma}$ (componentwise). In fact, it is sufficient to replace each $\gamma_b < \gamma_0$ in $\boldsymbol{\gamma}$ by 0, and each $\gamma_b \geq \gamma_0$ in $\boldsymbol{\gamma}$ by γ_0. Since by definition there are at most $d_{SB}(R) - 1$ components of $\boldsymbol{\gamma}$ smaller than γ_0, the resulting vector has Hamming weight at least $N_B - d_{SB}(R) + 1$ and therefore (up to the scaling by γ_0) it belongs to $\mathcal{W}(N_B, d_{SB}(R))$.

For $\boldsymbol{\gamma} \in \Gamma^c_{\gamma_0, d_{SB}(R)}$, it follows from the observation above that $G_\infty(\rho, \boldsymbol{\gamma})$ is lowerbounded by

$$G_\infty(\rho, \boldsymbol{\gamma}) \geq \inf_{\mathbf{x} \in \mathcal{W}(d_{SB}(R))} \inf_{\boldsymbol{\omega} \in [0,1]^{N_B}} \left\{ \rho \gamma_0 \sum_{b=1}^{N_B} x_b \omega_b - a(\boldsymbol{\omega}) \right\} \triangleq \widetilde{G}_\infty(\rho, \gamma_0) \quad \text{(C.43)}$$

Define $\widehat{s}$ as in Theorem 2 (eq. (3.26)) and suppose that, for the given ensemble, $\widehat{s}$ is finite. Then, we let $\gamma_0 = \widehat{s}/\rho$ and, continuing the chain of inequalities (C.42),

$$\begin{aligned} P_e(\rho) &\leq \mathcal{K}_1 \gamma_0^{d_{SB}(R)} + \exp(-L_B \widetilde{G}_\infty(\rho, \gamma_0)) \int_{\Gamma^c_{\gamma_0, d_{SB}(R)}} f_{\boldsymbol{\gamma}}(\mathbf{z}) d\mathbf{z} + \epsilon_{L_B} + o(\gamma_0^{d_{SB}(R)}) \\ &\leq \mathcal{K}_1 \left(\frac{\widehat{s}}{\rho} \right)^{d_{SB}(R)} + \epsilon'_{L_B} + o(\rho^{-d_{SB}(R)}) \\ &\leq \mathcal{K} \rho^{-d_{SB}(R)} + \epsilon'_{L_B} + o(\rho^{-d_{SB}(R)}) \end{aligned} \quad \text{(C.44)}$$

where $\epsilon'_{L_B} \to 0$ as $L_B \to \infty$. This shows the weak goodness of the ensemble.

Proof of Corollary 2. For ensembles of BICM codes with random bit interleaving we use the Bhattacharyya union bound with Bhattacharyya factor $B_b(\rho, \mu, \mathcal{X})$ defined in (3.19). Following the same approach as in the proof of Theorem 2, we see that the upper bound to the error probability takes on the same form of (C.42) – (C.44) provided that we replace γ_b by ζ_b defined in (3.20). For large ρ, we have that $\zeta_b \to \frac{d^2_{\min}}{4} \gamma_b$ where $d_{\min}$ is the minimum Euclidean distance of the normalized signal constellation $\mathcal{X}$ (see (3.20) and [17]). The scaling by the constant factor $d^2_{\min}/4$ (independent of ρ and of L) involves at most a scaling of the SNR threshold $\widehat{s}$. Therefore, weak goodness of the underlying binary code implies weak goodness of the corresponding BICM ensemble with random bit interleaving.

Trellis terminated CCs do not satisfy Theorem 2. For simplicity, we consider trellis terminated CCs of rate k/N_B, such that in each trellis transition the N_B coded bits are sent to the N_B fading blocks (this defines the blockwise partitioning of the CC codeword). Then, output blocks of length L correspond exactly to trellises of L_B trellis sections. Let

$$F_\epsilon^{L_B}(\boldsymbol{\omega}) \triangleq \frac{1}{L_B} \log \left| \mathcal{S}_\epsilon^{L_B}(\boldsymbol{\omega}) \right| \tag{C.45}$$

define the length-L_B MWEF exponential growth rate, such that $a(\boldsymbol{\omega})$ defined in (3.24) is given by $\lim_{\epsilon \to 0} \lim_{L_B \to \infty} F_\epsilon^{L_B}(\omega)$.

A *simple* error path in the trellis is a path that leaves state 0 and remerges for the first time to state 0 after a certain number of trellis sections, while it coincides with the all-zero path everywhere else. Take a fixed Hamming weight $\beta < \infty$ such that there exist a simple error path of length $\ell(\beta)$ (independent of L_B) in the code having Hamming weights per block $|w_b - \beta| \le \epsilon L_B/2$, for all $b = 1, \dots, N_B$. Such simple error path exists for any trellis terminated CC with given encoder. For example, we can take the minimum-length event corresponding to the free Hamming distance of the code. It follows that, for all $\epsilon > 0$, sufficiently large L_B and $\boldsymbol{\omega}$ such that $|\omega_b - \frac{\beta}{L_B}| \le \epsilon/2$, the lower bound

$$\left| \mathcal{S}_\epsilon^{L_B}(\boldsymbol{\omega}) \right| \ge L_B - \ell(\beta) \tag{C.46}$$

holds. In fact, $\mathcal{S}_\epsilon^{L_B}(\boldsymbol{\omega})$ must contain all the simple error events of length at most $\ell(\beta)$ starting at the $L_B - \ell(\beta) + 1$ distinct positions in the length-L_B trellis.

By continuity, we can write the sufficient condition of Theorem 2 as

$$\widehat{s} = \lim_{\epsilon \to 0} \lim_{L_B \to \infty} \sup_{\mathbf{x} \in \mathcal{W}(N_B, d_{SB}(R))} \sup_{\boldsymbol{\omega} \in [0,1]^{N_B}} \frac{F_\epsilon^{L_B}(\boldsymbol{\omega})}{\sum_{b=1}^{N_B} x_b \omega_b} \tag{C.47}$$

For any $\mathbf{x} \in \mathcal{W}(N_B, d_{SB}(R))$, take $\boldsymbol{\omega}$ in a box of side ϵ around the point $\frac{\beta}{L_B} \mathbf{1} \in [0,1]^{N_B}$. The lower bound (C.46) implies that

$$\widehat{s} \ge \lim_{\epsilon \to 0} \lim_{L_B \to \infty} \frac{\frac{1}{L_B} \log L_B + O(\frac{1}{L_B})}{\frac{\beta N_B}{L_B} + \epsilon} \tag{C.48}$$

which clearly shows that trellis terminated CCs cannot satisfy Theorem 2.

More in general, any code ensemble such that the MWEF $A_{w_1, \dots, w_{N_B}}$ increases *linearly* with the block length L_B for some Hamming weight vector $\mathbf{w}$ such that $\|\mathbf{w}\|_\infty \le \beta < \infty$ cannot satisfy Theorem 2. These code ensembles have an infinite ML decoding threshold [41] in the standard AWGN channel.

Proofs of the asymptotic union bounds (3.28) and (3.29). The M&L Bhattacharyya union bound for a code over the BPSK signal set can be written as

$$P_e(\rho) \le \mathbb{E}\left[\min\left\{ 1, \sum_{\boldsymbol{\omega} \in [0,1]^{N_B}} e^{-L_B \left(\rho \sum_{b=1}^{N_B} \gamma_b \omega_b \, - \, F(\boldsymbol{\omega}) \right)} \right\} \right]. \tag{C.49}$$

Since $\min(1, f(x))$ is continuous in x for continuous f and $\min(1, f(x)) \leq 1$, we can apply the dominated convergence theorem [103] and write,

$$\begin{aligned}
\lim_{L_B \to \infty} \quad & \mathbb{E}\left[\min\left\{1, \sum_{\boldsymbol{\omega} \in [0,1]^{N_B}} e^{-L_B\left(\rho \sum_{b=1}^{N_B} \gamma_b \omega_b \,-\, F(\boldsymbol{\omega})\right)}\right\}\right] \\
= \quad & \mathbb{E}\left[\min\left\{1, \lim_{L_B \to \infty} \sum_{\boldsymbol{\omega} \in [0,1]^{N_B}} e^{-L_B\left(\rho \sum_{b=1}^{N_B} \gamma_b \omega_b \,-\, F(\boldsymbol{\omega})\right)}\right\}\right] \\
= \quad & \mathbb{E}\left[\min\left\{1, \lim_{L_B \to \infty} \sum_{\boldsymbol{\omega} \in [0,1]^{N_B}} e^{-L_B\left(\rho \sum_{b=1}^{N_B} \gamma_b \omega_b \,-\, a(\boldsymbol{\omega})\right)}\right\}\right] \qquad \text{(C.50)}
\end{aligned}$$

The factor multiplying L_B in the exponent of the RHS of (C.50) is positive for a given channel realization $\boldsymbol{\gamma}$ if

$$\widehat{\rho} \triangleq \max_{\boldsymbol{\omega} \in [0,1]^{N_B}} \frac{a(\boldsymbol{\omega})}{\displaystyle\sum_{b=1}^{N_B} \gamma_b \omega_b} < \rho \qquad \text{(C.51)}$$

Conditioning with respect to $\boldsymbol{\gamma}$ we have that, in the limit of large L_B, $P_e(\rho|\boldsymbol{\gamma}) \to 0$ if $\rho > \widehat{\rho}$ while $P_e(\rho|\boldsymbol{\gamma}) \leq 1$ otherwise. It follows that in the limit for $L_B \to \infty$ the M&L Bhattacharyya bound takes on the form (3.28).

In the case of BICM codes with random bit interleaving (under the symmetrized mapping assumption), we can use the same M&L Bhattacharyya bound by replacing γ_b with ζ_b defined in (3.20), and (3.29) follows.

D

Tangential-Sphere Bound for Block-Fading Channels

In this appendix we illustrate how the tangential-sphere bound can be extended to block-fading channels. The tangential sphere bound, originally proposed by Poltyrev in [98], and further elaborated by Sason and Shamai for turbo-like code ensembles in [39]. Consider a binary code $\mathcal{C}$ of length $N = N_B L$ mapped over BPSK modulation over the block-fading channel,

$$\mathbf{y}_b = h_b \mathbf{x}_b + \mathbf{z}_b \quad b = 1, \ldots, N_B \tag{D.1}$$

where the entries of $\mathbf{z}_b$ are i.i.d. $\sim \mathcal{N}_{\mathbb{C}}(0, \frac{1}{\rho})$. Again, here we define $\gamma_b \overset{\Delta}{=} |h_b|^2$. Notice that this model represents exactly the same channel as (2.1), but now the noise variance is modified accordingly to have SNR ρ. This modification will turn out to be useful in the derivation of the tangential-sphere bound. We consider that the channel coefficients are perfectly known at the receiver. Moreover, let $A_{w_1,\ldots,w_{N_B}}$ denote the MWEF of $\mathcal{C}$, i.e., the number of pairwise error events in $\mathcal{C}$ with Hamming weight per block $w_1, , \ldots, w_{N_B}$.

We will now, derive the tangential-sphere bound for a given channel realization. The tangential-sphere bound is based on the following lemma,

Lemma 2 *Let $\mathcal{A}$ and $\mathcal{B}$ be two events in a given probability space. Then,*

$$\Pr(\mathcal{A}) \le \Pr(\mathcal{A} \cap \mathcal{B}) + \Pr(\bar{\mathcal{B}}). \tag{D.2}$$

In the tangential sphere bound, the event $\mathcal{A}$ represents the decoding error, and the event $\mathcal{B} = \mathsf{C}_{\theta,N}(\mathbf{x_h})$ represents an N-dimensional cone with vertex in the origin, half angle θ and with main axis $\mathbf{x_h}$ where $\mathbf{x_h} \triangleq (h_1\mathbf{x}_1, \dots, h_{N_B}\mathbf{x}_{N_B})$ denotes the faded codeword and $\mathcal{C}_\mathbf{h}$ the faded code. Notice that all codewords of $\mathcal{C}_\mathbf{h}$ lie in a multidimensional sphere of N dimensions, of squared radius

$$\mathsf{R}^2 = L \sum_{b=1}^{N_B} \gamma_b. \tag{D.3}$$

The only difference with the standard AWGN case is that now, some components of the faded code have different energies due to the different fading values in different blocks.

Following the derivation in [98, 39], we show in Figure D.1 the geometry of the tangential-sphere bound in the case of faded codewords. We now define the standard bound parameters (see Figure D.1 for their geometric interpretation). Let

$$\delta = 4 \sum_{b=1}^{N_B} \gamma_b w_b \tag{D.4}$$

be the squared Euclidean distance of a pairwise error event with Hamming weights per block $w_1, \dots, w_{N_B}$ [1],

$$\sigma^2 = \frac{1}{2\rho}, \tag{D.5}$$

$$r_{z_1} = \mathsf{r}\left(1 - \frac{z_1}{\mathsf{R}}\right), \tag{D.6}$$

$$\beta_\delta(z_1) = \frac{r_{z_1}}{\sqrt{1 - \frac{\delta^2}{4\mathsf{R}^2}}} \frac{\delta}{2\mathsf{r}}, \tag{D.7}$$

$$\alpha_\delta = \mathsf{r}\sqrt{1 - \frac{\delta^2}{4\mathsf{R}^2}} \tag{D.8}$$

and r, the radius of the cone with angle θ.

Then, the expression for the tangential-sphere bound for a given channel

---

[1] Notice that this implicitly defines the conditional Euclidean distance spectrum, A_δ accounting for the number of codewords with squared Euclidean distance δ

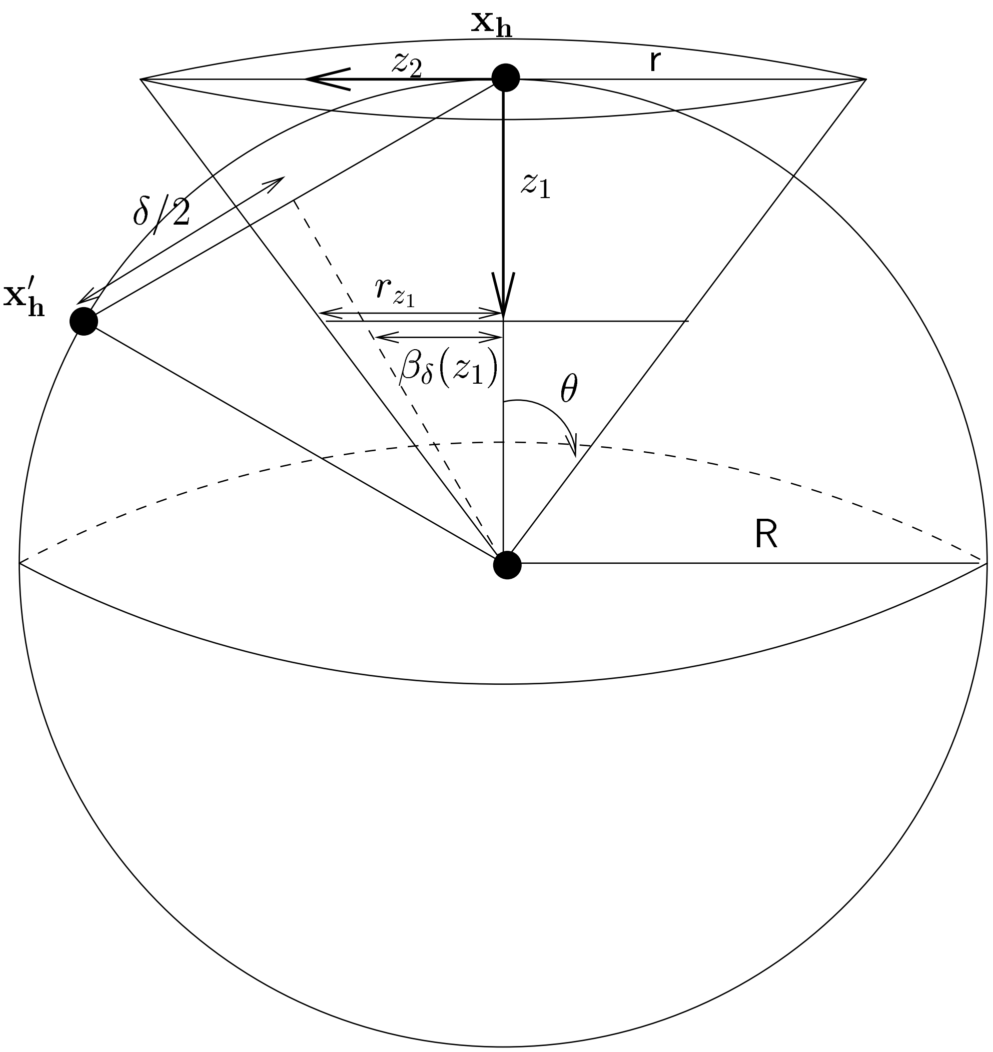

$\mathbf{x_h}$
z_2
r
z_1
$\delta/2$
$\mathbf{x}'_\mathbf{h}$
r_{z_1}
$\beta_\delta(z_1)$
θ
R

Figure D.1: Illustration of the geometry of the tangential sphere bound for the block-fading channel.

realization is given by,

$$\begin{aligned}P_e(\rho|\mathbf{h}) &\le P_e^{\mathrm{TSB}}(\rho|\mathbf{h})\\ &\triangleq \int_{-\infty}^{+\infty}\frac{dz_1}{\sqrt{2\pi\sigma^2}}e^{-z_1^2/2\sigma^2}\Bigg\{1-\bar{\Gamma}\left(\frac{N-1}{2},\frac{r_{z_1}}{2\sigma^2}\right)+\\ &+\sum_{\delta:\,\delta/2<\alpha_\delta}A_\delta\,\bar{\Gamma}\left(\frac{N-2}{2},\frac{r_{z_1}^2-\beta_\delta(z_1)^2}{2\sigma^2}\right)\cdot\\ &\cdot\left[\mathrm{Q}\left(\frac{\beta_\delta(z_1)}{\sigma}\right)-\mathrm{Q}\left(\frac{r_{z_1}}{\sigma}\right)\right]\Bigg\}\end{aligned} \tag{D.9}$$

where

$$\mathrm{Q}(x)\triangleq\frac{1}{\sqrt{2\pi}}\int_x^{\infty}e^{-\frac{t^2}{2}}dt \tag{D.10}$$

is the Gaussian tail function,

$$\bar{\Gamma}(a,x)\triangleq\frac{1}{\Gamma(a)}\int_0^x t^{a-1}e^{-t}dt \tag{D.11}$$

is the normalized incomplete gamma function,

$$\Gamma(x)\triangleq\int_0^{+\infty}t^{x-1}e^{-t}dt \tag{D.12}$$

is the gamma function and r, the cone radius, is the solution of

$$\sum_{\delta:\,\delta/2<\alpha_\delta}A_\delta\int_0^{\theta_k}\sin^{N-3}\phi d\phi=\frac{\sqrt{\pi}\Gamma\left(\frac{N-2}{2}\right)}{\Gamma\left(\frac{N-1}{2}\right)} \tag{D.13}$$

with

$$\theta_k=\cos^{-1}\left(\frac{\delta}{2\mathsf{r}}\frac{1}{\sqrt{1-\frac{\delta^2}{4\mathsf{R}^2}}}\right). \tag{D.14}$$

Then, the expectation which is taken over the joint distribution of $\boldsymbol{\gamma}=(\gamma_1,\dots,\gamma_{N_B})$ gives,

$$P_e(\rho)\le\mathbb{E}[P_e^{\mathrm{TSB}}(\rho|\mathbf{h})]. \tag{D.15}$$

Notice that as $P_e^{\mathrm{TSB}}(\rho|\mathbf{h})$ is a probability, there is no need to use the trick if $\mathbb{E}[\min\{1,P_e(\rho|\mathbf{h})\}]$ [25]. Note the high computational effort required to compute (D.15): we have to use Montecarlo techniques to compute the average, and at every channel realization one has to optimize the cone radius with (D.13) and compute (D.9).

However, we can compute an approximated tangential-sphere bound using the technique described in [51]. Define the modified squared Euclidean distance as

$$\delta' = 4\Delta_p^{1/N_B} \sum_{b=1}^{N_B} \gamma_b \tag{D.16}$$

where $\Delta_p \triangleq \prod_{b=1}^{N_B} w_b$ denotes the product weight corresponding to a pairwise error event with weights per block $w_1, \dots, w_{N_B}$. Then, if we replace d by d' in (D.9) and (D.13) and we redefine accordingly all the bound parameters, we can easily compute the expectation over the fading states since

$$\gamma = \sum_{b=1}^{N_B} \gamma_b \tag{D.17}$$

is central chi-squared with $2N_B$ degrees of freedom and mean N_B, such that

$$f_\gamma(z) = \frac{z^{N_B - 1}}{(N_B - 1)!} e^{-z} \tag{D.18}$$

is the pdf.

Then, we can compute a new approximated upper bound as,

$$\begin{aligned} P_e(\rho) \lessapprox \int_0^{+\infty} \Bigg[\int_{-\infty}^{+\infty} \frac{dz_1}{\sqrt{2\pi\sigma^2}} e^{-z_1^2/2\sigma^2} \Bigg\{ 1 - \bar{\Gamma}\left(\frac{N-1}{2}, \frac{r_{z_1}}{2\sigma^2}\right) + \\ + \sum_{\substack{\Delta_p \\ \delta' : \delta'/2 < \alpha_{\delta'}}} A_{\Delta_p} \bar{\Gamma}\left(\frac{N-2}{2}, \frac{r_{z_1}^2 - \beta_{\delta'}(z_1)^2}{2\sigma^2}\right) \cdot \\ \cdot \left[\mathrm{Q}\left(\frac{\beta_\delta(z_1)}{\sigma}\right) - \mathrm{Q}\left(\frac{r_{z_1}}{\sigma}\right) \right] \Bigg\} \Bigg] f_\gamma(\gamma) d\gamma \end{aligned} \tag{D.19}$$

with r' being now the solution of

$$\sum_{\substack{\Delta_p \\ \delta' : \delta'/2 < \alpha_{\delta'}}} A_{\Delta_p} \int_0^{\theta'_k} \sin^{N-3} \phi \, d\phi = \frac{\sqrt{\pi}\,\Gamma\left(\frac{L-2}{2}\right)}{\Gamma\left(\frac{N-1}{2}\right)} \tag{D.20}$$

with

$$\theta'_k = \cos^{-1}\left(\frac{\delta'}{2\mathsf{r}'} \frac{1}{\sqrt{1 - \frac{\delta'^2}{4\mathsf{R}^2}}} \right). \tag{D.21}$$

that has only a single variate expectation and only needs the product enumerator, which significantly reduces the computational complexity with respect to (D.9).

E

Approximated Density Evolution for MIMO IC receivers

In this appendix we describe a semi-analytical technique based on density evolution techniques first introduced to analyze the performance and capacity of low-density parity check codes (LDPC) [37] and applied in [67] to analyzed the performance of large system multiuser receivers in CDMA systems. The principle underlying DE techniques is that, for large codeword length, the pdf of the messages propagated by the iterative decoder at a given iteration, concentrates around the expected pdf resulting from a cycle-free graph. In this paper we provide a DE algorithm based on the Gaussian approximation and improved bounding techniques [98, 39] (see also Appendix D), in order to analyze the iterative IC decoding approaches described in Chapter 4. This method captures the diversity behavior of the code in the limit of infinite block length. We anticipate that the proposed method, although conceptually very simple, carries a computational burden comparable to simulation mainly due to the outer expectation over the fading states inherent in the non-ergodic nature of the channel (4.1). However, we think that using this basic idea, we can envisage more efficient algorithms that use fast density evolution techniques [91] in order to characterize in what system settings (antenna configurations, binary code, signal constellation) IC techniques can be efficiently employed. We describe the method for BICM NSTC with BPSK modulation, while it can be easily extended to the case of BICM TAST using the equivalent channel $\mathcal{H}$. The method can be extended using the results of Appendix A to general signal constellations.

For every realization of $\mathbf{H}$, the signal-to-noise ratios at the output of the IC filter $\boldsymbol{\beta}^{(i)} = (\beta_1^{(i)}, \ldots, \beta_{N_T}^{(i)})$ at iteration i of the equivalent parallel channels are given by,

$$\beta_t^{(i)} = \frac{\rho}{|\mathbf{f}_t^{(i)}|^2 + \rho \sum_{t' \neq t} |\mathbf{f}_t^{(i)} \mathbf{h}_{t'}|^2 v_t^{(i)}}. \tag{E.1}$$

where we recall that $\mathbf{f}_t^{(i)}$ is the IC filter corresponding to antenna t at iteration i and $v_t^{(i)}$ is the variance of the residual interference. The SNRs between the different antennas may be very different from each other depending on the channel realization. Recall that, as opposed to the CDMA case, large system arguments cannot be invoked, and therefore, we must track the evolution of a vector of parameters.

We characterize the error probability given by the soft-input soft-output decoder of the binary code $\mathcal{C}$ block as $\epsilon = f(\boldsymbol{\beta}^{(i)})$. In order to avoid exhaustive simulation, we resort to improved bounding techniques, i.e., the tangential-sphere bound, modified to handle a block-fading channel (see Appendix D). Recall that $A_{w_1,\ldots,w_{N_T}}$ denotes the MWEF of $\mathcal{C}$, and that, given $\boldsymbol{\beta}^{(i)}$, the squared Euclidean distance of a pairwise error event with weights per block $w_1, \ldots, w_{N_T}$ is

$$\delta = 4\rho \sum_{t=1}^{N_T} w_t \beta_t^{(i)}. \tag{E.2}$$

This defines the conditional Euclidean distance spectrum A_d. In order to model the error rate fed back from the decoder at each antenna and iteration, we compute the symbol error rate (SER) per block, we define $A^t_{w_1,\ldots,w_{N_T}} = \frac{w_t}{L} A_{w_1,\ldots,w_{N_T}}$. Then, the tangential-sphere bound on the SER of antenna t at the i-th iteration can be written using $A^t_{w_1,\ldots,w_{N_T}}$ in (D.9).

In order to avoid the computation of the log-likelihood ratios density function at the decoder's output, we use the so-called Gaussian Approximation (GA), and we consider that the a posteriori $\mathcal{L}^t_{\text{app}} \sim \mathcal{N}(\mu^t_{\text{app}}, 2\mu^t_{\text{app}})$. Then,

$$\epsilon_s = \Pr(\mathcal{L}_{\text{app}} < 0) = \mathrm{Q}\left(\sqrt{\frac{\mu_{\text{app}}}{2}}\right), \tag{E.3}$$

which gives that

$$\mu^t_{\text{app}} = 2[\mathrm{Q}^{-1}(\epsilon^t_s)]^2. \tag{E.4}$$

We model the extrinsic

$$\mathcal{L}^t_{\text{ext}} = \mathcal{L}^t_{\text{app}} - \mathcal{L}^t_{\text{in}} \tag{E.5}$$

as $\sim \mathcal{N}(\mu^t_{\text{ext}}, 2\mu^t_{\text{ext}})$, with $\mu^t_{\text{ext}} = \mu^t_{\text{app}} - \mu^t_{\text{in}}$, where $\mathcal{L}^t_{\text{in}} \sim \mathcal{N}(\mu^t_{\text{in}}, 2\mu^t_{\text{in}})$, is the LLR at the decoder input, with $\mu^t_{\text{in}} = 4\beta_t$. Therefore, the residual interference variances for the next iteration are

$$v_t^{(i+1)} = 1 - \mathbb{E}\left[\tanh^2\left(\frac{\mathcal{L}_{\text{ext}}^{t,(i)}}{2}\right)\right] \tag{E.6}$$

which can be easily computed using the Gauss-Hermite quadrature rules [45]. The bit error rate (BER) $\epsilon_b(i)$ at each iteration can be bounded using the same technique by replacing A_δ^t by $A'_\delta = \sum_h \frac{h}{K} A_{h,\delta}$ in (D.9) and (D.13), where $A_{h,\delta}$ is the number of pairwise error events with input Hamming weight h and Euclidean distance δ. We will nickname this density evolution method as GA-TSB-DE.

We here report some numerical examples of the proposed density evolution algorithm for a $N_T = N_R = 4$ MIMO channel with the MMSE-IC receiver. Figure E.1 shows the evolution with the iterations of the BER, for a given channel matrix $\mathbf{H}$ for $E_b/N_0 = 0$dB, for the convolutional code $(5,7,7,7)_8$. It can be observed that, for this given realization of $\mathbf{H}$ and SNR, there is a very good correspondence between the GA-TSB-DE method and the simulation for large codeword length (K=100000) and not-so-large codeword length (K=128). Moreover we also see that the MFB is achieved after 3 iterations. For the sake of completeness we also report the BER of the ML decoder of NSTC. Notice that, for lower SNR or a *bad* $\mathbf{H}$, the matching between simulation and GA-TSB-DE may not be so good, making the code work at the region where the TSB is not tight. Also, for some bad channel matrices, the MFB may not be achieved at all.

In Figure E.2 we show the evolution chart of the residual interference variance vector. As observed, there is a fixed point at a very low value, which indicates that the MFB will be closely approached. Notice that, due to the GA and TSB on SER per block and BER, GA-TSB-DE does not yield any upper nor lower bound on the BER performance, but only an approximation. We have computed the truncated MWEF with a modified version of the algorithm in [104], in order to handle multivariate simple error events. Notice that computing only simple error events is needed for convolutional codes, and that $A_{w_1,\ldots,w_{N_T}} \leq K A^{\text{simple}}_{w_1,\ldots,w_{N_T}}$ counts for its shifts, where $A^{\text{simple}}_{w_1,\ldots,w_{N_T}}$ is the MWEF for simple error events of the convolutional code.

We finally show in Figure E.3 the Montecarlo average BER and WER performance predicted by the GA-TSB-DE compared to simulation. Notice that, strictly speaking, DE methods yield WER = 1. Therefore, in order to estimate the WER, which is the relevant performance measure for the case of our study, we have computed the TSB for the WER by using the truncated MWEF $A_{w_1,\ldots,w_{N_T}} \leq K A^{\text{simple}}_{w_1,\ldots,w_{N_T}}$ for K = 128. As we observe from the figure, there is a very good matching between simulation and the prediction by GA-TSB-DE. Indeed, GA-TSB-DE predicts very well the diversity gain of the iterative MMSE-IC decoder and provides good performance estimation. As mentioned in a previous section, since BICM NSTC are not full rank, its performance will cross that pre-

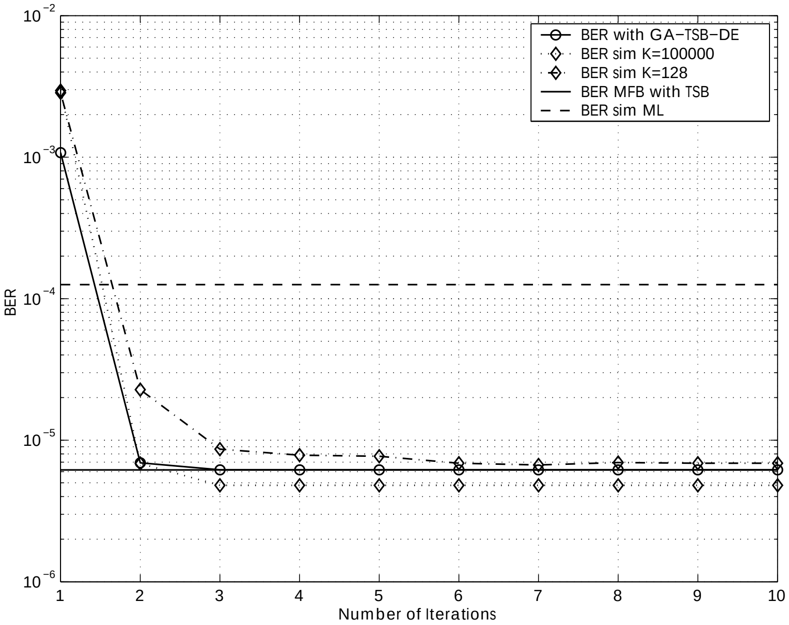


Figure E.1: BER evolution with the iterations for a fixed $\mathbf{H}$ for $E_b/N_0 = 0$dB and $(5,7,7,7)_8$ CC.

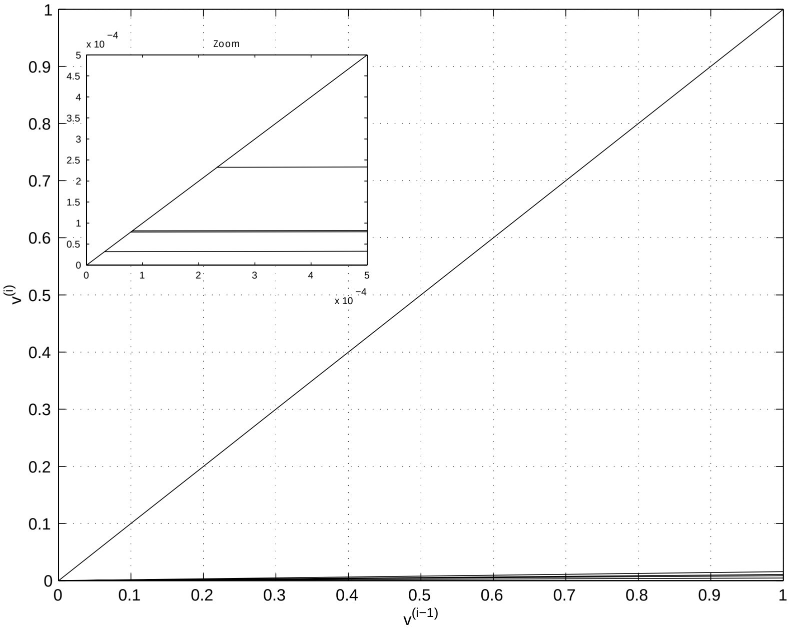


Figure E.2: Snapshot for fixed $\mathbf{H}$ (the same of Figure E.1) of the residual interference variance evolution for $E_b/N_0 = 0$dB and $(5,7,7,7)_8$ code.

dicted by GA-TSB-DE (infinite length BICM NSTC are full rank by definition), and will show a diversity floor effect, similar to that of turbo-codes.

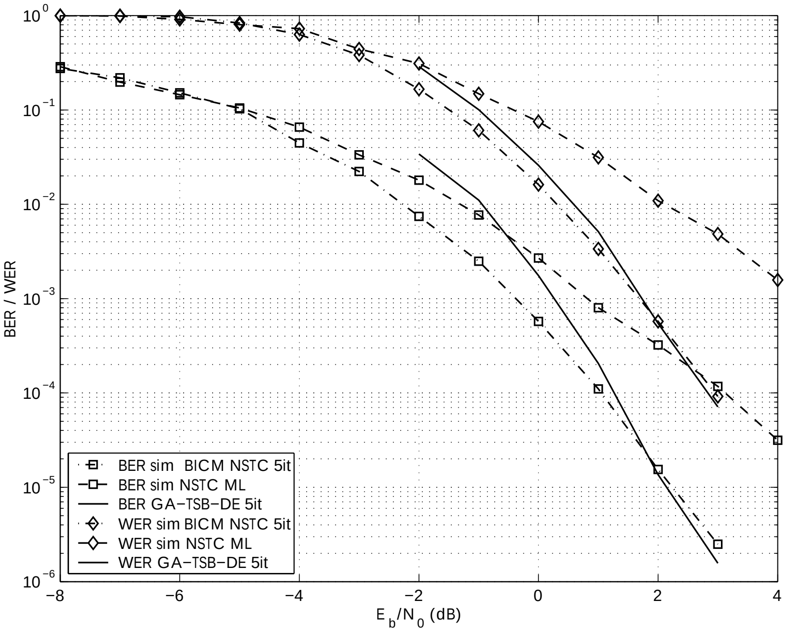


Figure E.3: Average BER and WER of BICM NSTC and NSTC $(5, 7, 7, 7)_8$ code.

BIBLIOGRAPHY

[1] S. W. Weinstein and P. M. Ebert, "Data transmission by frequency-division multiplexing using the discrete Fourier transform," *IEEE Trans. on Comm.*, vol. 19, no. 5, Oct. 1971.

[2] ETSI Normalization Comitee, "Radio Broadcasting Systems, Digital Audio Broadcasting (DAB) to mobile, portable and fixed receivers," *Norm ETSI doc. ETS300401, European Telecommunications Standards Institute, Sophia-Antipolis, France,*, 1005-1997.

[3] ETSI Normalization Comitee, "Digital Broadcasting Systems for Television Sound and Data Services; Framing structure, Channel Coding and Modulation for Digital Terrestrial Television," *Norm ETSI doc. ETS300744, European Telecommunications Standards Institute, Sophia-Antipolis, France,*, 1996.

[4] ANSI T1E1.4 Comitee Contribution, "The DWMT: a multicarrier transceiver for adsl using m-band wavelets," Tech. Rep., ANSI, 1993.

[5] ETSI Normalization Comitee, "Broadband Radio Access Networks (BRAN); HIPERLAN Type 2; Physical (PHY) Layer," *Norm ETSI doc. TS101475, European Telecommunications Standards Institute, Sophia-Antipolis, France,*, 2000.

[6] L. H. Ozarow, S. Shamai and A. D. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. on Vehicular Tech.*, vol. 43, no. 2, pp. 359–378, May 1994.

[7] E. Biglieri, J. Proakis and S. Shamai, "Fading channels: information-theoretic and communications aspects," *IEEE Trans. on Inform. Theory*, vol. 44, no. 6, pp. 2619 –2692, Oct. 1998.

[8] A. Guillén i Fàbregas and G. Caire, "Turbo-like codes over the block-fading channel: code design and construction," in *3rd International Symposium on Turbo Codes and Related Topics, Brest, France*, Sept. 2003.

[9] A. Guillén i Fàbregas and G. Caire, "Turbo-like codes over the block-fading channel," in *2004 International Symposium on Information Theory, Chicago, IL*, June-July 2004.

[10] A. Guillén i Fàbregas and G. Caire, "Turbo-like codes are good in the block-fading channel," in *2004 International Symposium on Information Theory and Applications, Parma Italy*, October 2004.

[11] A. Guillén i Fàbregas and G. Caire, "Coded modulation in the block-fading channel: Coding theorems and code construction," *submitted to IEEE Trans. on Information Theory*, 2004.

[12] A. Guillén i Fàbregas and G. Caire, "Impact of signal constellation expansion on the achievable diversity in quasistatic multiple-antenna channels," in *(invited paper) to appear in Proc. 2004 European signal Processing Conference (EUSIPCO), Vienna, Austria*, September 2004.

[13] A. Guillén i Fàbregas and G. Caire, "Analysis and design of natural and threaded space-time codes with iterative decoding," *36th Asilomar Conference*, 2002.

[14] A. Guillén i Fàbregas and G. Caire, "Design of space-time bicm codes for block fading channels with iterative decoding," *37th Conference on Sciences and Systems CISS, The Johns Hopkins University*, March 2003.

[15] G. Caire, M. Guillaud, A. Guillén i Fàbregas, K. Gosse, A. Ribeiro Dias, S. Rouquette and D. T. M. Slock, "Investigation of multiple transmit multiple receive antenna techniques without channel knowledge at the emitter, in the hiperlan/2 context," Tech. Rep., RNRT ANTIPODE project, deliverable 2.1, October 2003.

[16] G. Caire, M. Guillaud, A. Guillén i Fàbregas, K. Gosse, A. Ribeiro Dias, S. Rouquette and D. T. M. Slock, "Analysys of the impact of a realistic propagation environment at 5ghz on mtmr techniques: Simulation results," Tech. Rep., RNRT ANTIPODE project, deliverable 2.6, October 2003.

[17] A. Guillén i Fàbregas, A. Martinez and G. Caire, "Error probability of Bit-interleaved Coded Modulation using the Gaussian Approximation," *in Proc. 38th Annual Conference on Information Sciences and Systems, CISS 2004, Princeton, NJ, March 2004.*

[18] A. Martinez, A. Guillén i Fàbregas and G. Caire, "New simple evaluation of the the error probability of bit-interleaved coded modulation using the saddlepoint approximation," *to appear 2004 International Symposium on Information Theory and its Applications (ISITA), Parma, Italy*, 2004.

[19] R. Knopp, *Coding and Multiple-Access over Fading Channels*, Ph.D. thesis, Ecole Polytechnique Federale de Lausanne (EPFL) and Institut Eurecom, 1997.

[20] G. Caire, G. Taricco and E. Biglieri, "Optimum power control over fading channels," *IEEE Trans. on Inform. Theory*, vol. 45, no. 5, pp. 1468–1489, July 1999.

[21] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. on Inform. Theory*, vol. 40, no. 4, pp. 1147–1157, Jul. 1994.

[22] L. Zheng and D. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels," *to appear IEEE Trans. on Inform. Theory*, 2002.

[23] J. Proakis, *Digital Communications*, McGraw-Hill, 1995.

[24] R. Knopp and P. Humblet, "On coding for block fading channels," *IEEE Trans. on Inform. Theory*, vol. 46, no. 1, pp. 1643–1646, July 1999.

[25] E. Malkamaki and H. Leib, "Evaluating the performance of convolutional codes over block fading channels," *IEEE Trans. on Inform. Theory*, vol. 45, no. 5, pp. 1643–1646, Jul. 1999.

[26] M. Chiani, A. Conti and V. Tralli, "Further results on convolutional code search for block fading channels," *to appear IEEE Trans. on Inform. Theory*, 2004.

[27] E. Malkamaki and H. Leib, "Coded diversity on block-fading channels," *IEEE Trans. on Inform. Theory*, vol. 45, no. 2, pp. 771–781, March 1999.

[28] I. E. Telatar, "Capacity of multi-antenna gaussian channels," *European Trans. on Telecomm.*, vol. 10, no. 6, pp. 585–596, November 1999.

[29] G. J. Foschini and M. J. Gans, "On limits of wireless communication in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, pp. 311–335, March 1998.

[30] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. on Inform. Theory*, pp. 774–765, March 1998.

[31] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.

[32] M. O. Damen, H. El Gamal and N. C. Beaulieu, "Linear threaded algebraic space-time constellations," *IEEE Trans. on Inform. Theory*, vol. 49, no. 10, pp. 2372–2388, Oct. 2003.

[33] J.-C. Guey, M. P. Fitz, M. R. Bell and W.-Y. Kuo, "Signal design for transmitter diversity wireless communication systems over rayleigh fading channels," in *Proc. Vehicular Technology Conference, Atlanta, GA,*, April 1996.

[34] H. El Gamal and A. R. Hammons, "On the design of algebraic space-time codes for MIMO block fading channels," *IEEE Trans. on Inform. Theory*, vol. 49, no. 1, pp. 151–163, Jan. 2003.

[35] R. Hammons and H. El Gamal, "On the theory of space-time codes for PSK modulation," *IEEE Trans on Inform. Theory*, vol. 46, no. 2, pp. 524–542, Mar. 2000.

[36] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. on Inform. Theory*, vol. 44, no. 3, pp. 927–946, May 1998.

[37] T. J. Richardson and R. L. Urbanke, "The capacity of low-density parity-check codes under message-passing decoding," *IEEE Trans. on Inform. Theory*, vol. 47, no. 2, pp. 599–618, Feb. 2001.

[38] S. ten Brink, "Designing iterative decoding schemes with the extrinsic information transfer chart," *AEU Int. J. Electron. Commun.*, vol. 54, no. 6, pp. 389–398, Dec. 2000.

[39] I. Sason and S. Shamai, "Improved upper bounds on the ml decoding error probability of parallel and serial concatenated turbo codes via their ensemble distance spectrum," *IEEE Trans. on Inform. Theory*, vol. 46, no. 1, pp. 24–47, Jan. 2000.

[40] I. Sason and S. Shamai, "On improved bounds on the decoding error probability of block codes over interleaved fading channels, with applications to turbo-like codes," *IEEE Trans. on Inform. Theory*, vol. 47, no. 6, pp. 2275–2299, Sept. 2001.

[41] D. Divsalar, "A simple tight bound on error probability of block codes with application to turbo codes," *JPL TMO Progress Report 42-139*, November 1999.

[42] D. Divsalar and E. Biglieri, "Upper bounds to error probabilities of coded systems beyond the cutoff rate," *IEEE Trans. on Commun.*, vol. 51, no. 12, pp. 2011–2018, December 2003.

[43] S. Shamai I. Sason and D. Divsalar, "Tight exponential upper bounds on the ml decoding error probability of block codes over fully interleaved fading channels," *IEEE Trans. on Commun.*, vol. 51, no. 8, pp. 1296–1305, August 2003.

[44] F. R. Kschischang, B. J. Frey, and H-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. on Inform. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.

[45] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*, New York: Dover Press, 1972.

[46] W.H.Press, S.A.Teukolsky, W.T.Vetterling and B.P. Flannery, *Numerical Recepies in C*, Cambridge University Press, also available at `http://www.nr.com`, 1992.

[47] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley Series in Telecommunications, 1991.

[48] D. Divsalar, H. Jin and R. J. McEliece, "Coding theorems for turbo-like codes," in *Proc. 1998 Allerton Conference*, October 1998.

[49] L. Bahl, J. Cocke, F. Jelinek and J. Raviv, "Optimal decoding of linear codes for minimizing the symbol error rate," *IEEE Trans. on Inform. Theory*, vol. 20, pp. 284–287, Mar. 1974.

[50] S. Wilson Y. Leung and J. Ketchum, "Multifrequency trellis coding with low delay for fading channels," *IEEE Trans. on Commun.*, vol. 41, no. 10, pp. 1450–1458, October 1993.

[51] M-K. Byun, D. Park and B.G Lee, "Performance analysis of space-time codes in quasistatic fading channels," *to appear IEEE Trans. on Inform. Theory.*

[52] A. Viterbi, *CDMA : Principles of Spread Spectrum Communication*, Prentice Hall, 1995.

[53] I. Sason, E. Telatar and R. Urbanke, "On the asymptotic input-output weight distributions and thresholds of convolutional and turbo-like encoders," *IEEE Trans. on Inform. Theory*, vol. 48, no. 12, pp. 3052–3061, December 2002.

[54] S. Benedetto and G. Montorsi, "Unveiling turbo codes: Some results on parallel concatenated coding schemes," *IEEE Trans. on Inform. Theory*, vol. 42, no. 2, pp. 409–428, Mar. 1996.

[55] S. Benedetto, D. Divsalar, G. Montorsi and F. Pollara, "Serial concatenated of interleaved codes: Performance analysis, design and iterative decoding," *IEEE Trans. on Inform. Theory*, vol. 44, no. 3, pp. 909–926, May 1998.

[56] H. Jin, *Analysis and Design of Turbo-like Codes*, Ph.D. thesis, California Institute of Technology, 2001.

[57] B. Hassibi and B. Hochwald, "High-rate codes that are linear in space and time," *IEEE Transactions on Information Theory*, July 2002.

[58] G. Ungerboeck, "Channel coding with multilevel signals," *IEEE Trans. on Inform. Theory*, vol. 28, Jan. 1988.

[59] A. M. Tonello, "Space-time bit-interleaved coded modulation with an iterative decoding strategy," in *Proc. IEEE VTC 2000 Fall*, Sept 2000.

[60] S. H. Müller-Weinfurtner, "Coding for multiple antenna transmission in fast fading and in OFDM," in *Proc. IEEE ICC 2002*, May 2002.

[61] G. Taricco E. Biglieri and E. Viterbo, "Bit-interleaved time-space codes for fading channels," in *Proc. CISS 2000, Princeton, NJ*, March 2000.

[62] A. Conti M. Chiani and V. Tralli, "Bit-interleaved pragmatic space-time codes: Design and construction," in *Proc. IEEE Globecom 2002*, Nov 2002.

[63] J. Boutros, N. Gresset and L. Brunel, "Turbo coding and decoding for multiple antenna channels," *in the Proc 3th Int. Symp. Turbo Codes and Applications, Brest, France*, Sept. 2003.

[64] G. Caire and G. Colavolpe, "On low-complexity space-time coding for quasi-static fading channels," *IEEE Trans. on Inform. Theory*, vol. 49, no. 6, pp. 1400–1416, June 2003.

[65] M. O. Damen and H. El Gamal, "Universal space-time codes," *IEEE Trans. on Inform. Theory*, 2003.

[66] H. El Gamal and R. Hammons, "A new approach to layered space-time coding and signal processing," *IEEE Trans. on Inform. Theory*, vol. 47, no. 6, pp. 2321–2334, September 2001.

[67] J. Boutros and G. Caire, "Iterative joint decoding: Unified framework and asymptotic analysis," *IEEE Trans. on Inform. Theory*, vol. 48, no. 7, pp. 1772–1793, July 2002.

[68] C. Berrou, A. Glavieux and P. Thitimajshima, "Near shannon limit error correcting coding and decoding: Turbo codes," in *Proc. IEEE Int. Conference on Communications, Geneva, Switzerland*, 1993.

[69] M.Pohst, "On the computation of lattice vectors of minimal length, successive minima and reduced basis with applications," *ACM SIGSAM*, vol. 15, pp. 37–44, 1981.

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

[71] C.P.Schnorr and M. Euchner, "Lattice basis reduction: improved practical algorithms and solving subset sum problems," *Mathematical Programming*, vol. 66, pp. 181–191, 1994.

[72] E. Viterbo and J. Boutros, "A universal lattice decoder for fading channels," *IEEE Trans. on Inform. Theory*, vol. 45, no. 4, pp. 1639–1642, July 1999.

[73] J. Boutros, N. Gresset, L. Brunel and M. Fossorier, "Soft-input soft-output lattice sphere decoder for linear channels," *submitted to IEEE Global Communications Conference, March 2003.*

[74] H. Vikalo and B. Hassibi, "Towards closing the capacity gap on multiple antenna channels," in *Proc. IEEE ICASSP 2002*, May 2002.

[75] B. M. Hochwald and S. ten Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Trans. on Comm.*, vol. 51, no. 3, March 2003.

[76] M. O. Damen, H. El Gamal and G. Caire, "On maximum likelihood detection and the search for the closest lattice point," *IEEE Trans. on Inform. Theory*, vol. 49, no. 10, October 2003.

[77] A.Wiesel, X.Mestre, A.Pages and J.R. Fonollosa, "Efficient implementation of sphere demodulation," in *SPAWC 2003, Roma, Italy, June 15-18*, 2003.

[78] D. Wubben, R. Bohnke, J. Rinas, V. Kuhn and K.D. Kammeyer, "Efficient algorithm for decoding layered space-time codes," *IEEE Electronics Letters*, vol. 37, pp. 1348–1350, Oct. 2001.

[79] G. J. Foschini, G. D. Golden, R. A. Valenzuela, and P. W. Wolniansky, "Simplified processing for high spectral efficiency wireless communication employing multi-element arrays," *IEEE J. Select. Areas Commun.*, vol. 17, no. 11, pp. 1841–1852, Nov. 1999.

[80] G. Caire and R. Müller, "The optimal received power distribution for IC-based iterative multiuser joint decoders," in *39th Annual Allerton Conference on Commun., Control and Computing, Monticello, IL*, Oct. 3-5 2001.

[81] A. Nordio, M. Hernandez and G. Caire, "Low-complexity turbo equalization and multiuser decoding for td-cdma," *IEEE Trans. on Wireless Comm.*, vol. 3, no. 2, March 2004.

[82] L. K. Rasmussen, A. J. Grant and P. D. Alexander, "An Extrinsic Kalman Filter for Iterative Multiuser Decoding," *IEEE Trans. on Inform. Theory*, vol. 50, no. 4, April 2004.

[83] E. Biglieri, A. Nordio and G. Taricco, "Suboptimum receiver interfaces and space-time codes," *IEEE Trans. on Signal Processing*, vol. 51, no. 11, November 2003.

[84] R. Van Nee and R. Prasad, *OFDM for wireless communications*, Artech House Publishers, Boston, 2000.

[85] ETSI Normalization Comitee, "Channel models for HIPERLAN/2 in different indoor scenarios," Norm ETSI doc. 3ERI085B, European Telecommunications Standards Institute, Sophia-Antipolis, France, 1998.

[86] J.M Conrat and P. Laspougeas, "Modélisation de la propagation radio bande étroite 5 Ghz à l'intérieur des bâtiments," Tech. Rep., RNRT ANTIPODE project, France Telecom R&D, April 2002.

[87] H. Jafarkhani V. Tarokh and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. on Inform. Theory*, vol. 45, no. 5, pp. 1456–1467, 1999.

[88] A. Boariu O. Tirkkonen and A. Hottinen, "Minimal non-orthogonality rate 1 space-time block code for 3+ tx," in *Proc. International Symposium on Spread Spectrum Techniques and Applications*, Sept. 200.

[89] M. Guillaud and D. Slock, "Multi-stream coding for MIMO OFDM systems with space-time-frequency spreading," in *Proc. IEEE Int. Symp. on Wireless Personal Multimedia Communications, Honolulu, Hawaii*, 2002.

[90] 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 *Proc. ISSSE, Pisa, Italy*, Sept. 1998.

[91] H. Jin and T. Richardson, "Fast density evolution," *in Proc. 38th Annual Conference on Information Sciences and Systems, CISS 2004, Princeton, NJ, March 2004.*, Sept. 2004.

[92] M. Chiani,A. Conti and A.Ventura, "Evaluation of low-density parity-check codes over block fading channels," in *2000 IEEE International Conference on Communications*, June 2000.

[93] H. El Gamal, G. Caire and M. O. Damen, "Lattice coding and decoding achieve the optimal diversity-vs-multiplexing tradeoff of mimo channels," *submitted to IEEE Trans. on Inform. Theory*, November 2003.

[94] E. Zehavi, "8-PSK trellis codes for a rayleigh channel," *IEEE Trans. on Commun.*, vol. 40, no. 5, pp. 873–884, May 1992.

[95] X. Li and J. A. Ritcey, "Trellis-coded modulation with bit interleaving and iterative decoding," *IEEE JSAC*, pp. 715–724, April 1999.

[96] F. Brannstrom, L. K. Rasmussen and A. Grant, "Optimal scheduling for multiple serially concatenated systems," *in the Proc 3th Int. Symp. Turbo Codes and Applications, Brest, France*, Sept. 2003.

[97] A. J. Viterbi and Jim K. Omura, *Principles of Digital Communication and Coding*, McGraw-Hill, 1979.

[98] G. Poltyrev, "Bounds on decoding error probability of binary linear codes via their spectra," *IEEE Trans. on Inform. Theory*, vol. 40, pp. 1284–1292, July 1994.

[99] J. Zangl and R. Herzog, "Improved tangential-sphere bound on the bit-error probability of concatenated codes," *IEEE J. Select. Areas Commun.*, vol. 19, no. 5, pp. 825–830, May 2001.

[100] H. M. Tullberg and P. H. Siegel, "Serial concatenated trellis coded modulation with inner rate-1 accumulate code," *in the Proc Globecom 2001, San Antonio, Texas*, Nov. 2001.

[101] J.A. Anderson and J.M. Bell, *Number Theory with Applications*, Prentice Hall, 1997.

[102] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, Number 38 in Applications of Mathematics. Springer Verlag, 2nd edition, April 1998.

[103] R. Durrett, *Probability: Theory and Examples*, Duxbury Press, 1996.

[104] N. Phamdo X. Huang and L. Ping, "Recursive method for generating weight ennumerating functions for trellis codes," *IEE Electronic Letters*, vol. 37, no. 12, June 2001.