Optimisation de précodeurs linéaires pour les systèmes MIMO à récepteurs itératifs

Nhat-Quang Nhan

THÈSE / UNIVERSITÉ DE BRETAGNE OCCIDENTALE
sous le sceau *de l'Université* Bretagne Loire

pour obtenir le titre de
DOCTEUR DE L'UNIVERSITÉ DE BRETAGNE OCCIDENTALE
Mention : *Sciences et Technologies de l'Information et de la* Communication
Spécialité: Communications numériques
École Doctorale SICMA

présentée par
Nhat-Quang NHAN
Préparée à l'unité de recherche LAB-STICC, CNRS UMR 6285
Université de Bretagne Occidentale et TELECOM Bretagne

Optimization of linear precoders for coded MIMO systems with iterative receivers

Thèse soutenue le 05 octobre 2016
devant le jury composé de :

Benoît GELLER
Professeur, ENSTA ParisTech / rapporteur

Charly POULLIAT
Professeur, INP-ENSEEIHT Toulouse/ rapporteur

Olivier BERDER
Professeur, Université de Rennes 1 / examinateur

Telex M. N. NGATCHED
A. Professor, Memorial University of Newfoundland / examinateur

Karine AMIS
Maître de Conférences HDR, TELECOM Bretagne / co-directrice de thèse

Emanuel RADOI
Professeur, Université de Bretagne Occidentale / directeur de thèse

Philippe ROSTAING
Maître de Conférences, Université de Bretagne Occidentale / encadrant

"Opportunity is missed by most people because it is dressed in overalls and looks like work."

— T. A. Edison

Acknowledgements

I would like to express my sincere gratitude to my first supervisor, Prof. Emanuel Radoi, for encouraging my research and for allowing me to grow as a research scientist under his supervision. His constant support on my professional and personal life as well as his advice on my future career have been invaluable. His technical advice and encouragement gave me full of motivation, for every single day in the previous three years, to pursuit my research works.

Similar, profound gratitude goes to my second supervisor, Dr. Karine Amis. I am especially grateful to her for the encouragement and continuous support during my PhD study. At the beginning of my PhD, she spent a lot of time to teach and help me to build up my scientific background for this thesis. I have learnt and improved myself from her immense knowledge. Her valuable guidance helped me in all the time of doing research, writing research papers and writing this dissertation.

I would equally like to thank Dr. Philippe Rostaing, who is also my supervisor. I am particularly indebted to his advice, guidance, encouragement and inspiration, which have been invaluable over the years. He has patiently corrected my mistakes and guided me to the right direction. His ideas are always interesting and inspiring me. I still remember he used to drive me home after our technical discussions, which usually lasting until the late evening. This thesis would not have been completed without his guidance and support.

Thanks to my supervisors, the three year graduate work at Lab-STICC is one of the most wonderful period in my life. They gave me the opportunities to have two international collaborations in Canada and US. They also gave me the opportunity to improve my teaching experience in the last year of my PhD. It is not sufficient to express my gratitude to them with only few words.

I would like to thank the two reviewers of my PhD dissertation, Prof. Benoît Geller at ENSTA ParisTech and Prof. Charly Poulliat at INP-ENSEEIHT Toulouse, for their constructive comments and suggestions on my PhD report. I am also grateful to the first examiner of the jury, Prof. Olivier Berder at Université de Rennes 1. Prof. Berder was my master thesis supervisor. The experience working with him gave me a great motivation to pursuit a PhD and to develop my career as a researcher. I am equally very thankful to the second examiner of the jury, Dr. Telex M. N. Ngatched at Memorial University of Newfoundland (MUN), Canada. I was supervised by Dr. Ngatched when I was a visiting student at MUN. I very appreciate for his help and for the time he spent for the technical discussion with me via Skype, after I had comeback to France.

I would like to thank Prof. Octavia Dobre at MUN, Canada, for giving me a great opportunity to join her research group as a visiting student. It has been a great privilege to be a part of her research team. She gave me valuable help whenever I asked for assistance. Her supervision on my research works is invaluable. I have learnt from her many things that will be useful for my future scientific career and personal life. I would also like to thank Dr. Deyuan Chang, Dr. Yi Zhang, Dr. Yahia Eldemerdash and Mr. Trung Nguyen for their helps during the time I was at MUN.

I was lucky to be supervised by many experts in the field. I am very thankful to Prof. Yahong Rosa Zheng and Prof. Chengshan Xiao at Missouri University of Science and Technology (MST), USA, for allowing me to be a part of their research group as a visiting student. At MST, I have got the opportunity to study and work in a very dynamic laboratory. All of their excellent supervision, feedback, suggestions and kindness at various stages have been significantly improved my knowledge, scientific mind, and skills of critical thinking, writing and presentation. Many thanks to Mr. Leo David Fan, Mr. Juening Jin, Mr. Niaz Ahmed, Mr. Weimin Duan, Dr. Tu Nguyen, Dr. Dao Lam, Ms. Jenny Mai Huynh and the other Vietnamese friends at MST for their helps and for making my visit at MST such a great experience.

The warmest thank to my Lab-STICC colleagues at both University of Western Brittany and Telecom Bretagne for the stimulating discussions, and for providing the fun environment in which I have learnt and grown during the past three years. I also want to thank all of my friends in Brest who make my stay a memorial period.

I would also like to thank Prof. Stéphane Azou at École Nationale d'Ingénieurs de Brest (ENIB), for giving the kindly flexibility on my workload during the first period of my postdoc at ENIB. His understanding helps me to prepare my PhD defense easily.

I would like to express my special thanks to Dr. Tuan-Duc Nguyen, Prof. Tue Huynh, Dr. L. Q. Vinh Tran, Dr. T. K. Ngan Nguyen, Dr. Viet-Hoa Nguyen and Dr. Minh-Tan Pham for their constant support and encouragement. The discussions with them have strongly inspired me.

I would like to thank my parents, for their constant love, encouragement, and limitless support throughout my life. Finally, I would like to thank my fiancée, Thuy Tran, who is also a PhD, for her understanding, her helps, and her endless love.

Contents

Abbreviations

$\mathbf{a}$		vector notation (bold lowercase)
$\mathbf{A}$		matrix notation (bold uppercase)
$\mathbf{A}^\dagger$		Hermitian (conjugate transpose) of matrix $\mathbf{A}$
$\mathbf{A}^T$		transpose of matrix $\mathbf{A}$
$\mathrm{Tr}\{\mathbf{A}\}$		trace of matrix $\mathbf{A}$
$\|\mathbf{A}\|_F$		Frobenius norm of matrix $\mathbf{A}$
$\mathrm{E}[a]$		mathematical expectation of a
$\mathbf{I}_b$		identity matrix of size $b \times b$
$\|\mathbf{a}\|$		vector 2-norm
$\|\mathcal{A}\|$		cardinality of set $\mathcal{A}$
$\lfloor a \rfloor$		the greatest integer less than or equal to a
$\mathrm{diag}\{a_1, a_2, \cdots, a_n\}$		diagonal matrix with n diagonal elements $a_1, a_2, \cdots, a_n$
n_T		number of transmit antennas
n_R		number of receive antennas
b		number of datastreams
$\mathbf{H}$	$[n_R \times n_T]$	channel matrix
$\mathbf{F}$	$[n_T \times b]$	precoding matrix
$\mathbf{G}$	$[b \times n_R]$	postcoding matrix
$\mathbf{s}$	$[b \times 1]$	transmitted symbol vector
$\mathbf{H}_v$	$[b \times b]$	virtual channel matrix
$\mathbf{F}_d$	$[b \times b]$	precoding matrix for virtual channel
$\boldsymbol{\eta}$	$[b \times 1]$	additive virtual noise vector
γ		angle of the virtual channel in case $b = 2$
ρ		gain of the virtual channel in case $b = 2$
$I(\mathbf{y}, \mathbf{s})$		channel mutual information between channel output $\mathbf{y}$ and input symbol $\mathbf{s}$
$d_{\min}$		minimum Euclidean distance of the received constellation
ℓ_1		minimum squared Euclidean distance between the pair of symbol vectors, whose associated binary mappings differ by exactly one bit

List of Figures

List of Tables

Introduction

With the dramatic increase of mobile users along with the development of social media in this decade, the challenge of improving the capacity of modern radio-cellular networks must be taken up. Long Term Evolution (LTE) and LTE-Advanced (LTE-A) standards were developed in 2004 and 2008 respectively to tackle it. In fact, the development of the fifth-generation (5G) radio mobile networks is ongoing to replace the current fourth-generation (4G) LTE and 4.5G LTE-A. Nevertheless, the transition from 4G to 5G, which has effectively already begun, could take a decade or longer. Though several technologies appeared to be key ingredients for the next generation, LTE-A is expected to continue playing a vital role in the 5G era [1]. The main output of the studies into LTE and LTE-A were the standardizations for the air interface, in which the most important requirements were high data rate and high quality of service, which assure low error-rate and low latency [2–5]. Besides, as discussed in the recent surveys [6, 7], low complexity communication systems are also essential in the next 5G mobile networks. To adapt with the modern trend of technology, in this PhD thesis, we investigate the wireless communication schemes, in which simple forward error correction (FEC) codes are used for low complexity and latency. We then optimize the error-rate performance of these systems in order to be used in LTE and LTE-A.

One of the vital technologies that is exploited to catch up the specifications proposed in LTE and LTE-A is the multiple-input multiple-output (MIMO) technology. Indeed, multiple antenna wireless systems, commonly referred to as MIMO systems, have become increasingly popular since the late of 1990s after the practical demonstration and the theoretical prediction of very high spectral efficiencies in [8] and [9] respectively. The advantages of using multiple antennas at the transmitter and the receiver of a wireless MIMO system have been well exploited in the recent years [10]. By using multiple antenna transceivers, MIMO technology not only offers multiplexing and diversity gains,

but it also achieves higher conventional point-to-point link reliability in comparison with single transceiver systems [11]. The main challenge is to design a MIMO scheme that fully exploits the presence of multiple antennas. In fact, in [12, 13], multiple copies of transmitted data symbols have been proposed to map across antennas for diversity and transmission robustness. The association of this technique with iterative receivers has shown promising performance [14, 15]. More importantly, in time domain duplex closed-loop schemes, the channel state information (CSI) is readily available at the transmitter through a feedback link, which allows us to further design a precoder that is able to adapt to the channel conditions. Indeed, several kinds of linear precoders have been proposed in the literature. They were designed according to different criteria such as maximization of the minimum Euclidean distance on the received constellation (referred to as max-$d_{\min}$ [16–21]), globally maximization of the channel mutual information [22] (referred to as GOPT) and power allocation optimization (referred to as minimization of bit-error-rate (BER) [23]), weighted MMSE [24], max-SNR (maximization of the received SNR or beamforming) [25], water filling (WF) [9] or mercury/water-filling (m/WF) [26].

Unfortunately, neither the outer FEC code nor the receiver structure was taken into account in most of the designs of linear precoder except for the design of space time block codes, which only considers CSI at the receiver side [27]. The low density parity check code was considered in [22], however, the precoder is not designed for any specific receiver. Its design criterion is the global maximization of the mutual information between the finite alphabet input and the corresponding channel output. Since this precoder is globally optimized, it is referred to as GOPT precoder. A drawback of the GOPT precoder is that it requires the search for optimal precoder for each channel realization and signal-to-noise ratio, with an extremely high computational complexity. This drawback makes the GOPT precoder implementation infeasible in practice, but its performance can be used as a lower bound for other precoder design purpose. In this thesis, we consider the concatenation of the FEC encoder and the MIMO linear precoder. The investigations are twofold.

On one hand, we optimize the MIMO linear precoder assuming simple binary FEC encoder and iterative receivers. Firstly, turbo detection is considered at the receiver side. We propose, in **Chapter 3**, a precoder (referred to as $\mathbf{F}_{\ell_1}$ precoder), which significantly improves the system error-rate performance by maximizing the minimum Euclidean distance between the pair of symbols, whose associated binary patterns are different by only

one bit. In addition, by considering a direct mapping at the received constellation, we introduce a novel precoder (referred to as *EXIT-based* precoder), which adapts to the optimal mapping, to improve the system error-rate performance. We compare the proposed precoders with the max-$d_{\min}$ precoder, which shows good error-rate performance compared to other precoders of literature when maximum likelihood detection and uncoded system are considered. However, in FEC encoded systems, a good optimization criterion for MIMO precoder is channel mutual information maximization. Unfortunately, the best precoder that maximizes the channel mutual information, which is the previously mentioned GOPT precoder, can not be used over random channels due to the high computational complexity. Therefore, in **Chapter 4**, we introduce an algorithm that combines the design criteria of max-$d_{\min}$ and GOPT to propose a new low complexity optimized precoder (referred to as LCOPT precoder), which asymptotically maximizes the channel mutual information in a complexity competitive way. The proposed LCOPT precoder has fixed received constellation forms and, therefore, it can easily apply the mappings proposed in Chapter 3. Secondly, we focus our study on precoder optimization assuming turbo equalization at the receiver, which has lower complexity compared to turbo detection. However, with turbo equalization, the received symbols are initially processed by an interference canceller, where they are decomposed into parallel substreams, before being converted into soft-messages and entering the decoder. Therefore, the mapping with respect to the received constellation is not essential for this receiver. In addition, since the soft-messages fed to the decoder come from the interference canceller (not channel output), it is important to maximize the mutual information between the transmitted symbols and the symbols at interference canceller output rather than maximizing the channel mutual information. Therefore, we propose, in **Chapter 5**, a new precoder that maximizes the mutual information at the interference canceller output (referred to as Genie-optimized precoder). Simulations assuming a turbo equalization at the receiver indeed show a significant performance gain of the proposed Genie-optimized precoder compared to the GOPT precoder, which aims to maximize the channel mutual information.

On the other hand, we also consider the case when non-binary low density parity check (NB-LDPC) FEC codes are used. We firstly investigate the concatenation of NB-LDPC codes with MIMO linear precoders. Conventionally, high-order Galois field (GF) is used at the NB-LDPC encoder to increase the data rate. Each GF symbol is mapped onto one

MIMO symbol vector. Though several algorithms have been proposed in the literature to reduce the decoding complexity, the computational complexity spent for decoding the high-order field codes is still painfully expensive. Therefore, in **Chapter 6**, we propose to map multiple low-order GF symbols onto one MIMO symbol vector assuming non-binary turbo detection at the receiver side. It is proved that this mapping significantly reduces the computational complexity at the receiver. Additionally, we propose to apply MIMO linear precoding to enhance the system error-rate performance as well as to further reduce the complexity by limiting the number of internal iterations inside the NB-LDPC decoder. To complete our study about the complexity reduction for the communication systems that use NB-LDPC codes, we propose, in **Chapter 7**, a novel low complexity reliability-based hard output decoding algorithm for NB-LDPC codes.

Chapter 1

Channel coding and MIMO iterative receivers

In most of the precoder designs, channel coding was not taken into account. In other words, the systems were considered as uncoded systems. In this study, we consider the channel coding. As mentioned in the introduction, this thesis contains two main parts. In the first part, we aim to optimize the MIMO linear precoder assuming simple FEC codes (for low complexity) and iterative receivers. The recursive systematic convolutional (RSC) code is used in the first part. In the second part, we optimize the complexity and performance for the NB-LDPC encoded MIMO precoded systems. Following this structure, we respectively introduce, in Section 1.1 and Section 1.2 of this chapter, the forward error correction codes and the iterative receivers that will be considered throughout this thesis. Section 1.3 briefly introduces the extrinsic information (EXIT) chart, which is a useful tool to analyze the convergence behavior as well as the performance of iterative receivers. Section 1.4 wraps up this chapter with a conclusion.

1.1 Channel coding

We recall hereinafter the encoding and decoding processes of RSC and NB-LDPC codes. The structure of RSC codes is simple and has been well-exploited. In contrast, the presentation of NB-LDPC codes is more complicated and the studies about NB-LDPC

codes are still ongoing. In this section, after a short introduction to RSC codes, we present in more details the NB-LDPC codes.

1.1.1 Recursive systematic convolutional (RSC) codes

1.1.1.1 RSC encoder

In contrast to a block code, which encodes a finite-length input message, a convolutional encoder works as a finite-state machine that takes in a continuous sequence of information bits and produces a continuous sequence of encoded bits. Hence, a convolutional encoder can be represented by a linear finite-state register circuit. The number of register elements in the circuit is called the memory order of the encoder. The coding rate of a convolutional code is calculated by $R = \frac{k}{n}$, where, at each time instance t, k is the number of input bits and n is the number of output bits.

A recursive systematic convolutional code is a convolutional code that takes into account the output from previous state to produce the output for next state, *i.e.* the encoder output is fed back into the encoder state. Let us denote by m the memory order of the RSC code, and by $[S^{(1)}, \ldots, S^{(m)}]$ the contents of the shift-register, which shifts data from left to right. In this thesis, we mostly focus on precoder design for MIMO systems assuming outer FEC code and iterative receiver. With the purpose of considering simple FEC codes, we restrict the memory order of the RSC code used in this study to $m = 3$. The comparison in terms of error-rate performances of the considered system over different memory orders will also be illustrated in Chapter 4. An encoder with $m = 3$ is demonstrated in FIGURE 1.1. a is the input information bit, $c^{(1)}$ is the systematic bit, $c^{(2)}$ is the parity bit after encoding. The termination bit e is used to make sure that all of memory elements will return to zeros at the end of every block sequence before encoding a new block. The number of termination bits is equal to the memory order m. In FIGURE 1.1, the generator functions of the feedback and feed-forward links at input and output of the shift-register are $1 + D^2 + D^3$ and $1 + D + D^3$ respectively, where D is interpreted as a delay of one unit of time. This pair of polynomials can be represented in octal form as $(13, 15)_{octal}$.

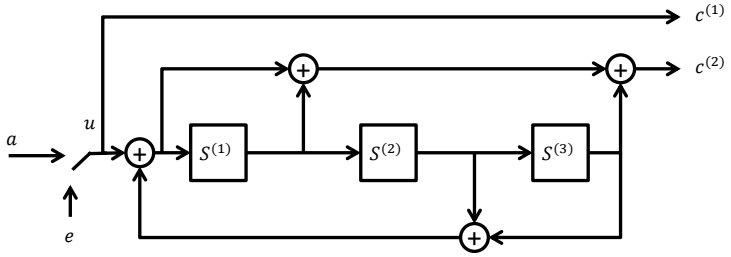


FIGURE 1.1: A rate $1/2$, $(13,15)_{octal}$ recursive systematic convolutional encoder.

1.1.1.2 BCJR soft-decoder

The BCJR algorithm was invented by Bahl, Cocke, Jelinek and Raviv in 1972 [28]. It can be applied to convolutional codes to obtain a soft-input soft-output decoder. In order to reduce the computational cost and memory requirement, the traditional BCJR algorithm can be transfered into logarithmic domain and named Log-BCJR algorithm [29]. We herein briefly recall the Log-BCJR algorithm in a practical way by using Jacobian logarithm. We restrict ourselves to binary codes. Let us start with the definition of notations.

- The information message is denoted by

$$\mathbf{u} = [u_1 \dots u_K] = [\underbrace{u_1^{(1)} \dots u_1^{(k)}}_{\substack{\underline{u}_1 \\ t=1}}; \dots; \underbrace{u_t^{(1)} \dots u_t^{(k)}}_{\substack{\underline{u}_t \\ t}}; \dots; \underbrace{u_T^{(1)} \dots u_T^{(k)}}_{\substack{\underline{u}_T \\ t=T}}]$$

  where K stands for the message binary length. We define $T = \frac{K}{k}$.

- The codeword is denoted by

$$\mathbf{c} = [c_1 \dots c_N] = [\underbrace{c_1^{(1)} \dots c_1^{(n)}}_{\substack{\underline{c}_1 \\ t=1}}; \dots; \underbrace{c_t^{(1)} \dots c_t^{(n)}}_{\substack{\underline{c}_t \\ t}}; \dots; \underbrace{c_T^{(1)} \dots c_T^{(n)}}_{\substack{\underline{c}_T \\ t=T}}]$$

  where $N = \frac{K}{R} = nT$ is the codeword length.

- The noisy message received from the channel is denoted by

$$\mathbf{v} = [v_1 \dots v_N] = [\underbrace{v_1^{(1)} \dots v_1^{(n)}}_{\substack{\underline{v}_1 \\ t=1}}; \dots; \underbrace{v_t^{(1)} \dots v_t^{(n)}}_{\substack{\underline{v}_t \\ t}}; \dots; \underbrace{v_T^{(1)} \dots v_T^{(n)}}_{\substack{\underline{v}_T \\ t=T}}]$$

- Given the memory order m, there are 2^m possible states. Let us consider a time instant t. We denote by S_r the encoder state at time instant $t-1$ and by S_s the encoder state at time instant t. In the trellis representation, a branch will connect S_r to S_s.

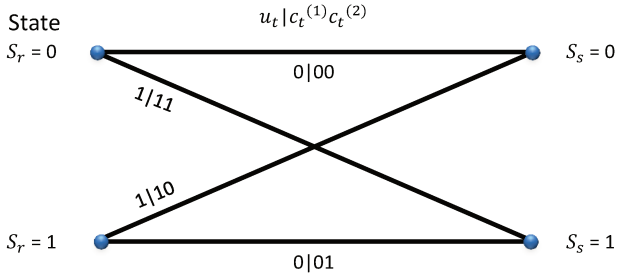


FIGURE 1.2: Trellis representation of the $(1,3)_{octal}$ RSC code. $k = 1$, $n = 2$, $c_t^{(1)} = u_t$, $c_t^{(2)} = u_t + c_{t-1}^{(2)}$ and $u_t^{(1)} = u_t$

For illustration purpose, we have represented in FIGURE 1.2. the trellis of the (1,3) RSC code. The constraint length equals 2 and the number of states is 2. A branch connecting two states is labeled by the encoder input bit value and the corresponding output bit values. In this example, u_t is original input bit, $c_t^{(1)} = u_t$ and $c_t^{(2)}$ are encoded bits. Let us define the Jacobian logarithm $\overset{*}{\max}$ as

$$\overset{*}{\max}(x,y) = \ln\left(e^x + e^y\right) = \max(x,y) + \ln\left(1 + e^{-|x-y|}\right). \tag{1.1}$$

Then

$$\ln\left(\sum_{i=1}^{n} e^{a_i}\right) = \overset{*}{\max}\left(\ldots \overset{*}{\max}\left(\overset{*}{\max}(a_1,a_2),a_3\right)\ldots,a_n\right). \tag{1.2}$$

The Log-BCJR algorithm used with $\overset{*}{\max}$ can be summarized by the following steps:

1. Calculating the log-probability of all transition paths.

   In FIGURE 1.2, S_r and S_s can take two values each, yielding four possible transitions. Let us consider a given pair (S_r, S_s). If we denote by γ_t the log of the probability of the transition between S_r and S_s at time instant t, then, in case $k = 1$ and $n = 2$ (as shown in FIGURE 1.2), γ_t is calculated as follows

$$\begin{aligned}\gamma_t(S_r,S_s) &= \ln\left[P\left(\underline{u}_t = \underline{u}_{rs}\right)P\left(\underline{v}_t|\underline{c}_{rs}\right)\right],\\ &= \ln\left[P\left(u_t = u_{rs}\right)P\left(v_t^{(1)}|c_{rs}^{(1)}\right)P\left(v_t^{(2)}|c_{rs}^{(2)}\right)\right],\end{aligned} \tag{1.3}$$

where $\underline{u}_{rs}$ is the value of $\underline{u}_t$ defining the transition from S_r to S_s in the trellis and $\underline{c}_{rs}$ is the associated encoder output. Calculation yields

$$\gamma_t(S_r, S_s) = \underbrace{\operatorname{sign}(u_t)\frac{A_t}{2} + \operatorname{sign}\left(c_t^{(1)}\right)\frac{R_t^{(1)}}{2} + \operatorname{sign}\left(c_t^{(2)}\right)\frac{R_t^{(2)}}{2}}_{\gamma'_t(S_r,S_s)} + \underbrace{\Upsilon^{u_t} + \Upsilon^{c_t^{(2)}} + \Upsilon^{c_t^{(1)}}}_{\Upsilon_t}, \tag{1.4}$$

where Υ_t is a constant at time instant t, A_t is the *a priori* log-likelihood ratio (LLR) of u_t, $R_t^{(1)}$ and $R_t^{(2)}$ are the LLRs of $v_t^{(1)}$ and $v_t^{(2)}$ respectively and sign(.) is the sign of the LLR according to each bit (*e.g.* $\operatorname{sign}(u_t = 0) = -1$ and $\operatorname{sign}(u_t = 1) = +1$).

2. Calculating the log-probability of the current state based on the previous states.

   Let us denote by α_t the log of the probability such that the current state is S_s based on the previous states, then, α_t is calculated as follows.

$$\begin{aligned}\alpha_t(S_s) = \ln\left[P\left(S^{(t)} = S_s\right)\right] &= \ln\left[\sum_i P\left(S^{(t-1)} = S_i\right) P\left(S^{(t-1)} = S_i, S^{(t)} = S_s\right)\right], \\ &= \ln\left[\sum_i e^{\alpha_{t-1}(S_i)} e^{\gamma_t(S_i,S_s)}\right],\end{aligned} \tag{1.5}$$

   where $S^{(t)}$ stands for the encoder state at time instant t. Based on FIGURE 1.2, calculation yields

$$\begin{cases} \alpha_t(0) = \sum_{n=1}^{t} \Upsilon_n + \overset{*}{\max}\left(\alpha'_{t-1}(0) + \gamma'_t(0,0), \alpha'_{t-1}(1) + \gamma'_t(1,0)\right), \\ \alpha_t(1) = \sum_{n=1}^{t} \Upsilon_n + \overset{*}{\max}\left(\alpha'_{t-1}(0) + \gamma'_t(0,1), \alpha'_{t-1}(1) + \gamma'_t(1,1)\right). \end{cases} \tag{1.6}$$

   In general,

$$\alpha_t(S_s) = \sum_{n=1}^{t} \Upsilon_n + \underbrace{\overset{*}{\max}_i\left(\alpha'_{t-1}(S_i) + \gamma'_t(S_i, S_s)\right)}_{\alpha'_t(S_s)}. \tag{1.7}$$

3. Calculating the log-probability of the current state based on the future states.

   Let us denote by β_t the log of the probability such that the current state is S_r based on the future states, then, β_t is calculated as follows

$$\begin{aligned}\beta_t(S_r) = \ln\left[P\left(S^{(t)} = S_r\right)\right] &= \ln\left[\sum_i P\left(S^{(t+1)} = S_i\right) P\left(S^{(t)} = S_r, S^{(t+1)} = S_i\right)\right], \\ &= \ln\left[\sum_i e^{\beta_{t+1}(S_i)} e^{\gamma_{t+1}(S_r,S_i)}\right].\end{aligned} \tag{1.8}$$

Calculation yields

$$\begin{cases} \beta_t(0) = \sum_{n=t}^{T} \Upsilon_n + \overset{*}{\max}\left(\beta'_{t+1}(0) + \gamma'_{t+1}(0,0), \beta'_{t+1}(1) + \gamma'_{t+1}(0,1)\right), \\ \beta_t(1) = \sum_{n=t}^{T} \Upsilon_n + \overset{*}{\max}\left(\beta'_{t+1}(0) + \gamma'_{t+1}(1,0), \beta'_{t+1}(1) + \gamma'_{t+1}(1,1)\right). \end{cases} \tag{1.9}$$

In general,

$$\beta_t(S_r) = \sum_{n=t}^{T} \Upsilon_n + \underbrace{\overset{*}{\max}_{i}\left(\beta'_{t+1}(S_i) + \gamma'_{t+1}(S_r, S_i)\right)}_{\beta'_t(S_r)}. \tag{1.10}$$

4. Calculating *a posteriori* log-probability of the current transition.

   Let us denote by $\delta_t(S_r, S_s)$ the log-probability of the transition from S_r to S_s. Calculation yields

$$\begin{aligned} \delta_t(S_r, S_s) &= \ln\left[P\left(S^{(t-1)} = S_r, S^{(t)} = S_s, \mathbf{v}\right)\right], \\ &= \ln\left[e^{\alpha_{t-1}(S_r)} e^{\gamma_t(S_r,S_s)} e^{\beta_t(S_s)}\right], \\ &= \underbrace{\sum_{n=1}^{T} \Upsilon_n}_{\Upsilon} + \underbrace{\alpha'_{t-1}(S_r) + \gamma'_t(S_r, S_s) + \beta'_t(S_s)}_{\delta'_t(S_r,S_s)}. \end{aligned} \tag{1.11}$$

5. Calculating the *a posteriori* LLR of input bit u_t.

   The LLR of the input bit u_t given the received bits $\mathbf{v}$ is calculated by

$$\begin{aligned} L(u_t|\mathbf{v}) &= \ln\left[\frac{P(u_t = 1|\mathbf{v})}{P(u_t = 0|\mathbf{v})}\right], \\ &= \ln\left[\frac{\sum_{i|u_t=1} P\left(S^{(t-1)} = S_r^{(i)}, S^{(t)} = S_s^{(i)}|\mathbf{v}\right) P(\underline{v})}{\sum_{i|u_t=0} P\left(S^{(t-1)} = S_r^{(i)}, S^{(t)} = S_s^{(i)}|\mathbf{v}\right) P(\underline{v})}\right], \\ &= \ln\left[\sum_{i|u_t=1} e^{\delta_t\left(S_r^{(i)}, S_s^{(i)}\right)}\right] - \ln\left[\sum_{i|u_t=0} e^{\delta_t\left(S_r^{(i)}, S_s^{(i)}\right)}\right]. \end{aligned} \tag{1.12}$$

   Substitution yields

$$L(u_t|\mathbf{v}) = \overset{*}{\max}_{i|u_t=1}\left(\delta'_t(S_r^{(i)}, S_s^{(i)})\right) - \overset{*}{\max}_{i|u_t=0}\left(\delta'_t(S_r^{(i)}, S_s^{(i)})\right). \tag{1.13}$$

Note that $u_t = c_t^{(1)}$. The *a posteriori* LLRs of $c_t^{(2)}$ can be computed in a similar way. In that case, the $\overset{*}{\max}$ operation is done with $c_t^{(2)}$ instead of u_t. In summary, the Log-BCJR algorithm can be carried out directly from $\alpha'_t, \gamma'_t, \beta'_t$ and δ'_t.

1.1.2 NB-LDPC code

Low density parity check (LDPC) code was pioneered by Robert Gallager in his doctoral dissertation in 1962 [30], twelve years after error correction codes were firstly introduced by Hamming in 1950. In 1981, R. Michael Tanner generalized LDPC codes and introduced a graphical representation for LDPC codes, which is widely used later and known as Tanner graph. Since 1993, with the invention of Turbo codes, researchers focused on finding a low complexity code that can approach Shannon channel capacity. Consequently, the LDPC was re-invented with the works of David Mackay [31], [32]. In 1998, Davey and Mackay proposed LDPC codes over high order Galois field GF(q), where q is the field order [33]. When $q = 2$, the code is known as binary LDPC code. When $q > 2$ the code is called Non-Binary LDPC code [34]. It is shown that, the NB-LDPC codes achieve better performance compared to their binary counterpart. On the other hand, for the moderate code lengths, the error-rate performance can be improved by increasing q [35]. Nowadays, LDPC codes are popularly used in many communication applications.

NB-LDPC codes are investigated in the second part of this document. We herein provide an overview about NB-LDPC codes. The well-known log-domain Belief-Propagation decoding algorithm for NB-LDPC codes is presented at the end of this subsection.

1.1.2.1 Parity check matrix and Tanner graph

Let us firstly introduce the code construction for binary LDPC codes. The extension to non-binary (NB-LDPC) is straightforward and consequently presented at the end of this subsection. As suggested by their name, LDPC codes are block codes with parity-check matrices that contain only a very small number of non-zero elements. The essential of this sparseness property is the complexity reduction for iterative decoders. In LDPC codes, the complexity increases linearly with the code length and the computations of decoding algorithms. Unfortunately, finding such a sparse matrix for an existing code is not practical. Therefore, for LDPC codes, the parity check matrices with low density distribution of non-zero entries are firstly constructed, the encoders are determined afterwards.

An LDPC matrix is denoted by an $M \times N$ matrix $\boldsymbol{\mathcal{H}}$. M is the number of rows (number of check-sums) and N is the number of columns (codewords length). The sparseness of the matrix is generally represented by d_v^i and d_c^j, which respectively stand for the degree of the i^{th} variable (or column weight) and the degree of the j^{th} check (or row weight). The matrix $\boldsymbol{\mathcal{H}}$ is sparse if $d_v << M$ and $d_c << N$. An LDPC matrix is called *regular* when d_v^i and d_c^j keep constant for all columns and rows. An LDPC matrix is called *irregular* when d_v^i and d_c^j do not keep constant. The construction for an irregular LDPC matrix normally based on the degree distributions, which are the percentages of different values of d_v^i and d_c^j. Finally, an LDPC matrix is called *partly regular* when only d_v^i or d_c^j keeps constant while the other is varied.

An LDPC matrix can be represented in a graphical form known as Tanner graph. Most of decoding algorithms will be based on the information exchanges in this graph to perform the decoding. Let us consider an example of regular $d_v = 2, d_c = 4, M = 4, N = 8$ binary parity check matrix $\boldsymbol{\mathcal{H}}$ as shown in (1.14). The Tanner graph representation of this matrix is presented in FIGURE 1.3.

$$\boldsymbol{\mathcal{H}} = \begin{bmatrix} 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \end{bmatrix}. \tag{1.14}$$

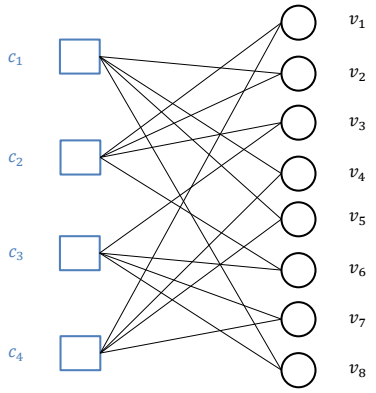


FIGURE 1.3: Tanner graph representation of a (4,8) parity check matrix.

As shown in FIGURE 1.3, the Tanner graph of a parity check matrix $\boldsymbol{\mathcal{H}}$ contains M check nodes (number of rows) and N variable nodes (number of columns). The connections between check nodes and variable nodes correspond to the positions of non-zero entries in the parity check matrix. Another important factor to qualify a good LDPC matrix construction is the *girth*. The girth of an LDPC matrix is defined as the minimum propagation steps among all of the closed-loops in the Tanner graph. In other words, we should design an LDPC matrix such that the path starting from a node and ending up at the same node is as long as possible in the Tanner graph. We should avoid *short cycle* in the design of LDPC matrix, *i.e.* girth equals 4.

When the non-zero elements in $\boldsymbol{\mathcal{H}}$ are order-q Galois field ($\mathrm{GF}(q > 2)$) symbols, then we obtain a NB-LDPC code. All calculations related to the NB-LDPC matrix are performed in $\mathrm{GF}(q)$. The NB-LDPC parity check matrix can be structurally constructed using some algorithms such as Progressive Edge Growth [36] or defining Non-Binary Cyclic and Quasi Cyclic LDPC codes [37].

1.1.2.2 NB-LDPC encoder

Construction of GF(q) symbols

We can construct codes using symbols from any Galois field $\mathrm{GF}(q)$. However, the symbols from the binary field $\mathrm{GF}(2)$ and its extension $\mathrm{GF}(2^m)$ are most widely used in digital data transmission and storage systems because information in these systems is universally coded in binary form. This subsection presents the method to construct the $\mathrm{GF}(2^m)$ symbols. The details of field algebra can be founded in [38, Chapter 2].

Let us start with a definition of polynomials, whose coefficients are taken from the binary field $\mathrm{GF}(2)$. A polynomial $f(X)$ with variable X and binary coefficients f_i is represented by

$$f(X) = f_0 + f_1 X + f_2 X^2 + \cdots + f_n X^n \tag{1.15}$$

where f_i is 0 or 1 for $0 \leq i \leq n$. The degree of the polynomial, which is denoted by m, is the largest power of X with a non-zero coefficient. In general, there are 2^n polynomials over $\mathrm{GF}(2)$ with degree $m = n$.

A polynomial $p(X)$ over $\mathrm{GF}(2)$ of degree m is called *irreducible* over $\mathrm{GF}(2)$ if $p(X)$ is not divisible by any polynomial over $\mathrm{GF}(2)$ of degree less than m but greater than zero.

An irreducible polynomial $p(X)$ of degree m is said to be *primitive polynomial* if the smallest positive integer n for which $p(X)$ divides $X^n + 1$ is $n = 2^m - 1$. For example, $p(X) = X^4 + X + 1$ divides $X^{15} + 1$ but does not divide any $X^n + 1$ such that $1 \leq n < 15$. List of primitive polynomials for different degrees is shown in [38, pp. 47].

The GF(2^m) symbols are constructed from a primitive polynomial with degree m over GF(2). Let us consider an example for the construction of GF($2^4 = 16$) symbols with degree $m = 4$ primitive polynomial $p(X) = 1 + X + X^4$. Set $p(x) = 1 + x + x^4$, we then have $x^4 = 1 + x$. From this relation we can define the polynomials for each symbol in GF(2^4): $\alpha_0 = x^{-Inf} = 0, \alpha_1 = x^0 = 1, \alpha_2 = x^1 = x, \alpha_3 = x^2, \alpha_4 = x^3, \alpha_5 = x^4 = 1 + x$. We can denote by 0 the symbol α_0 in GF(q). From x^5 we can do the back substitutions

$$\begin{aligned} \alpha_6 &= x^5 = x(1 + x) = x + x^2 \\ \alpha_7 &= x^6 = x(x^5) = x(x + x^2) = x^2 + x^3 \\ &\dots \end{aligned}$$

We finally get the representations for the elements of GF(16) which are generated from primitive polynomial $p(X) = 1 + X + X^4$.

NB-LDPC codes encoding

In high-order Galois fields, the Gauss-Jordan elimination shows its limitation to find the systematic form of $\boldsymbol{\mathcal{H}}$. In general, the NB-LDPC code encoding can be done as follows [39].

The matrix $\boldsymbol{\mathcal{H}}$ is written as $\boldsymbol{\mathcal{H}} = [\mathbf{A} \quad \mathbf{B}]$, with $\mathbf{A}$ an $M \times M$ matrix and $\mathbf{B}$ an $M \times K$ matrix. Under the systematic form, a codeword $\mathbf{v}$ is written as $\mathbf{v} = [\mathbf{r} \quad \mathbf{i}]$, with $\mathbf{r}$ the redundant symbol part and $\mathbf{i}$ the information symbol part. Therefore, the redundant vector $\mathbf{r}$ can be found by

$$\begin{aligned} &\boldsymbol{\mathcal{H}}\mathbf{v}^T = 0, \\ &\Leftrightarrow [\mathbf{A} \quad \mathbf{B}].[\mathbf{r} \quad \mathbf{i}]^T = 0, \\ &\Leftrightarrow \mathbf{r} = \mathbf{A}^{-1}\mathbf{B}\mathbf{i}, \end{aligned} \tag{1.16}$$

where $\mathbf{A}^{-1}$ is the inverse of matrix $\mathbf{A}$. However, there is an issue that the matrix $\mathbf{A}$ is not always invertible. A solution for this problem is to introduce a permutation matrix $\mathbf{P}$ of size $N \times N$ such that

$$\check{\boldsymbol{\mathcal{H}}} = \boldsymbol{\mathcal{H}}\mathbf{P} = [\check{\mathbf{A}} \quad \check{\mathbf{B}}]. \tag{1.17}$$

Matrix $\check{\mathbf{A}}$ is now invertible, therefore, there exists a vector $\check{\mathbf{v}} = [\check{\mathbf{r}} \quad \mathbf{i}]$ with $\check{\mathbf{r}} = \check{\mathbf{A}}^{-1}\check{\mathbf{B}}\mathbf{i}$.

Another approach to encode NB-LDPC codes is back-substitution. In this approach, the parity check matrix is reformed into upper triangularization form by applying columns permutation. From the upper-triangular matrix, the redundant part $\mathbf{r}$ can be calculated step by step thanks to the back-substitution method.

1.1.2.3 Log-BP soft-decoder

Variants of the belief-propagation (BP) algorithm for NB-LDPC codes can be found in [34] and [35]. An extension of the BP algorithm for NB-LDPC codes over log-domain, which is referred to as Log-BP algorithm, was proposed in [40]. The Log-BP algorithm is less expensive in practical implementation compared to the conventional one. Therefore, the Log-BP decoding algorithm for NB-LDPC codes is introduced in this subsection.

Let us start with the definition of log-likelihood ratio (LLR) in Galois field. LLR of a GF symbol is defined as a vector. Thus, we refer to LLR vector (LLRV) as the LLR of a GF symbol. LLRV of a GF symbol at variable node v_0 is defined by

$$\mathbf{L}(v_0) = [L(v_0 = \alpha_1) \dots L(v_0 = \alpha_{q-1})]^T, \tag{1.18}$$

where the elements read

$$L(v = \alpha_i) = \ln \frac{P(v = \alpha_i)}{P(v = \alpha_0)}, \tag{1.19}$$

with $P(v = \alpha_i)$ is the probability that v takes the value $\alpha_i \in GF(q)$. Box-plus operator ($\boxplus$) was introduced in [40] to compute (in terms of LLRV of each variable node) the LLRV of the linear combination of variable nodes connected to one check node. Considering $\mathbf{L}_1$ and $\mathbf{L}_2$ the LLRVs of v_1 and v_2 respectively and $\{A_1, A_2\} \in GF(q)$ two GF symbols, the box-plus operator for NB-LDPC codes is denoted by

$$\mathbf{L}(A_1 v_1 + A_2 v_2) = \boxplus(\mathbf{L}_1, \mathbf{L}_2, A_1, A_2). \tag{1.20}$$

The LLRV at the output of parity check nodes can then be simply expressed in terms of box-plus operators. The Log-BP for NB-LDPC codes in [40] is split into four main steps.

Step 1: Initialization

Let us denote by $\mathbf{L}_{ch}(v_i)$ the LLRV that a variable node v_i receives from channel. In addition, we denote by $\mathcal{N}_i$ and $\mathcal{M}_j$ the set of check nodes connected to the variable node v_i and the set of variable nodes connected to the check node c_j respectively (see FIGURE 1.4 and FIGURE 1.5). With $1 \leq i \leq |\mathcal{N}_i|$ and $1 \leq j \leq |\mathcal{M}_j|$, the messages that the variable node v_i sends to a check node c_j and vice versa are respectively defined by

$$\begin{aligned} \mathbf{V}_{ij}^{(0)} &= \mathbf{L}_{ch}(v_i), \\ \mathbf{C}_{ji}^{(0)} &= \mathbf{0}, \quad \forall j. \end{aligned} \tag{1.21}$$

Step 2: Tentative decoding

At this step, we compute the a posteriori LLRV for each variable node as

$$\mathbf{L}_{post}(v_i) = \mathbf{L}_{ch}(v_i) + \sum_{j' \in \mathcal{N}_i} \mathbf{C}_{j'i}^{(l)}. \tag{1.22}$$

From $\mathbf{L}_{post}(v_i)$, the hard decision can be easily determined at each variable node. With $k \in \{1, \ldots, q-1\}$ the component of the LLRV, we find $k_{\max} = \arg\max_k (\mathbf{L}_{post}(v_i)_k)$, then, $v_i = \alpha_0$ if $\mathbf{L}_{post}(v_i)_{k_{\max}} < 0$, otherwise $v_i = \alpha_k$. After hard decision, the most likely codeword $\tilde{\mathbf{v}}$ is determined. The syndrome is calculated by

$$\mathbf{s} = \boldsymbol{\mathcal{H}} \tilde{\mathbf{v}}^T. \tag{1.23}$$

If $\mathbf{s} = \mathbf{0}$, $\tilde{\mathbf{v}}$ is a codeword and the algorithm is stopped at this step.

Step 3: Horizontal step

This step calculates the messages from variable nodes to check nodes. The message $\mathbf{V}_{ij}^{(l)}$ from v_i to c_j at iteration l reads

$$\mathbf{V}_{ij}^{(l)} = \mathbf{L}_{ch}(v_i) + \sum_{j' \in \mathcal{N}_i \setminus j} \mathbf{C}_{j'i}^{(l-1)}. \tag{1.24}$$

FIGURE 1.4 shows the graphical representation of step 3.

Step 4: Vertical step

This step calculates the messages from check nodes to variable nodes. In NB-LDPC codes, each check node c_j should take into account the GF values of the non-zero elements

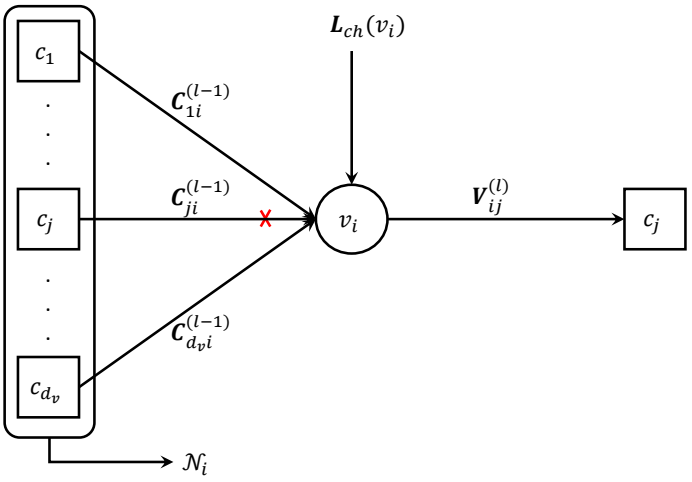


FIGURE 1.4: Graphical representation of messages calculation at variable nodes.

in $\boldsymbol{\mathcal{H}}$ that connect to it as shown in FIGURE 1.5. The message that the check node c_j sends to the adjacent variable node v_j is a LLRV, which satisfies the check sum condition. We have

$$\begin{aligned} \mathbf{C}_{ji}^{(l)} &= \mathbf{L}(h_{ji}v_i + \sum_{i' \neq i} h_{ji'}v_{i'} = 0), \\ &= \mathbf{L}(v_i = h_{ji}^{-1} \sum_{i' \neq i} h_{ji'}v_{i'}). \end{aligned} \tag{1.25}$$

To apply the definition of GF box-plus operator in (1.20), two variables were introduced in [40] as follows

$$\begin{aligned} \sigma_{ji} &= \sum_{i' \leq i} h_{ji'}v_{i'}, \\ \rho_{ji} &= \sum_{i' \geq i} h_{ji'}v_{i'}. \end{aligned} \tag{1.26}$$

Then,

$$\begin{aligned} \mathbf{L}\left(\sigma_{ji}\right) &= \mathbf{L}\left(\sigma_{j(i-1)} + h_{ji}v_i\right), \\ &= \boxplus\left(\mathbf{L}(\sigma_{j(i-1)}), \mathbf{L}(v_i), 1, h_{ji}\right). \\ \mathbf{L}\left(\rho_{ji}\right) &= \mathbf{L}\left(\rho_{j(i+1)} + h_{ji}v_i\right), \\ &= \boxplus\left(\mathbf{L}(\rho_{j(i+1)}), \mathbf{L}(v_i), 1, h_{ji}\right), \end{aligned} \tag{1.27}$$

where $\mathbf{L}(v_i)$ is $\mathbf{V}_{ij}^{(l-1)}$. This means that we split $\mathcal{M}_j$ in FIGURE 1.5 into upper and lower parts and recursively calculate $\mathbf{L}(\sigma_{ji})$ and $\mathbf{L}(\rho_{ji})$ as in (1.27) with $i \in \{1, \ldots, d_v\}$. The

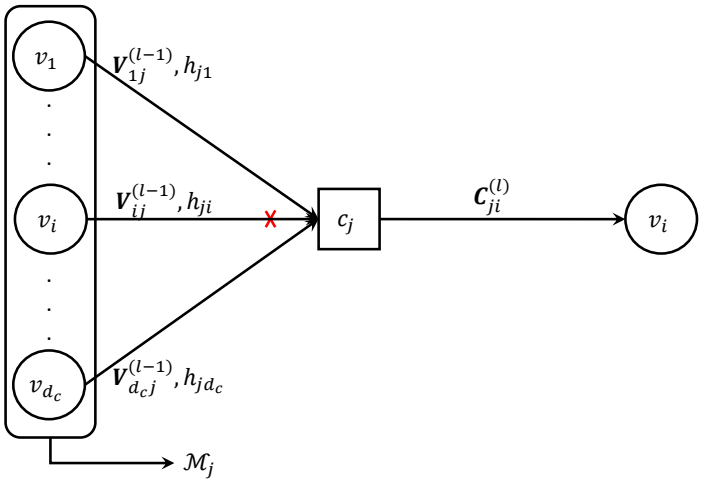


FIGURE 1.5: Graphical representation of messages calculation at check nodes.

message from check node c_j to variable node v_i is finally calculated by

$$\begin{aligned} \mathbf{C}_{ji}^{(l)} &= \mathbf{L}\left(h_{ji}^{-1}\sigma_{j(i-1)} + h_{ji}^{-1}\rho_{j(i+1)}\right), \\ &= \boxplus\left(\mathbf{L}\left(\sigma_{j(i-1)}\right), \mathbf{L}\left(\rho_{j(i+1)}\right), h_{ji}^{-1}, h_{ji}^{-1}\right). \end{aligned} \tag{1.28}$$

The algorithm will run until it meets the stopping condition at step 2 or until the maximum number of iterations is reached.

The complexity of this algorithm will be later discussed in Chapter 6. The most time consuming step of the Log-BP algorithm is the vertical step (step 4). Several decoding algorithms were proposed to reduce the decoding complexity, especially at the vertical step, namely Log-BP-FFT [41] (fast Fourier transform (FFT)-based) and EMS [42](extended min-sum, which significantly reduces the decoding complexity over high-order GF, *e.g.* $q = 64, 128, \ldots$). However, by considering a low to moderate order of GF (*e.g.* $q = 16$ as will be considered in this thesis), the complexity of the presented Log-BP algorithm is acceptable. Moreover, it is quasi-optimal for large-girth parity check matrices. Therefore, Log-BP algorithm is considered in Chapter 6 of this thesis. The study about complexity of decoding algorithms for NB-LDPC codes will be carried out in Chapter 7.

1.2 Iterative receivers for MIMO wireless communications

Let us consider a communication scheme that takes into account the concatenation of FEC encoder and symbol mapping. The mapped symbols are then precoded before the

transmission. The optimal receiver is the maximum likelihood (ML) detection of the information message from the channel output. However, it is infeasible in practice to implement this kind of optimal receiver. Therefore, iterative receiver is used to asymptotically achieve the performance of the ML solution. In this section, we introduce the MIMO wireless communication systems and the two iterative receivers, turbo detection and turbo equalization, which will be later exploited in this study.

1.2.1 MIMO transmission

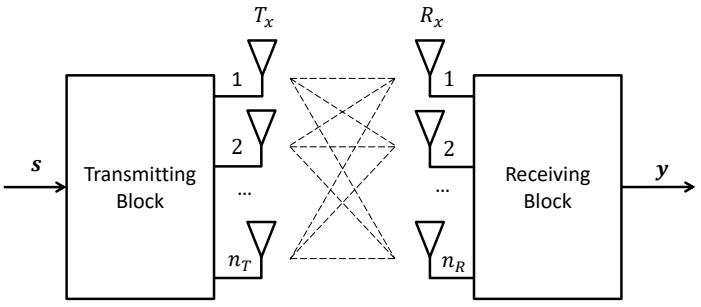


FIGURE 1.6: MIMO model with n_T transmit antennas and n_R receive antennas

MIMO technology has become essential in the recent years to take up the challenges of higher data rate and increasing data traffic that radio-cellular networks have to face up. It is one of the most crucial distinction between 3G and 4G wireless systems [10]. The idea of using multiple transceiver antennas not only offers the multiplexing and diversity gains, but it also achieves higher conventional point-to-point link reliability in comparison with single transmitter and single receiver systems [11]. Because of these properties, MIMO has become an important part of modern wireless communication standards such as IEEE 802.11ac/n (WiFi), 3GPP LTE & LTE-A (4G & 4.5G) and the upcoming 5G.

Let us consider a MIMO transmission with n_T transmit and n_R receive antennas. We assume that the channel is time-invariant and non-frequency selective over the data transmission. The basic MIMO system model is illustrated in FIGURE 1.6. The received signal at antenna j reads

$$y_j = \sum_{i=1}^{n_T} h_{j,i} s_i + n_j, \tag{1.29}$$

where $h_{j,i}$ is the channel gain of the path from the transmit antenna i to the receive antenna j, s_i is the complex transmit signal at antenna i, and n_j is the noise at the

receive antenna j. In general, the received vector $\mathbf{y}$ reads

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{n}, \tag{1.30}$$

where $\mathbf{y} = [y_1, y_2, ..., y_{n_R}]^T$, and $\mathbf{s} = [s_1, s_2, ..., s_{n_T}]^T$ is the transmitted symbol vector, $\mathbf{H}$ is the channel matrix, and $\mathbf{n}$ is the noise vector. The channel matrix $\mathbf{H}$, which represents $n_R \times n_T$ connection paths between n_T transmitting and n_R receiving antennas, is defined by

$$\mathbf{H} = \begin{pmatrix} h_{1,1} & \cdots & h_{1,n_T} \\ \vdots & \ddots & \vdots \\ h_{n_R,1} & \cdots & h_{n_R,n_T} \end{pmatrix}. \tag{1.31}$$

The channel matrix components are random and chosen based on different statistical models. The noise is considered as an additive white Gaussian noise (AWGN) and its elements n_j are independent from each other and have a complex circularly-symmetric Gaussian distribution.

1.2.2 Turbo detection

Turbo detection was firstly introduced in [43]. Later applied in many MIMO systems [44], it shows significant error-rate performance improvement compared to non-iterative MIMO detection. It is based on the well-known turbo principle [45]. It consists of a MIMO symbol soft demapper and a soft decoder, which iteratively perform the maximum *a posteriori* (MAP) detection and the soft decoding to enhance the system error-rate performance. FIGURE 1.7 shows the turbo detection model assuming bit-interleaved coded modulation and the MIMO transmission as shown in (1.30). $L^1_{\mathrm{A}}, L^1_{\mathrm{P}}$ and L^1_{E} respectively stand for the *a priori*, the *a posteriori* and the extrinsic log likelihood ratios (LLRs) of the soft demapper, while those values for the soft decoder are $L^2_{\mathrm{A}}, L^2_{\mathrm{P}}$ and L^2_{E}.

From (1.30), let us assume $\mathrm{E}[\mathbf{n}\mathbf{n}^\dagger] = \sigma_n^2 \mathbf{I}_{n_R}$ and $\mathrm{E}[\mathbf{s}\mathbf{s}^\dagger] = \sigma_s^2 \mathbf{I}_{n_T}$ and let $\mathbf{I}_{n_R}$ be the identity matrix of size n_R. In addition, let $\mathcal{Q}^{n_T}$ be the set of symbol vectors with the mapping rule defined by

$$\left(\alpha_1^k, \ldots, \alpha_{\mathfrak{q}}^k\right)_{\alpha_\ell^k \in \{0,1\}} \to \mathbf{s}_k \in \mathcal{Q}^{n_T},$$

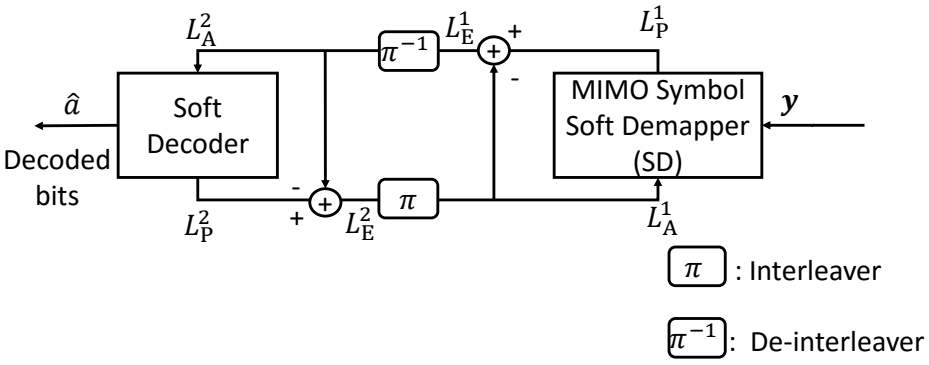


FIGURE 1.7: Turbo detection.

where $\mathtt{q} = \log_2(|\mathcal{Q}^{n_T}|)$. The conditional probability density function (pdf) of the received vector $\mathbf{y}$ is defined by

$$p(\mathbf{y}|\mathbf{s} = \mathbf{s}_k) = \frac{1}{(\pi\sigma_n^2)^{n_R}} \exp\left(-\frac{\|\mathbf{y} - \mathbf{H}\mathbf{s}_k\|^2}{\sigma_n^2}\right). \tag{1.32}$$

Given $\mathbf{y}$, the *a posteriori* LLR of the bit at position i is calculated by

$$L_P^1(i) = \ln \frac{P(\alpha_i = 1|\mathbf{y})}{P(\alpha_i = 0|\mathbf{y})}, \tag{1.33}$$

where both nominator and denominator are defined by

$$\begin{aligned} P(\alpha_i = \varepsilon|\mathbf{y}) &= \sum_{\mathbf{s}_k \in \mathcal{Q}^{n_T}} P(\alpha_i = \varepsilon, \mathbf{s} = \mathbf{s}_k|\mathbf{y}), \\ &= \sum_{\mathbf{s}_k \in \mathcal{Q}^{n_T}|\alpha_i = \varepsilon} P(\mathbf{s} = \mathbf{s}_k|\mathbf{y}), \quad \varepsilon \in \{0, 1\}. \end{aligned} \tag{1.34}$$

Applying (1.32) with the Bayesian rule, we obtain

$$L_P^1(i) = \ln \frac{\sum_{\mathbf{s}_k \in \mathcal{Q}^{n_T}|\alpha_i = 1} p(\mathbf{y}|\mathbf{s} = \mathbf{s}_k) P(\mathbf{s} = \mathbf{s}_k)}{\sum_{\mathbf{s}_k \in \mathcal{Q}^{n_T}|\alpha_i = 0} p(\mathbf{y}|\mathbf{s} = \mathbf{s}_k) P(\mathbf{s} = \mathbf{s}_k)}. \tag{1.35}$$

Hence, from (1.32) and (1.35)

$$L_P^1(i) = \ln \frac{\sum\limits_{\mathbf{s}_k \in \mathcal{Q}^{n_T}|\alpha_i = 1} \exp\left(-\frac{\|\mathbf{y} - \mathbf{H}\mathbf{s}_k\|^2}{\sigma_n^2}\right) P(\mathbf{s} = \mathbf{s}_k)}{\sum\limits_{\mathbf{s}_k \in \mathcal{Q}^{n_T}|\alpha_i = 0} \exp\left(-\frac{\|\mathbf{y} - \mathbf{H}\mathbf{s}_k\|^2}{\sigma_n^2}\right) P(\mathbf{s} = \mathbf{s}_k)}. \tag{1.36}$$

At first iteration, the *a priori* probability is set to $P(\mathbf{s} = \mathbf{s}_k) = \frac{1}{|\mathcal{Q}^{n_T}|}$ (equiprobability). From the second iteration, it is computed from L_{A}^1 (which reads value from the interleaved L_{E}^2, see FIGURE 1.7) by

$$P\left(\mathbf{s} = \mathbf{s}_k | L_{\mathrm{A}}^1\right) = \prod_{i=1}^{q} P\left(\alpha_i = \alpha_i^k | L_{\mathrm{A}}^1\right), \tag{1.37}$$

where the probabilities of each bit are calculated by

$$P(\alpha_i = \varepsilon | L_{\mathrm{A}}^1) = \frac{\exp\left((2\varepsilon - 1)\frac{L_A^1(i)}{2}\right)}{\exp\left(\frac{L_A^1(i)}{2}\right) + \exp\left(-\frac{L_A^1(i)}{2}\right)}, \quad \varepsilon \in \{0, 1\}. \tag{1.38}$$

The computational complexity for $L_P^1(i)$ can be reduced by using the Jacobian logarithm as follows

$$\begin{aligned} L_P^1(i) = &\max_{\mathbf{s}_k \in \mathcal{Q}^{n_T} | \alpha_i = 1}{}^{*} \left(-\frac{\|\mathbf{y} - \mathbf{H}\mathbf{s}_k\|^2}{\sigma_n^2} + \sum_{j=1}^{q} (2\alpha_j^k - 1)\frac{L_A^1(j)}{2} \right) \\ &- \max_{\mathbf{s}_k \in \mathcal{Q}^{n_T} | \alpha_i = 0}{}^{*} \left(-\frac{\|\mathbf{y} - \mathbf{H}\mathbf{s}_k\|^2}{\sigma_n^2} + \sum_{j=1}^{q} (2\alpha_j^k - 1)\frac{L_E^1(j)}{2} \right). \end{aligned} \tag{1.39}$$

1.2.3 Turbo equalization

Turbo equalization, which was first introduced in [46] and subsequently investigated in [47, 48], has become essential to take up the challenge of data transmission over a channel with intersymbol interference. A turbo equalizer consists of an interference canceller and a soft decoder, which iteratively exchange extrinsic information through symbol-to-binary (SBC) and binary-to-symbol converters (BSC). In modern turbo equalization, the iterative information exchanges are taken into account not only between the minimum mean square error (MMSE)-based interference canceller (MMSE IC) and the soft decoder, but also between the soft decoder and the symbol-to-binary converter. Assuming bit-interleaved coded modulation and MIMO transmission, the turbo equalization model at the receiver side is shown in FIGURE 1.8, where $L_{\mathrm{A}}^1, L_{\mathrm{P}}^1$ and L_{E}^1 respectively stand for the *a priori*, the *a posteriori* and the extrinsic log likelihood ratios (LLRs) of the SBC, while the equivalent notations for the soft decoder are $L_{\mathrm{A}}^2, L_{\mathrm{P}}^2$ and L_{E}^2.

By applying the turbo principle under an intersymbol interference cancellation criterion, turbo equalization achieves a good system error-rate performance [49–52]. In addition, the complexity of turbo equalization is less than that of turbo detection. Indeed, thanks to the interference canceller, the received signal can be separated into parallel substreams

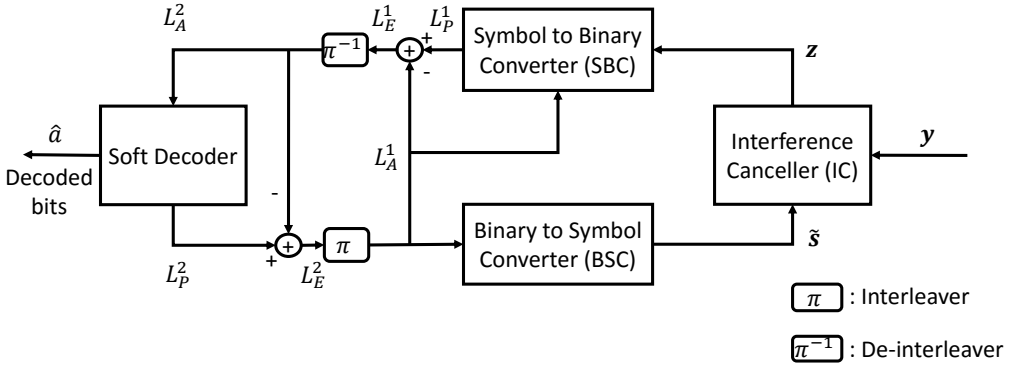


FIGURE 1.8: Turbo Equalization.

and demapped independently as will be presented hereinafter. Since the MAP detection (SBC block) is performed for each substream, it significantly reduces the computational complexity.

1.2.3.1 Interference canceller

The interference canceller consists of a feed-forward and a feedback filters, which are respectively denoted by $\mathbf{W}$ and $\mathbf{Q}$. At the output of the Interference Canceller, the detected vector, which is denoted by $\mathbf{z}$, reads

$$\mathbf{z} = \mathbf{W}\mathbf{y} - \mathbf{Q}\tilde{\mathbf{s}}, \tag{1.40}$$

where $\tilde{\mathbf{s}}$ is an estimation of the transmitted symbol vector $\mathbf{s}$ (BSC output) and $\mathbf{W}$ $(n_T \times n_R)$, $\mathbf{Q}$ $(n_T \times n_T)$ are obtained by using the MMSE criterion. Let us denote by $\Lambda^{\mathrm{IC,in}}$ the *a priori* LLR input of BSC and by $\omega^{\mathrm{IC,in}}$ the *a priori* LLR input of SBC. We assume $\mathrm{E}[\tilde{\mathbf{s}}] = \mathrm{E}[\mathbf{s}] = 0$ (zero mean modulation) and $\mathrm{E}[\tilde{\mathbf{s}}\mathbf{s}^\dagger] = \mathrm{E}[\tilde{\mathbf{s}}\tilde{\mathbf{s}}^\dagger] = \sigma_{\tilde{\mathbf{s}}}^2 \times \mathbf{I}_{n_T}$. The mean square error function is defined by $\epsilon = \mathrm{E}\big[\|\mathbf{z}-\mathbf{s}\|^2\big] = \mathrm{E}\big[\mathrm{Tr}\big\{(\mathbf{z}-\mathbf{s})(\mathbf{z}-\mathbf{s})^\dagger\big\}\big]$. The problem is to minimize the mean square error ϵ under constraint $\mathbf{Q}_{ii} = 0 \ \forall i$. The constraint implies that the diagonal elements of matrix $\mathbf{Q}$ are zeros, *i.e.* the calculation of symbol $\mathbf{z}_i$ does not depend on the estimated symbol $\tilde{\mathbf{s}}_i$ at the same time. In other words, the constraint means that only the inter-symbol interference has to be canceled. The optimization problem can be written as

$$\begin{cases} \min\limits_{\mathbf{W},\mathbf{Q}} \mathrm{E}\big[\|\mathbf{z}-\mathbf{s}\|^2\big], \\ \text{subject to} \quad \mathbf{Q}_{ii} = 0 \quad \forall i. \end{cases} \tag{1.41}$$

Let us define $\mathbf{B} = (\sigma_s^2 - \sigma_{\tilde{s}}^2)\mathbf{H}\mathbf{H}^\dagger + \sigma_n^2\mathbf{I}_{n_R}$. Using the Lagrangian multipliers, the optimization problem yields

$$\mathbf{W}_{k,:} = \sigma_s^2\mathbf{H}_{:,k}^\dagger\left(\mathbf{B} + \sigma_{\tilde{s}}^2\mathbf{H}_{:,k}\mathbf{H}_{:,k}^\dagger\right)^{-1}, \tag{1.42}$$

and

$$\mathbf{Q}_{k,:} = \mathbf{W}_{k,:}\mathbf{H} - \mathbf{W}_{k,:}\mathbf{H}_{:,k}e_k, \tag{1.43}$$

where e_k is the k^{th} row of $\mathbf{I}_{n_T}$, $\mathbf{H}_{:,k}$ and $\mathbf{H}_{k,:}$ respectively denote the k^{th} column and k^{th} row of $\mathbf{H}$.

Let us define $\left(\mathbf{B}+\sigma_{\tilde{s}}^2\mathbf{H}_{:,k}\mathbf{H}_{:,k}^\dagger\right)^{-1} = \mathbf{C}$. Then, the computation cost of $\mathbf{W}_{k,:}$ can be reduced by using the Woodbury's theorem, which yields

$$\mathbf{C} = \mathbf{B}^{-1} - \frac{\sigma_{\tilde{s}}^2\mathbf{B}^{-1}\mathbf{H}_{:,k}\mathbf{H}_{:,k}^\dagger\mathbf{B}^{-1}}{1 + \sigma_{\tilde{s}}^2\mathbf{H}_{:,k}^\dagger\mathbf{B}^{-1}\mathbf{H}_{:,k}}. \tag{1.44}$$

Hence, we can also deduce the following expression

$$0 < \mathbf{W}_{k,:}\mathbf{H}_{:,k} = \frac{\sigma_s^2\mathbf{H}_{:,k}^\dagger\mathbf{B}^{-1}\mathbf{H}_{:,k}}{1 + \sigma_{\tilde{s}}^2\mathbf{H}_{:,k}^\dagger\mathbf{B}^{-1}\mathbf{H}_{:,k}} = \mu_k < 1. \tag{1.45}$$

From (1.40) we deduce that

$$\begin{aligned} \mathbf{z} &= \mathbf{W}\mathbf{H}\mathbf{s} - \mathbf{Q}\tilde{\mathbf{s}} + \mathbf{W}\mathbf{n}, \\ &= \mathbf{W}\mathbf{H}\mathbf{s} - (\mathbf{W}\mathbf{H} - \boldsymbol{\mu})\tilde{\mathbf{s}} + \mathbf{W}\mathbf{n}, \end{aligned} \tag{1.46}$$

where $\boldsymbol{\mu}$ is a $n_T \times n_T$ diagonal matrix, whose diagonal elements are μ_k, $k \in \{1, \ldots, n_T\}$. Hence,

$$z_k = \mu_k s_k + \underbrace{\sum_{m \neq k} \mathbf{W}_{k,:}\mathbf{H}_{:,m}(s_m - \tilde{s}_m) + \mathbf{W}_{k,:}\mathbf{n}}_{\xi_k}. \tag{1.47}$$

Finally, the IC output at the k^{th} substream can be modeled as follows

$$z_k = \mu_k s_k + \xi_k \quad \text{for } k \in \{1, \ldots, n_T\}, \tag{1.48}$$

where ξ_k is independent from s_k, has Gaussian distribution with zero mean and variance $\sigma_{\xi_k}^2 = \sigma_s^2\mu_k(1 - \mu_k)$. Thus, the signal-to-noise ratio at IC output of the k^{th} substream,

which is denoted by p_k, reads

$$p_k = \frac{\mu_k^2 \sigma_s^2}{\sigma_{\xi_k}^2} = \frac{\mu_k}{1-\mu_k}, \tag{1.49}$$

where, from (1.42),(1.44) and (1.45), calculation yields

$$1-\mu_k = \frac{1+\sigma_{\tilde{s}}^2 \mathbf{H}_{:,k}^{\dagger} \mathbf{B}^{-1} \mathbf{H}_{:,k} - \sigma_s^2 \mathbf{H}_{:,k}^{\dagger} \mathbf{B}^{-1} \mathbf{H}_{:,k}}{1+\sigma_{\tilde{s}}^2 \mathbf{H}_{:,k}^{\dagger} \mathbf{B}^{-1} \mathbf{H}_{:,k}}. \tag{1.50}$$

Then, the SNR p_k is calculated by

$$p_k = \frac{\sigma_s^2 \mathbf{H}_{:,k}^{\dagger} \mathbf{B}^{-1} \mathbf{H}_{:,k}}{1+(\sigma_{\tilde{s}}^2 - \sigma_s^2) \mathbf{H}_{:,k}^{\dagger} \mathbf{B}^{-1} \mathbf{H}_{:,k}}. \tag{1.51}$$

At the optimum convergence state, *i.e.* $\sigma_s^2 = \sigma_{\tilde{s}}^2, \tilde{\mathbf{s}} = \mathbf{s}$, the following expressions hold

$$z_k = \mu_k s_k + \underbrace{\mathbf{W}_{k,:}\mathbf{n}}_{\xi_k}, \tag{1.52}$$

$$\mathbf{W}_{k,:} = \sigma_s^2 \mathbf{H}_{:,k}^{\dagger} \left(\sigma_{\tilde{s}}^2 \mathbf{H}_{:,k} \mathbf{H}_{:,k}^{\dagger} + \sigma_n^2 \mathbf{I}_{n_R}\right)^{-1}, \tag{1.53}$$

$$\mu_k = \frac{\frac{\sigma_s^2}{\sigma_n^2} \mathbf{H}_{:,k}^{\dagger} \mathbf{H}_{:,k}}{1+\frac{\sigma_s^2}{\sigma_n^2} \mathbf{H}_{:,k}^{\dagger} \mathbf{H}_{:,k}}, \tag{1.54}$$

and

$$p_k = \frac{\sigma_s^2}{\sigma_n^2} \mathbf{H}_{:,k}^{\dagger} \mathbf{H}_{:,k}. \tag{1.55}$$

1.2.3.2 SBC and BSC converters

Let $\mathcal{Q}$ be the set of Q-ary modulation symbols, with the mapping rule defined by $\left(\alpha_1^{\ell}, \ldots, \alpha_q^{\ell}\right)_{\alpha_i^{\ell} \in \{0,1\}} \rightarrow s_{\ell} \in \mathcal{Q}$, where $q = \log_2(Q)$. Then, thanks to (1.48), the LLRs at the output of SBC (see FIGURE 1.8) can be calculated by a low-complexity procedure, in a similar way as presented in (1.39), as follows

$$\begin{aligned} L_{P,k}^1(i) = & \max_{s_{\ell} \in \mathcal{Q} | \alpha_i = 1}^{*} \left(-\frac{|y_k - \mu_k s_{\ell}|^2}{\sigma_{\xi_k}^2} + \sum_{j=1}^{q} (2\alpha_j^{\ell} - 1) \frac{L_{A,k}^1(j)}{2} \right) \\ & - \max_{s_{\ell} \in \mathcal{Q} | \alpha_i = 0}^{*} \left(-\frac{|y_k - \mu_k s_{\ell}|^2}{\sigma_{\xi_k}^2} + \sum_{j=1}^{q} (2\alpha_j^{\ell} - 1) \frac{L_{A,k}^1(j)}{2} \right). \end{aligned} \tag{1.56}$$

On the other hand, the symbol $\tilde{s}_k$ on the k^{th} stream of the IC output is estimated by

$$\begin{aligned}\tilde{s}_k = \mathrm{E}\left[s_k | L^1_{A,k}\right] &= \sum_{s_\ell \in \mathcal{Q}} s_\ell P(s_k = s_\ell | L^1_{A,k}), \\ &= \sum_{s_\ell \in \mathcal{Q}} s_\ell \prod_{i=1}^{q} P(\alpha_i = \alpha_i^\ell | L^1_{A,k}),\end{aligned} \tag{1.57}$$

where $P(\alpha_i = \alpha_i^\ell | L^1_{A,k})$ can be calculated from $L^1_{A,k}$, which reads value from the interleaved $L^2_{E,k}$, in a similar way as presented in (1.38). The estimated vector $\tilde{\mathbf{s}}$ is then used in the next iteration to find the IC output as shown in (1.40).

1.3 EXIT chart

Introduced by Stephan ten Brink in 1999 [53, 54] and later widely applied in iterative concatenated system analyses [55, 56], the extrinsic information transfer (EXIT) chart is a useful tool to analyze the convergence behavior of a soft-in/soft-out iterative receiver by tracking its mutual information (MI) transfer characteristic. For example, EXIT chart allows us to predict the minimum required number of iterations for convergence as well as the minimum SNR required for convergence. EXIT chart is constructed from two extrinsic transfer functions (EXIT functions) of two components of a iterative receiver. Each curve plots the MI of the extrinsic log-likelihood ratios (LLRs) versus the MI of the *a priori* LLRs of each receiver component. Note that the two EXIT functions in the EXIT chart are measured independently.

Let us denote by L_P the *a posteriori* LLR at the output of a receiver component, while the *a priori* LLR, which is fed back from the other receiver component, is denoted by L_A. The extrinsic LLR (L_E) is calculated by subtracting L_A from L_P. In this thesis, we only investigate the serial concatenations, whose inner receiver component is the detector and the outer receiver component is the decoder. For the iterative receivers presented in Section 1.2, the detectors are the soft-demapper (in the case of turbo detection) and the MMSE IC revolved around the SBC and BSC (in the case of turbo equalization). One should note that, to measure the EXIT function of detector, we have to take into account not only the *a priori* LLR from the decoder, but also the LLR from channel output, which depends on the signal-to-noise ratio (SNR).

In order to measure the EXIT function of one receiver component (detector or decoder), we need to select a set of I_A (*e.g.* I_A ranges from 0 to 1 in the binary case) that we want to measure for the corresponding set of I_E. For each value of I_A, we apply the following procedure, which is presented by 3 main steps.

Step 1: The inputs of a soft-in/soft-out receiver component are LLR values. Thus, I_A should be transformed into L_A. The *a priori* information from the partner decoder could be modeled using an AWGN channel. This assumption follows two conditions. Firstly, for a large interleaver, the *a priori* values are uncorrelated from the channel observation. Secondly, the probability density functions of L_A can be approached from a Gaussian distribution. Thanks to the Gaussian assumption, the following symmetry condition [57] holds $\mathrm{E}[L_A] = \sigma_A^2/2$. We can then generate the *a priori* LLRs as

$$L_A = \frac{\sigma_A^2}{2} X + B_A, \tag{1.58}$$

where B_A is Gaussian distributed with zero mean and variance σ_A^2, $X \in \{-1, +1\}$ are symbols of the binary sequence at input of the corresponding encoding component. With the symmetry condition, for any pair of binary random variable X and the corresponding LLR L, the mutual information between X and L can be calculated by

$$I(X, L) = 1 - \int_{-\infty}^{+\infty} p_L(\tau | X = +1) \log_2[1 + e^{-\tau}] \, d_\tau. \tag{1.59}$$

Therefore, from (1.58) and (1.59), we can express I_A as a function of σ_A, which is known as $J(.)$ function. $J(.)$ reads

$$I_A = J(\sigma_A) \triangleq 1 - \int_{-\infty}^{+\infty} \frac{e^{-\frac{(\tau - \sigma_A^2/2)^2}{2\sigma_A^2}}}{\sqrt{2\pi}\sigma_A} \log_2[1 + e^{-\tau}] \, d_\tau. \tag{1.60}$$

Let us denote by $J^{-1}(.)$ the reverse function of $J(.)$. Hence, σ_A can be calculated from I_A by

$$\sigma_A = J^{-1}(I_A). \tag{1.61}$$

The $J(.)$ and $J^{-1}(.)$ functions could be closely approximated as follows

$$J(\sigma) \approx \left(1 - 2^{H_1 \sigma^{2H_2}}\right)^{H_3}, \tag{1.62}$$

$$J^{-1}(I_A) \approx \left(-\frac{1}{H_1}\log_2\left(1 - I_A^{\frac{1}{H_3}}\right)\right)^{\frac{1}{2H_2}}. \tag{1.63}$$

The numerical optimization by minimizing the total squared difference between (1.60) and (1.62) gives $H_1 = 0.3073$, $H_2 = 0.8935$, and $H_3 = 1.1064$ [58]. With this approximation, we can firstly compute σ_A from I_A by (1.61) and subsequently generate L_A from σ_A by (1.58).

Step 2: The *a posteriori* LLRs L_P will be obtained at the output of the component. For the decoder in serial concatenated schemes, L_P is only calculated from the L_A generated in step 1. For the detector in serial concatenated schemes, L_P is calculated from both the L_A and L_{ch}, which is the channel output LLR calculated by taking the input X at the transmitter. Thus, the extrinsic LLRs are calculated as follows

$$L_E = L_P - L_A. \tag{1.64}$$

Step 3: This step aims to calculate the extrinsic mutual information I_E from L_E. Due to the nonlinearity of the component, the LLR distribution of the extrinsic output is unknown and no longer Gaussian. However, we can rewrite (1.59) as $I(X,L) = 1 - E\{\log_2(1+e^{-L})\}_{X=+1}$. The expectation can be replaced by the time average and we can measure the mutual information from a large number N of samples even for non-Gaussian or unknown distributions as follows.

$$\begin{aligned} I(X,L) &= 1 - \mathrm{E}\left[\log_2(1+e^{-\mathrm{sign}(L)|L|})\right]_{X=+1}, \\ &\approx 1 - \frac{1}{N}\sum_{n=1}^{N}\left[(1-P_{e_n})\log_2(\underbrace{1+e^{-|L|}}_{1/(1-P_{e_n})}) + P_{e_n}\log_2(\underbrace{1+e^{|L|}}_{1/P_{e_n}})\right], \\ &\approx 1 - \frac{1}{N}\sum_{n=1}^{N} H_b(P_{e_n}). \end{aligned} \tag{1.65}$$

where H_b is the binary entropy and $P_{e_n} = P(\mathrm{sign}(L) = -1|X = +1) = \frac{1}{1+e^{|L|}}$ and $P(\mathrm{sign}(L) = +1|X = +1) = 1 - P_{e_n}$. The extrinsic MI I_E is then measured by (1.65).

To plot an EXIT chart for any iterative receiver presented in Section 1.2, we firstly apply the procedure above for both of the detector and decoder (the EXIT function of detector takes into account not only the *a priori* LLR from the decoder, but also the LLR from channel output). Let us denote by (I_A^1, I_E^1) and (I_A^2, I_E^2) the pairs of *a priori*

and extrinsic mutual informations of the detector and decoder respectively. A complete EXIT chart is consequently obtained by plotting I_A^1 versus I_E^1 and I_E^2 versus I_A^2 in only one figure.

1.4 Conclusion

The primary purpose of this chapter is to review briefly the principal characteristics of the systems and tools that will be used in this document. Firstly, we described RSC and NB-LDPC codes and the corresponding soft decoding algorithms. After that, MIMO iterative receivers were briefly introduced. These iterative receivers include turbo detection and turbo equalization. Finally, we presented the extrinsic information transfer (EXIT) chart, which is a useful tool to analyze the characteristic of these iterative receivers. Thanks to the EXIT chart, we can easily analyze the convergence behavior of the iterative receivers when they are used with MIMO precoding techniques. Next chapter focuses on MIMO linear precoding and schemes used as references in this PhD are described in details.

Chapter 2

MIMO linear precoding techniques

2.1 Precoded MIMO systems

In MIMO wireless communications, several techniques have been studied to exploit the presence of multiple transceiver antennas. These techniques can be categorized into three main groups: spatial multiplexing (SM), diversity coding, and precoding.

- Spatial multiplexing splits a high data rate signal into n_T independent data-streams and each stream is transmitted by a transmit antenna. Spatial multiplexing can be used without CSI at transmitter (CSI-T) but CSI at receiver (CSI-R) is required. A drawback of this technique is the error-rate performance loss. In practice, it is also limited to small Tx and Rx numbers due to receiver complexity and antenna correlation [59].

- Diversity coding applies a static space-time coding form onto the transmitting symbols before transmission. It exploits the diversity gain to achieve a higher reliability compared to single transceiver systems. Diversity coding does not require the CSI-T but it requires the CSI at the receiver. The two well-known techniques in this group are space-time block codes (STBCs) and space-time trellis codes (STTCs) [60].

- Precoding is a technique of pre-processing the modulated symbol vector by multiplying it with a precoding matrix before transmission. The precoding matrix exploits the CSI-T and is designed according to different performance criteria [61].

In LTE and LTE-A specifications, the CSI is expected to be readily available at the transmitter and the receiver. Therefore, precoding plays an essential role in the LTE and LTE-A standards in order to provide high quality of services.

A basic precoding system structure, which contains a precoder $\mathbf{F}$ and a postcoder $\mathbf{G}$, is shown in FIGURE 2.1. A codeword c, which is output from a forward error correction (FEC) encoder, is grouped and mapped onto complex modulation symbols s. These symbols are then converted by a serial-to-parallel converter to form a modulated symbol vector $\mathbf{s}$. The precoder processes the symbol vector $\mathbf{s}$ before transmission according to the channel state information. At the receiver side, a postcoder is considered for post-processing. We assume average unit transmit power. Thus, to conserve the total average transmit power, the precoder must satisfy the following condition

$$\mathrm{Tr}\{\mathbf{F}\mathbf{F}^{\dagger}\} = 1. \tag{2.1}$$

The power allocations are different according to the design criterion, the signal to noise ratio, and the CSI. The received symbol vector at output of the postcoder, which is

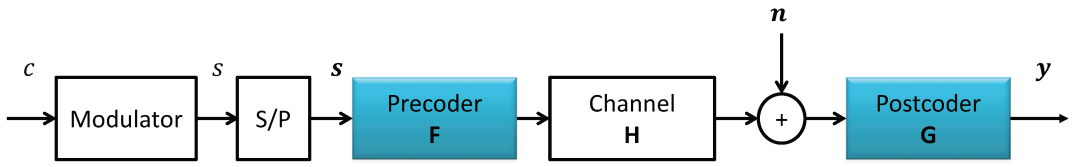


FIGURE 2.1: Precoding schema.

denoted by $\mathbf{y}$, reads

$$\mathbf{y} = \mathbf{GHFs} + \mathbf{Gn}, \tag{2.2}$$

where $\mathbf{n}$ is a vector of additive noise samples. The received symbol vector $\mathbf{y}$ is then detected to obtain an estimation of the transmitted codeword c.

2.2 Channel transformation technique

Thanks to the precoding matrix, we can control the number of independent data streams that are transmitted over the MIMO system. Let us consider the MIMO system presented in (2.2) thanks to which we want to transmit b independent data streams. Hence, the

modulated symbol vector $\mathbf{s}$ has size $b \times 1$, $\mathbf{F}$ is the $n_T \times b$ precoding matrix, $\mathbf{H}$ is the $n_R \times n_T$ channel matrix, $\mathbf{G}$ is the $b \times n_R$ postcoding matrix, and $\mathbf{n}$ is the $n_R \times 1$ additive noise vector. One should note that $b \leq \text{rank}(\mathbf{H}) \leq \min(n_T, n_R)$, so n_T and n_R can be larger than b. We assume $\text{E}[\mathbf{ss}^\dagger] = \mathbf{I}_b, \text{E}[\mathbf{sn}^\dagger] = 0$ and $\text{E}[\mathbf{nn}^\dagger] = \mathbf{R}_n$, where $\mathbf{I}_b$ is the identity matrix of size $b \times b$ and $\mathbf{R}_n$ is the noise covariance matrix.

Assuming the channel state information (CSI) is perfectly known at both the transmitter and receiver, channel diagonalization and noise whitening techniques can be applied to transform the model presented in (2.2) into a simpler form. The operation is decomposed in three steps and is referred to as virtual transformation. Let us define $\mathbf{F}_d$ and $\mathbf{F}_v$ such that $\mathbf{F} = \mathbf{F}_v\mathbf{F}_d$. The decompositions of the two matrices $\mathbf{F}_v$ and $\mathbf{G}$ as the product of three matrices yield

$$\mathbf{F}_v = \mathbf{F}_1\mathbf{F}_2\mathbf{F}_3 \qquad \text{and} \qquad \mathbf{G} = \mathbf{G}_1\mathbf{G}_2\mathbf{G}_3, \tag{2.3}$$

where $(\mathbf{F}_i, \mathbf{G}_i)$ performs the particular operations which are detailed hereinafter.

2.2.1 Noise whitening

Let us consider the eigenvalue decomposition of the noise covariance matrix, which yields

$$\mathbf{R}_n = \text{E}[\mathbf{nn}^\dagger] = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\dagger, \tag{2.4}$$

where $\mathbf{Q}$ is an unitary matrix and $\mathbf{\Lambda}$ is a diagonal matrix. This step aims to transform the additive noise vector into white circularly-symmetric complex gaussian noise vector. To achieve that goal, we impose that the correlation matrix $\mathbf{R}_{v_1} = \text{E}[\mathbf{G_1 nn}^\dagger\mathbf{G_1}^\dagger] = \mathbf{G_1Q\Lambda Q}^\dagger\mathbf{G_1}^\dagger$ becomes $\sigma_\eta^2\mathbf{I}_{n_R}$. The matrix $\mathbf{G_1}$ is therefore defined by

$$\mathbf{G_1} = \sigma_\eta\mathbf{\Lambda}^{-1/2}\mathbf{Q}^\dagger. \tag{2.5}$$

The intermediate channel matrix at this step, which is denoted by $\mathbf{H}_{v_1}$, reads

$$\mathbf{H}_{v_1} = \mathbf{G}_1\mathbf{H}\mathbf{F}_1, \tag{2.6}$$

where $\mathbf{F}_1$ is an identity matrix of size n_T.

2.2.2 Channel diagonalization

At this step, in order to diagonalize the channel, we apply the singular value decomposition (SVD) to the intermediate matrix $\mathbf{H}_{v_1}$ as follows.

$$\mathbf{H}_{v_1} = \mathbf{A}\mathbf{\Sigma}\mathbf{B}^{\dagger}, \tag{2.7}$$

where $\mathbf{A}$ and $\mathbf{B}^{\dagger}$ are unitary matrices, and $\mathbf{\Sigma}$ is a diagonal matrix whose elements represent the square roots of all eigenvalues of the matrix $\mathbf{H}_{v_1}\mathbf{H}_{v_1}^{\dagger}$. Note that these eigenvalues are real positive numbers and sorted in decreasing order. The number of non-null eigenvalues depends on the rank of the matrix $\mathbf{H}_{v_1}$

$$k = \text{rank}(\mathbf{H}_{v_1}) \leq \min(n_T, n_R). \tag{2.8}$$

The diagonal matrix $\mathbf{\Sigma}$ can be then expressed by separating the non-null eigenvalues from the null ones.

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \tag{2.9}$$

where the matrix $\mathbf{\Sigma}_k$ contains all of the non-null eigenvalues. In order to diagonalize the intermediate channel matrix $\mathbf{H}_{v_1}$, we must select

$$\mathbf{F}_2 = \mathbf{B} \quad \text{and} \quad \mathbf{G}_2 = \mathbf{A}^{\dagger}. \tag{2.10}$$

Hence, the second intermediate channel matrix, which is denoted by $\mathbf{H}_{v_2}$, is a diagonal matrix that reads

$$\mathbf{H}_{v_2} = \mathbf{G}_2\mathbf{H}_{v_1}\mathbf{F}_2 = \mathbf{\Sigma}. \tag{2.11}$$

In addition, the correlation matrix $\mathbf{R}_{v_2}$ is then given by

$$\mathbf{R}_{v_2} = \mathbf{G}_2\mathbf{R}_{v_1}\mathbf{G}_2^{\dagger} = \sigma_{\eta}^2\mathbf{G}_2\mathbf{G}_2^{\dagger} = \sigma_{\eta}^2\mathbf{I}_{n_R}, \tag{2.12}$$

since $\mathbf{G}_2$ is an unitary matrix.

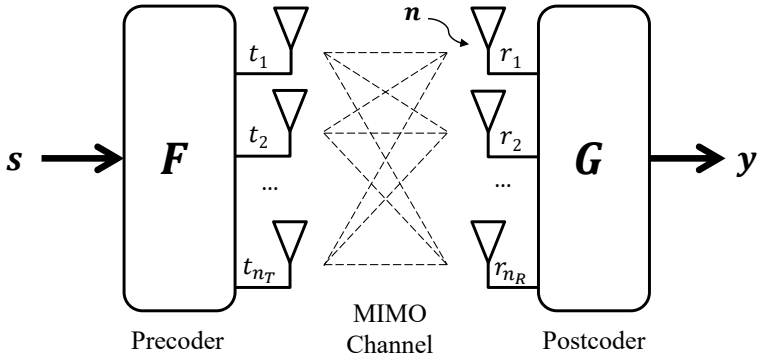


FIGURE 2.2: MIMO precoding system structure.

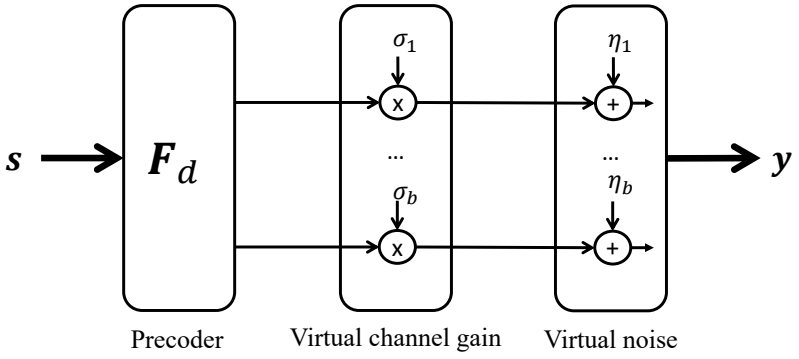


FIGURE 2.3: Equivalent MIMO precoding system structure.

2.2.3 Dimensionality reduction

The diagonal matrix $\mathbf{H}_{v_2}$ consists of the subchannel gains, which appear in decreasing order on its main diagonal. The goal of this operation is to adjust the dimension to the number of desired data-streams b in case $b \leq k$. The matrices $\mathbf{F}_3$ and $\mathbf{G}_3$ are then defined by

$$\mathbf{F}_3 = \begin{pmatrix} \mathbf{I}_b \\ \mathbf{0} \end{pmatrix} \quad \text{and} \quad \mathbf{G}_3 = (\mathbf{I}_b \quad \mathbf{0}). \tag{2.13}$$

The resulting matrix is given by

$$\mathbf{H}_v = \mathbf{G}_3 \mathbf{H}_{\mathbf{v_2}} \mathbf{F}_3 = \mathbf{\Sigma}_b, \tag{2.14}$$

where $\mathbf{\Sigma}_b$ represents the b largest singular values of $\mathbf{H}_{v_2}$. The correlation matrix of the noise is also truncated, which reads

$$\mathbf{R}_\eta = \sigma_\eta^2 \mathbf{I}_b. \tag{2.15}$$

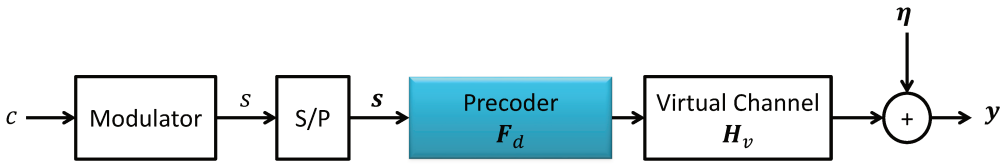


FIGURE 2.4: Precoding schema after channel transformation.

Finally, we obtain the following equivalent writing of (2.2)

$$\mathbf{y} = \mathbf{H}_v\mathbf{F}_d\mathbf{s} + \boldsymbol{\eta}, \tag{2.16}$$

where $\boldsymbol{\eta}$ is the $b \times 1$ additive white circularly-symmetric complex gaussian virtual noise vector with $\mathrm{E}[\boldsymbol{\eta}\boldsymbol{\eta}^\dagger] = \sigma_\eta^2\mathbf{I}_b$. The matrix $\mathbf{H}_v = \mathrm{diag}(\sigma_1, ..., \sigma_b)$ is the $b \times b$ eigen-channel matrix, where $\{\sigma_1, ..., \sigma_b\}$ are the b most significant singular values of $\mathbf{H}$ sorted in descending order. $\mathbf{F}_d$ is the precoding matrix to be optimized according to one or several criteria. It satisfies the power constraint $\|\mathbf{F}_d\|_F^2 = 1$. FIGURE 2.2 and FIGURE 2.3 respectively show the structures of the MIMO precoded system before and after applying the channel transformation. The resulting equivalent transmission schema after channel transformation is presented in FIGURE 2.4.

2.3 Existing precoders

The precoding techniques can be classified into two categories: diagonal and non-diagonal schemes. A precoder is called diagonal when the precoding matrix $\mathbf{F}_d$ in (2.16) is a diagonal matrix. The general working structure of the diagonal precoders is illustrated in the Figure 2.5. The principle is to find the power allocation expressed by the diagonal entries of $\mathbf{F}_d$, which are denoted by f_i, to optimize a particular criterion. In contrast, a precoder is called non-diagonal when the precoding matrix $\mathbf{F}_d$ is not a diagonal matrix.

There are a variety of diagonal and non-diagonal precoders that have been investigated so far in the literature. In this section, we present only the selected precoders that aim to maximize the channel capacity, which is an essential optimization criterion for precoder in MIMO encoded systems as will be illustrated hereinafter in Section 2.3.3.

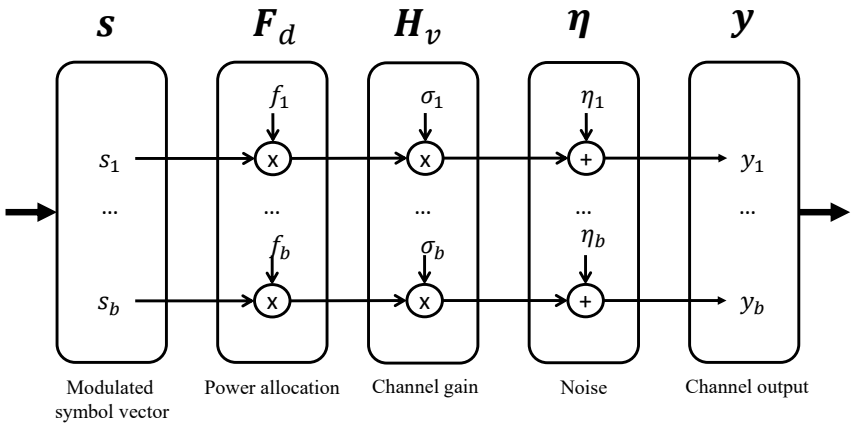


FIGURE 2.5: Diagonal precoding structure.

2.3.1 Diagonal precoders

Let us denote by $\mathbf{F}_d = \text{diag}(f_1, \ldots, f_b)$ the diagonal precoding matrix. Many researches have been carried out to optimize this diagonal precoding matrix according to different optimization criteria such as the received signal to noise ratio (SNR) maximization (max-SNR precoder [25]), the weighted mean square error minimization (MMSE precoder [24]), power allocation optimization (minimization of bit-error-rate (BER) [23]), and quality of service [24]. Another criterion, on which we focused during this PhD is the mutual information maximization. Diagonal precoders optimized with respect to this criterion are water-filling (WF) precoder [9] and its variant mercury/water-filling (M/WF) precoder [26, 62].

2.3.1.1 Water-filling precoder

This precoder aims to maximize the capacity of the precoded MIMO system with Gaussian inputs. The capacity of a virtual channel can be simplified as

$$C = \sum_{i=1}^{b} \log_2(1 + f_i^2 \sigma_i^2), \quad \text{with} \quad \sum_{i=1}^{b} f_i^2 = 1. \tag{2.17}$$

The optimized solution is given by

$$f_i^2 = \begin{cases} \Psi_{\text{WF}} - \frac{1}{\sigma_i^2} & \text{if } \Psi_{\text{WF}} > \frac{1}{\sigma_i^2} \\ 0 & \text{otherwise} \end{cases} \quad \text{with} \quad i = 1, \ldots, b \tag{2.18}$$

where the threshold Ψ_{WF} depends on the virtual channel and is defined by

$$\Psi_{\mathrm{WF}} = \frac{1+\gamma_{\mathrm{WF}}}{b_{\mathrm{WF}}} \quad \text{with} \quad \gamma_{\mathrm{WF}} = \sum_{i=1}^{b_{\mathrm{WF}}} \frac{1}{\sigma_i^2} \tag{2.19}$$

with b_{WF} the number of subchannels used by the water-filling precoder.

The water-filling precoder removes some subchannels and spreads power on the others to improve the channel capacity. As implied by its name, the water-filling precoder can be interpreted by the simple concept illustrated in FIGURE 2.6. The principle is to pour water into each unit vessel, which has initial solid level $\frac{1}{\sigma_i^2}$, up to a fixed water level Ψ_{WF}. The amount of water in each unit vessel then represents f_i^2. Finally, the power f_i^2 will be allocated to the corresponding subchannel.

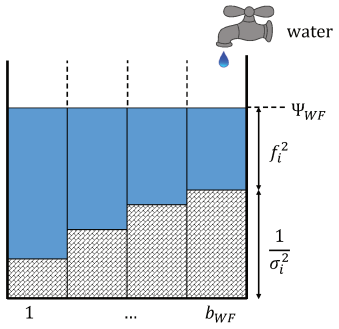


FIGURE 2.6: Water-filling concept.

2.3.1.2 Mercury/water-filling precoder

Similar to the ordinary water-filling precoder, the mercury/water-filling precoder searches for the best power allocation on every subchannel to maximize the channel capacity for Q-ary inputs. We firstly define a minimum mean square error (MMSE) in function of f_i, which reads $\mathrm{MMSE}_i(f_i) = \mathrm{E}\left[|s_i - \hat{s}_i(y_i, f_i)|^2\right]$, where $\hat{s}_i(y_i, f_i) = \mathrm{E}[s_i|y_i, f_i]$. Let us denote by $\mathrm{MMSE}_i^{-1}(.)$ the inverse function of the $\mathrm{MMSE}_i(.)$ function, with domain equal to $[0,1]$ and $\mathrm{MMSE}_i^{-1}(1) = 0$, *i.e.* $\mathrm{MMSE}_i^{-1}(\mathrm{MMSE}_i(f_i)) = f_i$. The power for each subchannel is then allocated by

$$f_i^2 = \frac{1}{\sigma_i^2}\mathrm{MMSE}_i^{-1}\left(\min\left\{1, \frac{\kappa}{\sigma_i^2}\right\}\right), \tag{2.20}$$

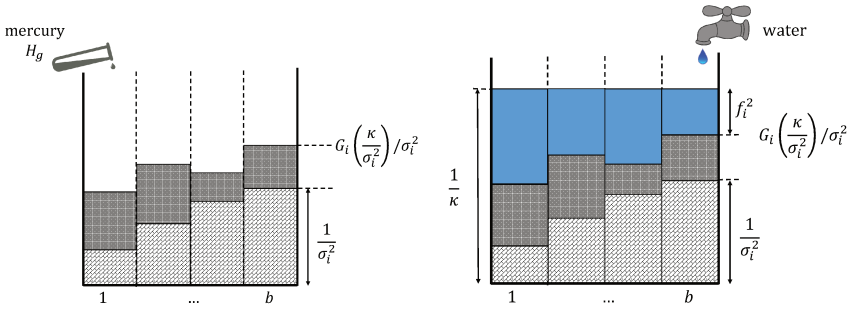
where κ is found by solving the following equation

$$\sum_{\substack{i=1 \\ \sigma_i^2 > \kappa}}^{b} \frac{1}{\sigma_i^2} \mathrm{MMSE}_i^{-1}\left(\frac{\kappa}{\sigma_i^2}\right) = 1. \tag{2.21}$$

Let us define a function $G_i(\zeta)$ such that

$$G_i(\zeta) = \begin{cases} 1/\zeta - \mathrm{MMSE}_i^{-1}(\zeta), & \zeta \in [0,1] \\ 1. & \zeta > 1 \end{cases} \tag{2.22}$$

Thanks to the $G_i(.)$ function, the mercury/water-filling precoder can be interpreted by a simple concept illustrated in FIGURE 2.7. The principle is to firstly pour mercury into each unit vessel, which has initial solid level $\frac{1}{\sigma_i^2}$, up to a mercury level $G_i(\frac{\kappa}{\sigma_i^2})/\sigma_i^2$. Consequently, water is poured into each unit vessel up to a fixed water level $\frac{1}{\kappa}$. The amount of water in each unit vessel then represents f_i^2. Finally, the power f_i^2 will be allocated to the corresponding subchannel.



(a) Mercury-filling (b) Water-filling

FIGURE 2.7: Mercury/water-filling concept.

2.3.2 Non-diagonal precoders

In this subsection we present two non-diagonal precoders. The first precoder, which globally maximizes the channel mutual information, is referred to as globally optimal (GOPT) precoder [22]. The GOPT precoder searches for the optimal precoding solution by using an optimization algorithm. However, the computational complexity of the algorithm is painfully high. The second precoder is max-$d_{\min}$ precoder, which was firstly introduced by Collin *et al.* in [16]. This precoder aims to maximize the minimum Euclidean distance

($d_{\min}$) between the received constellation symbols. Maximizing $d_{\min}$ asymptotically maximizes the lower bound of the channel mutual information. Therefore, this precoder can force the channel mutual information to a higher value and achieve a channel capacity that is close to the one exhibited by GOPT precoder.

2.3.2.1 Globally optimal precoder

For non-diagonal precoders, the SVD of $\mathbf{F}_d$ gives $\mathbf{F}_d = \mathbf{U}_\mathbf{F}\mathbf{\Sigma}_\mathbf{F}\mathbf{V}_\mathbf{F}^\dagger$. It will be proved in Chapter 4 that the matrix $\mathbf{U}_\mathbf{F}$ can always be chosen to coincide with the identity matrix $\mathbf{I}_b$ in order to maximize the channel mutual information, *i.e.* $\mathbf{U}_\mathbf{F} = \mathbf{I}_b$. Hence, by rewriting $\mathbf{\Psi} = \mathbf{\Sigma}_\mathbf{F}$, $\mathbf{\Theta} = \mathbf{V}_\mathbf{F}^\dagger$, we deduce

$$\mathbf{F}_d = \mathbf{\Psi}\mathbf{\Theta}. \tag{2.23}$$

The matrix $\mathbf{\Psi}$ controls the power allocation on each subchannel, while $\mathbf{\Theta}$ concerns itself with the form of the received constellation.

With the equivalent model (2.16), the channel mutual information between the discrete input $\mathbf{s}$ and the channel output $\mathbf{y}$ is given by

$$\mathcal{I}(\mathbf{y},\mathbf{s}) = b\log_2 Q - \frac{1}{Q^b}\sum_{m=1}^{Q^b} \mathrm{E}\left[\log_2 \sum_{k=1}^{Q^b} e^{-\zeta_{m,k}}\right], \tag{2.24}$$

where $\zeta_{m,k} = \left(\|\mathbf{H}_v\mathbf{F}_d(\mathbf{s}_m - \mathbf{s}_k) + \boldsymbol{\eta}\|^2 - \|\boldsymbol{\eta}\|^2\right)/\sigma_\eta^2$ and Q is the cardinality of the Q-ary modulation. Let us define $\mathbf{W} = \mathbf{F}_d^\dagger\mathbf{H}_v^\dagger\mathbf{H}_v\mathbf{F}_d$. It is proved that the mutual information $\mathcal{I}(\mathbf{y},\mathbf{s})$ is a concave function with respect to $\mathbf{W}$ and $\mathbf{\Psi}^2$.

With $\mathbf{F}_d = \mathbf{\Psi}\mathbf{\Theta}$, the authors in [22] proposed an iterative algorithm that respectively updates $\mathbf{\Theta}$ and $\mathbf{\Psi}$ based on the gradient ascent method (instead of directly updating $\mathbf{F}_d$, which seems to be infeasible). Though $\mathbf{\Psi}$ can be updated directly using its gradient, we however must rely on the incremental of $\mathbf{W}$ to update $\mathbf{\Theta}$, which significantly increases the computational complexity of the algorithm. The algorithm is referred to as GOPT algorithm and will be presented in further details in Chapter 4.

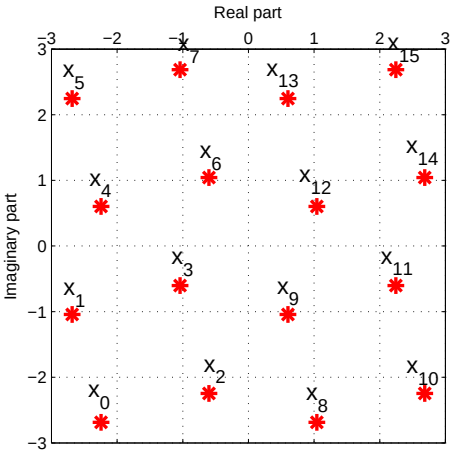


FIGURE 2.8: Received constellation $\mathbf{x}_i$ on the first sub-channel in case $\mathbf{F}_d = \mathbf{F}_{r_1}$.

2.3.2.2 max-$d_{\min}$ precoder

max-$d_{\min}$ precoder aims to optimize the matrix $\mathbf{F}_d$ so as to maximize the minimum Euclidean distance, denoted by $d_{\min} = \min_{m \neq \ell} \|\mathbf{x}_m - \mathbf{x}_\ell\|$ where $\mathbf{x} = \mathbf{H}_v \mathbf{F}_d \mathbf{s}$, between the received constellation symbols.

Solution for $b = 2$ data streams

The max-$d_{\min}$ precoder was firstly proposed in [16] and the solution was given for the case $b = 2$ data streams with 4-QAM modulation. With $b = 2$, the two eigenvalues of the $\mathbf{H}_v$ are rewritten as

$$\begin{cases} \sigma_1 = \rho \cos\gamma \\ \sigma_2 = \rho \sin\gamma \end{cases} \Leftrightarrow \begin{cases} \rho = \sqrt{\sigma_1^2 + \sigma_2^2} \\ \gamma = \arctan \frac{\sigma_1}{\sigma_2} \end{cases}. \tag{2.25}$$

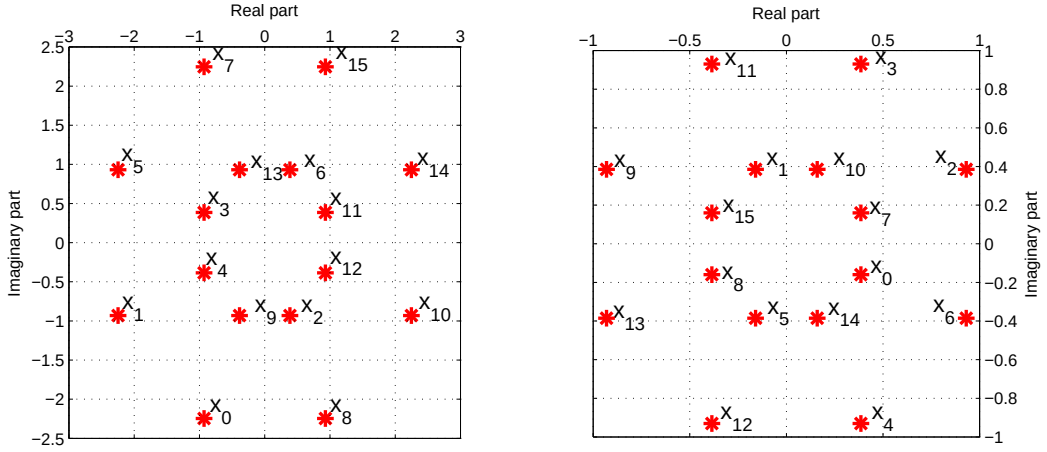
Hence, the conversion from cartesian to polar form of $\mathbf{H}_v$ gives

$$\mathbf{H}_v = \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} = \rho \begin{pmatrix} \cos\gamma & 0 \\ 0 & \sin\gamma \end{pmatrix}, \tag{2.26}$$

where ρ and γ respectively represent the channel gain and angle. As $\sigma_1 \geq \sigma_2 > 0$, we have $0 < \gamma \leq \pi/4$. Hence, the optimal solution depends on γ and by defining the threshold $\gamma_0 = \arctan \sqrt{\frac{\sqrt{2}-1}{2\sqrt{2}+\sqrt{6}-1}}$ ($\approx 17,28^0$), $\mathbf{F}_d$ reads

- if $0 \leq \gamma \leq \gamma_0$

$$\mathbf{F}_d = \mathbf{F}_{r_1} = \begin{pmatrix} \sqrt{\frac{3+\sqrt{3}}{6}} & \sqrt{\frac{3-\sqrt{3}}{6}} e^{i\frac{\pi}{12}} \\ 0 & 0 \end{pmatrix}, \tag{2.27}$$



(a) First sub-channel. (b) Second sub-channel.

FIGURE 2.9: Received constellation $\mathbf{x}_i$ on the first and second sub-channels in case $\mathbf{F}_d = \mathbf{F}_{octa}$.

- if $\gamma_0 < \gamma \leq \pi/4$

$$\mathbf{F}_d = \mathbf{F}_{octa} = \frac{1}{\sqrt{2}} \begin{pmatrix} \cos\psi & 0 \\ 0 & \sin\psi \end{pmatrix} \begin{pmatrix} 1 & e^{i\frac{\pi}{4}} \\ -1 & e^{i\frac{\pi}{4}} \end{pmatrix}, \tag{2.28}$$

where $\psi = \arctan\frac{\sqrt{2}-1}{\tan\gamma}$. In the case $\gamma \leq \gamma_0$, *i.e.* $\mathbf{F}_d = \mathbf{F}_{r_1}$, the precoder only spreads power on the first sub-channel. FIGURE 2.8 shows the received constellation on the first sub-channel of precoder max-$d_{\min}$ in this case. The constellation is similar to the one of 16-QAM modulation with a rotation by 15^o in each quadrant. On the other hand, in the case $\gamma > \gamma_0$, *i.e.* $\mathbf{F}_d = \mathbf{F}_{octa}$, the precoder spreads power on both sub-channels, where the received constellations are shown in FIGURE 2.9. It is shown that, thanks to the optimization criterion, a pair of neighbor symbols in the first sub-channel, *e.g.* $\mathbf{x}_3$ and $\mathbf{x}_4$, is separated in the second sub-channel.

FIGURE 2.10 shows the received $d_{\min}$ normalized by ρ for $\mathbf{F}_{r1}$ and $\mathbf{F}_{octa}$. We can see that, in order to keep the high value of the received normalized $d_{\min}$, the max-$d_{\min}$ precoder uses γ_0 as a threshold to switch between $\mathbf{F}_{r1}$ and $\mathbf{F}_{octa}$. Note that γ_0 is signal-to-noise ratio (SNR)-independent and designed for uncoded system. The extension of max-$d_{\min}$ precoder for higher-order modulation in case $b = 2$ data streams was proposed in [17, 63]. We will later recall in Chapter 6 the solution for 16-QAM modulation.

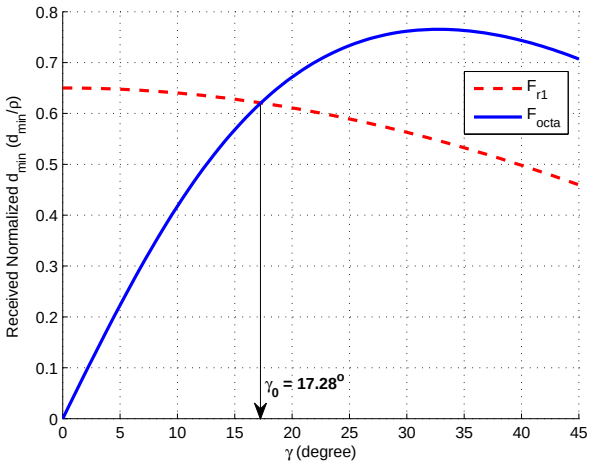


FIGURE 2.10: The received normalized $d_{\min}$ versus γ.

Solution for $b = 3$ data streams

An extension of max-$d_{\min}$ precoder for the case $b = 3$ data streams was proposed in [20, 63]. With $b = 3$, $\mathbf{H}_v$ can be rewritten as

$$\mathbf{H}_v = \rho \begin{pmatrix} \cos\gamma_1 & 0 & 0 \\ 0 & \sin\gamma_1 \cos\gamma_2 & 0 \\ 0 & 0 & \sin\gamma_1 \sin\gamma_2 \end{pmatrix}, \tag{2.29}$$

where ρ, γ_1 and γ_2 represent the channel gain and the channel angles respectively. As the diagonal elements of $\mathbf{H}_v$ are sorted in decreasing order, $0 \leq \gamma_2 \leq \pi/4$ and $\cos\gamma_2 \leq \operatorname{cotan}\gamma_1$.

The parameterized form of max-$d_{\min}$ precoder is given by

$$\mathbf{F}_d = \mathbf{\Psi} \underbrace{\mathbf{B}_\theta \mathbf{B}_\varphi}_{\mathbf{\Theta}}, \tag{2.30}$$

where $\mathbf{B}_\varphi = \begin{pmatrix} 1 & 0 & 0 \\ 0 & e^{i\varphi_2} & 0 \\ 0 & 0 & e^{i\varphi_3} \end{pmatrix}$, $\mathbf{B}_\theta = \begin{pmatrix} \mathfrak{c}_1 & \mathfrak{s}_1\mathfrak{c}_2 & \mathfrak{s}_1\mathfrak{s}_2 \\ \mathfrak{s}_1\mathfrak{c}_3 & -\mathfrak{c}_1\mathfrak{c}_2\mathfrak{c}_3 - e^{i\varphi_1}\mathfrak{s}_2\mathfrak{s}_3 & -\mathfrak{c}_1\mathfrak{s}_2\mathfrak{c}_3 + e^{i\varphi_1}\mathfrak{c}_2\mathfrak{s}_3 \\ \mathfrak{s}_1\mathfrak{s}_3 & -\mathfrak{c}_1\mathfrak{c}_2\mathfrak{s}_3 + e^{i\varphi_1}\mathfrak{s}_2\mathfrak{c}_3 & -\mathfrak{c}_1\mathfrak{s}_2\mathfrak{s}_3 - e^{i\varphi_1}\mathfrak{c}_2\mathfrak{c}_3 \end{pmatrix}$ and $\mathbf{\Psi}$ is power allocation matrix. We define $\mathfrak{c}_i = \cos\theta_i$ and $\mathfrak{s}_i = \sin\theta_i$ for $i = 1, .., 3$ with $0 \leq \theta_i \leq 90^o$ and $\varphi_i \leq 360^o$. If we introduce $\mathbf{\Theta} = \mathbf{B}_\theta \mathbf{B}_\varphi$, the expression of $\mathbf{F}_d$ becomes similar to (2.23).

In the case of 4-QAM modulation, the authors in [20, 63] proposed a precoder that switches among three following precoders: $\mathbf{F}_{qc1}$, $\mathbf{F}_{qc2}$ and $\mathbf{F}_{qc3}$. The $\mathbf{F}_{qc1}$ precoder

$\mathbf{F}_{qc_2}$	θ_1	θ_2	θ_3	φ_1	φ_2	φ_3
(a)	44.49197	30.59366	39.65316	0	161.56505	0
(b)	32.34322	37.85164	56.71270	180	0	45
(c)	62.52239	22.59606	66.97236	85.31834	21.52669	118.15496
(d)	37.42924	22.5	38.45324	180	90	135
$\mathbf{F}_{qc_3}$	θ_1	θ_2	θ_3	φ_1	φ_2	φ_3
(a)	42.33339	45	50.63553	90	155.25922	24.74077
(b)	52.86439	40.77576	53.32112	115.27892	145.43734	72.71867
(c)	52.01812	45	90	0	45	135

TABLE 2.1: Optimized angles in degree for the precoders $\mathbf{F}_{qc_2}$ and $\mathbf{F}_{qc_3}$ [20].

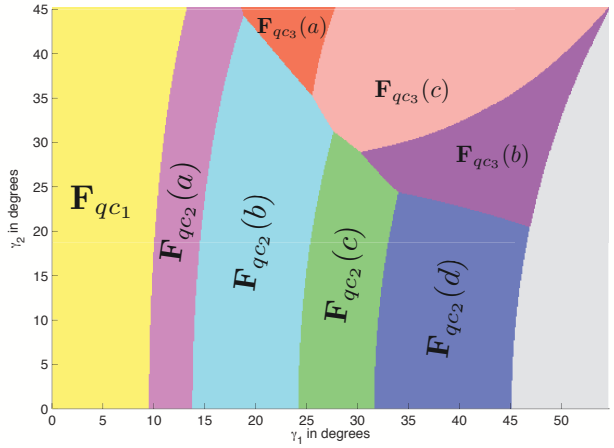


FIGURE 2.11: Range of precoder selection for 4-QAM modulation [20].

spreads power only on the first (strongest) subchannel and has only one precoding form. The optimal solution of $\mathbf{F}_{qc1}$ reads $(\theta_1 = \arctan\frac{\sqrt{5}(\sqrt{3}+1)}{\sqrt{2}}, \theta_2 = \arctan\frac{1}{2}, \theta_3 = 0)$, and $(\varphi_1 = 0, \varphi_2 = \varphi_3 = 15^o)$. The $\mathbf{F}_{qc2}$ precoder spreads power on the two first (two strongest) subchannels and has four different precoding forms with the defined parameters of (θ_i, φ_i) given in TABLE 2.1. The $\mathbf{F}_{qc3}$ precoder spreads power on all the three subchannels and has three different precoding forms with the defined parameters of (θ_i, φ_i) given in TABLE 2.1. The power allocation for each of the three precoders is given in [20, 63]. The selection among these precoders depends on the channel, characterized by (γ_1, γ_2). Thus, based on γ_1 and γ_2, the range of precoder definition for 4-QAM modulation in the case $b = 3$ is presented in FIGURE 2.11.

Generalization of max-$d_{\min}$ precoder

The main difficulty of the precoder design using the minimum Euclidean distance maximization criterion is twofold. Firstly, the space of solutions is large and its dimension is exponentially proportional to the number of data streams b. Secondly, the exact expression of max-$d_{\min}$ precoder depends on many parameters. Therefore, the authors in [21]

proposed a generalized max-$d_{\min}$ precoder. This precoder is a suboptimal solution that can come close to the desired goal, which is the $d_{\min}$ maximization. The principle of this precoder is as follows. Firstly, the authors in [21] proved that, by selecting $\mathbf{\Theta}$ to be equal to the discrete Fourier transform (DFT) matrix, $d_{\min}$ can be forced to a higher value. Given the suboptimal DFT matrix $\mathbf{\Theta}$, the power allocation matrix $\mathbf{\Psi}$ is searched so as to maximize $d_{\min}$.

2.3.3 Comparison between precoders

We mostly consider the case $b = 2$ in this thesis, since we mainly focus on precoder design for different iterative receivers. In addition, the case $b = 2$ is suitable for FEC encoded precoding system used with iterative receivers. Because the smaller b, the lower complexity at soft-demapper (see Section 1.2.2). The extensions of precoders proposed in this thesis to the case $b > 2$ will be considered as perspectives. Note that we, however, have a discussion in case $b > 2$ for our precoder design in Chapter 4, which focuses on channel mutual information maximization criterion.

In this subsection, we present the EXIT charts of the water filling, mercury/water filling, GOPT and max-$d_{\min}$ precoders in case $b = 2$. Firstly, we observe that diagonal precoders are less suitable than non-diagonal precoders to be used in concatenation with outer FEC codes. Secondly, we point out that the non-diagonal precoders, whose optimization criterion is to maximize the channel capacity, achieve good error-rate performances in FEC encoded system assuming iterative receivers.

Let us consider the turbo detection receiver, which is introduced in Section 1.2.2 of Chapter 1, to detect the channel output $\mathbf{y}$ from (2.16). We consider the case $b = 2$ data streams and pick a fixed channel with $\gamma = 30^o$ to plot the EXIT charts of the iterative receiver when the presented precoders are used at the transmitter. In FIGURE 2.12, the solid lines represent the EXIT functions of the soft demapper in case different precoders are used at the transmitter (or, in short, the solid lines represent the EXIT functions of different precoders). The dashed line stands for the EXIT function of the soft decoder, which is fixed for all SNRs and channel realizations. The *a priori* and extrinsic mutual information for soft demapper are denoted by I_A^1 and I_E^1. The equivalent notations for soft decoder are I_A^2 and I_E^2. The iterative detection-decoding converges at the crossing

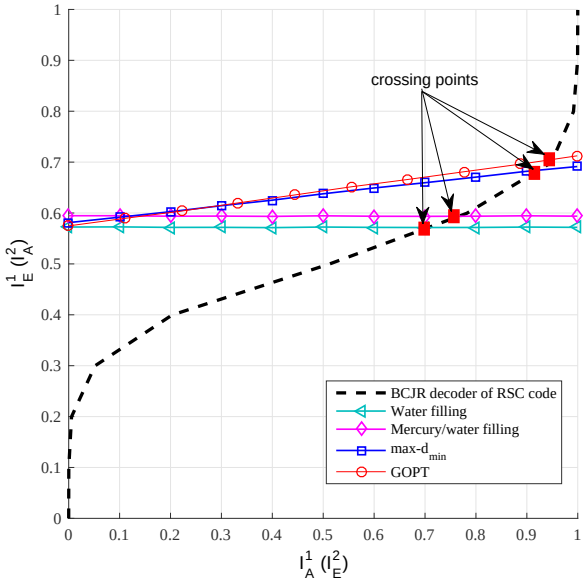


FIGURE 2.12: EXIT chart for turbo detection, SNR = 8 dB, $\gamma = 30^o$ and $(13, 15)_{octal}$ RSC code.

point, where the two EXIT functions intersect each other. The higher the crossing point, the better system error-rate performance.

As illustrated in FIGURE 2.12, the EXIT functions of water-filling and mercury/water-filling precoders remain constant because of their diagonal precoding forms. Therefore, there is no iterative improvement when these precoders are used. Hence, it can be concluded that the use of diagonal precoders does not allow us to exploit the channel diversity attained through the outer FEC code in combination with the coding diversity. In constrast, the non-diagonal precoders enable to take advantage of available diversity to iteratively improve the crossing point, which is equivalent to the convergence point, and then improve the system error-rate performance.

One should note that, (only) in case maximum *a posteriori* soft demapper is used, the area beneath the EXIT function of soft demapper is proportional to the channel mutual information. Hence, by maximizing the channel mutual information, we can force the crossing point to a higher value. For this reason, the GOPT precoder (globally maximizing the channel mutual information) and max-$d_{\min}$ precoder (asymptotically maximizing the lower bound of the channel mutual information) are used regularly as references throughout this work.

2.4 Conclusion

In LTE and LTE-A systems, the channel state information (CSI) is expected to be available at the transmitter through a feedback link. Hence, by assuming full CSI at both transmitter and receiver, linear precoders can be designed to optimize the MIMO system according to various criteria. In this chapter, we firstly recalled the virtual transformation technique. The technique helps to transform any MIMO system into a simpler structure, which transmits the modulated symbol vectors through a desired number of subchannels b. Secondly, we briefly described several selected diagonal and non-diagonal precoders that aim to maximize the channel mutual information. Finally, we relied on the extrinsic information transfer (EXIT) chart to compare the selected precoders. We observed from the EXIT chart that, when used in concatenation with an outer FEC code, the diagonal precoders do not provide a full diversity exploitation as the non-diagonal precoder do. Additionally, we pointed out that maximizing the channel capacity is an essential criterion for MIMO linear precoder design. Therefore, the non-diagonal precoders that maximize (GOPT) or asymptotically maximize (max-$d_{\min}$) the channel mutual information will be considered as references in this thesis. Next chapter investigates the concatenation of FEC codes with MIMO linear precoder assuming turbo detection at the receiver. Two new precoders will be proposed for two different symbol mappings. The first mapping is the conventional Q-ary modulation followed by MIMO conversion. The second mapping takes into account a direct mapping on the received constellation.

Chapter 3

Joint Optimization of MIMO Precoding and Symbol Mapping for Turbo Detection

The content of this chapter is mainly based on the following papers:

- Nhat-Quang Nhan, Philippe Rostaing, Karine Amis, Ludovic Collin, and Emanuel Radoi. "Joint Optimization of MIMO Precoding and Symbol Mapping for Turbo Detection". Submitted to *IEEE Transaction on Wireless Communications*, pending for review.
- Nhat-Quang Nhan, Philippe Rostaing, Karine Amis, Ludovic Collin, and Emanuel Radoi. "Optimized MIMO symbol mapping to improve the turbo cliff region of iterative precoded MIMO detection". In *the 2015 European Signal Processing Conference (EUSIPCO)*, pages 909 - 913, 2015.
- Nhat-Quang Nhan, Philippe Rostaing, Karine Amis, Ludovic Collin, and Emanuel Radoi. "Optimized maxdmin precoder assuming maximum squared Euclidean weight-mapping and turbo detection". To appear in *the International Symposium on Turbo Codes and Iterative Receivers (ISTC)*, 2016.

3.1 Introduction

In this chapter, we consider the concatenation of the MIMO precoder with a binary convolutional code. The turbo detection, introduced in Section 1.2.2, is applied at the receiver. We assume perfect CSI at both the transmitter and the receiver. Our main contributions are threefold. First, we propose a new precoder, which is referred to as $\mathbf{F}_{\ell_1}$ precoder. This precoder is particularly designed for the turbo detection with the usual mapping scheme consisting of binary to Q-ary symbol conversion followed by Q-ary symbol to MIMO symbol conversion (referred to as Gray-M mapping). The parameters of the proposed $\mathbf{F}_{\ell_1}$ precoder are fixed for all channels and, therefore, it simplifies the complexity for practical design. Second, we introduce a MIMO symbol mapper to replace the usual Gray-M mapping. This MIMO symbol mapper can be considered as a direct mapping onto the received constellation. We then demonstrate that the robustness of the linear precoder is improved by applying the MSEW mapping strategy at the mapper. As for the third contribution, we rely on the EXIT chart analysis to propose another new precoder (referred to as *EXIT-based* precoder), which is adapted to be used with the MSEW mapping at the MIMO symbol mapper. The max-$d_{\min}$ precoder, introduced in Section 2.3.2.2, is considered for comparison. Numerical results show that the novel precoders significantly outperform the max-$d_{\min}$ precoder.

The remainder of this chapter is organized as follows. Section 3.2 introduces the system model along with a definition of signal-to-noise ratio (SNR). In Section 3.3, a new precoder ($\mathbf{F}_{\ell_1}$ precoder) for turbo detection used with the usual Gray-M mapping is proposed. Firstly, the upper bound of the turbo detection assuming a precoder with full CSI is defined. Secondly, theoretical analysis is given to look for the defining parameters of the new precoder. Finally, numerical results are presented to validate the analysis and demonstrate the advantages of the new precoder in terms of error rate performance. In Section 3.4, a MIMO symbol mapper is investigated to replace the usual Gray-M mapping. The MSEW mapping technique is considered. Thanks to EXIT chart analysis, another precoder (*EXIT-based* precoder), which is adapted to be used with the MSEW mapping, is proposed. Error-rate comparisons for the association of the precoders and MSEW mapping are also presented at the end of the section to validate the analyses. Section 3.5 concludes the chapter and gives some perspectives.

3.2 Preliminaries

3.2.1 System Model

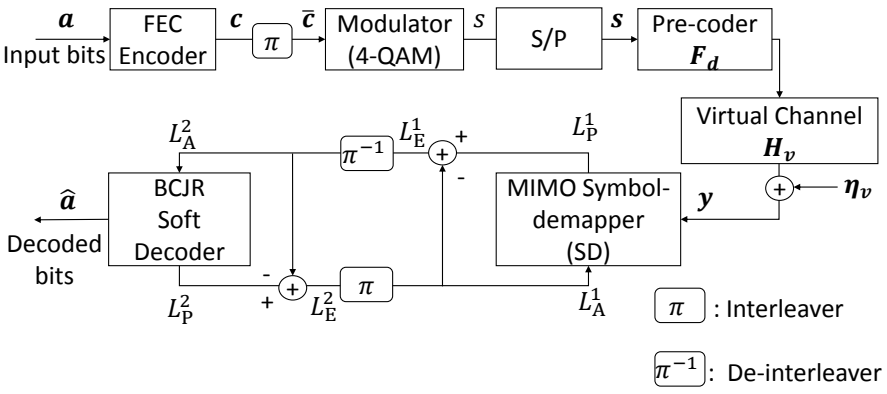


FIGURE 3.1: Equivalent system model.

Let us consider a MIMO system with n_R receive, n_T transmit antennas and b independent data streams to be transmitted. We assume full-CSI at both the transmitter and the receiver. A binary recursive-systematic convolutional (RSC) code is used at the outer FEC encoder to encode information data bits. The FEC codeword is then interleaved before entering a modulator. In the modulator, the interleaved FEC-encoded binary sequence $\bar{\mathbf{c}}$ is grouped and mapped onto a sequence of Q-ary quadrature amplitude modulation (QAM) symbol s, which is then converted into b parallel streams, *i.e.* every b symbols of s are grouped and transposed to a MIMO symbol $\mathbf{s}$ of size $b \times 1$. The vector $\mathbf{s}$ is then precoded with a matrix $\mathbf{F}$ and transmitted through the MIMO channel. At the receiver side, the turbo detection, which has been introduced in Section 1.2.2, is investigated. According to (2.16), the channel output $\mathbf{y}$ after the channel transformation reads $\mathbf{y} = \mathbf{H}_v\mathbf{F}_d\mathbf{s} + \boldsymbol{\eta}$.

We consider the case $b = 2$, which is widely used in the fourth-generation (4G) cellular networks. We would like to point out that the transformation in Section 1.2.2 requires $b \leq \text{rank}(\mathbf{H}) \leq \min(n_T, n_R)$, so n_T and n_R can be larger than b. Therefore, though the proposed precoder is derived by considering $b = 2$, its applications are not limited to 2×2 MIMO systems. We remind the singular value decomposition of $\mathbf{F}_d$ expressed as

$$\mathbf{F}_d = \mathbf{U}_{F_d}\mathbf{\Sigma}_{F_d}\mathbf{V}_{F_d}^{\dagger}.$$

Let us rewrite the diagonal matrix $\mathbf{\Sigma}_{F_d}$ in a polar form as follows

$$\mathbf{\Sigma}_{F_d} = \begin{pmatrix} \cos\psi & 0 \\ 0 & \sin\psi \end{pmatrix}.$$

This polar form satisfies the power constraint $\|\mathbf{F}_d\|_F^2 = 1$. With a given ψ, it is proved that the best selection of $\mathbf{U}_{F_d}$ to maximize the singular values of $\mathbf{H}_v\mathbf{F}_d$ is $\mathbf{U}_{F_d} = \mathbf{I}_b$ [16]. It is known that the remaining 2×2 unitary matrix $\mathbf{V}_{F_d}^{\dagger}$ can also be written as

$$\mathbf{V}_{F_d}^{\dagger} = \mathbf{D}\begin{pmatrix} \cos\theta & \sin\theta e^{i\phi} \\ -\sin\theta & \cos\theta e^{i\phi} \end{pmatrix},$$

where $\mathbf{D}$ is a diagonal unitary matrix [22]. Without loss of generality, let us select $\mathbf{D} = \mathbf{I}_2$. Hence, a parameterized definition of $\mathbf{F}_d$ can be defined as

$$\mathbf{F}_d = \begin{pmatrix} \cos\psi & 0 \\ 0 & \sin\psi \end{pmatrix}\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & e^{i\phi} \end{pmatrix}, \tag{3.1}$$

where ψ ($0^o \leq \psi \leq 90^o$) is linked to the power allocation on the eigen-subchannels, θ ($0^o \leq \theta \leq 90^o$) and ϕ ($0^o \leq \phi \leq 90^o$) allow us to respectively mix and rotate the symbols onto the two eigen-subchannels.

The equivalent scheme is shown in FIGURE 3.1, where L_A^1, L_P^1 and L_E^1 respectively stand for the *a priori*, the *a posteriori* and the extrinsic log likelihood ratios (LLRs) of the soft demapper, while the equivalent notations for the BCJR soft decoder are L_A^2, L_P^2 and L_E^2.

3.2.2 SNR definition

With $b = 2$, the conversion from cartesian to polar form of $\mathbf{H}_v$ gives

$$\mathbf{H}_v = \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} = \rho\begin{pmatrix} \cos\gamma & 0 \\ 0 & \sin\gamma \end{pmatrix}, \tag{3.2}$$

where ρ and γ respectively represent the channel gain and angle. As $\sigma_1 \geq \sigma_2 > 0$, $0 < \gamma \leq 45^o$. Therefore, any random MIMO channel can be simply characterized by the pair (ρ, γ) thanks to the virtual transformation. Let us define the instantaneous received

SNR as

$$\text{SNR} = \frac{\sigma_s^2}{\sigma_\eta^2}\|\mathbf{H}\|_F^2 = \frac{\sigma_s^2}{\sigma_\eta^2}\rho^2. \tag{3.3}$$

This SNR definition will be considered in all chapters of the thesis. With this definition of SNR, a channel is only characterized by its angle γ. For instance, the channel $\mathbf{H} = [2\ 1; 1\ 1]$, which was also used in [22, 64], has $\gamma = \arctan\frac{\sigma_2}{\sigma_1} \simeq 8.3^o$. In the remainder of this chapter, we classify channels by value of γ. The channel $\mathbf{H} = [2\ 1; 1\ 1]$ is referred to as Channel A. We also denote by Channel B another channel with $\gamma = 30^o$.

3.3 Optimized precoder for the conventional mapping

In this section, we consider the conventional Gray-M mapping as illustrated in FIGURE 3.1 (*i.e.* the encoded codeword $\bar{\mathbf{c}}$ is modulated by a modulator before being converted to MIMO symbol vectors $\mathbf{s}$ by using a serial-to-parallel converter). Let us denote by ℓ_1 the minimum squared Euclidean distance between any pair of MIMO symbol vectors with associated binary conversions differing by only one bit. The goal of this section is to propose a novel precoder that maximize ℓ_1 in the case of the conventional Gray-M mapping. We firstly show that maximizing ℓ_1 results in a low asymptotic BER. Consequently, we propose a new precoding solution (referred to as $\mathbf{F}_{\ell_1}$) that maximizes ℓ_1 for the case $b = 2$. Simulation is then carried out to validate the analysis.

3.3.1 Upper bound of Turbo detection assuming precoder with perfect CSI

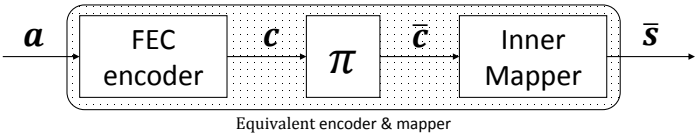


FIGURE 3.2: Equivalent encoder and mapper block.

Let us consider the concatenation of the modulator and the serial-to-parallel converter in FIGURE 3.1 as a binary to MIMO symbol mapper (inner mapper). The chain of the outer FEC encoder, the interleaver and this inner mapper can be considered as an equivalent encoder and mapper block, which maps every Hamming weight w input binary sequence $\mathbf{a}$ of length W into a sequence of MIMO symbols $\bar{\mathbf{s}}$ of size $b \times U$ as shown in FIGURE 3.2.

Denoting by $\mathbf{B} = \mathbf{H}_v\mathbf{F}_d$ the new channel matrix of size $b \times b$. The received vector sequence, denoted by $\bar{\mathbf{y}}$, reads

$$\bar{\mathbf{y}} = \underbrace{\mathbf{H}_v\mathbf{F}_d}_{\mathbf{B}}\bar{\mathbf{s}} + \bar{\boldsymbol{\eta}}, \tag{3.4}$$

where $\bar{\boldsymbol{\eta}}$ is the additional Gaussian noise vector sequence of size $b \times U$. Elements of $\bar{\boldsymbol{\eta}}$ are independent and identically distributed as $\eta_{i,j} \sim \mathcal{CN}(0, \sigma_\eta^2)$.

In the case of maximum likelihood detection, assuming that $\mathbf{B}$ is known at the transmitter, the pairwise error probability of the sequences $\bar{\mathbf{s}}$ and $\bar{\mathbf{s}}'$ is then given by

$$P\left(\bar{\mathbf{s}}' \to \bar{\mathbf{s}} \mid \mathbf{B}, \bar{\mathbf{s}}'\right) \leq P\left(\|\bar{\mathbf{y}} - \mathbf{B}\bar{\mathbf{s}}\|_F^2 \leq \|\bar{\mathbf{y}} - \mathbf{B}\bar{\mathbf{s}}'\|_F^2\right). \tag{3.5}$$

From (3.4), we can deduce that given $\mathbf{B}$ and $\bar{\mathbf{s}}'$,

$$\|\bar{\mathbf{y}} - \mathbf{B}\bar{\mathbf{s}}\|_F^2 - \|\bar{\mathbf{y}} - \mathbf{B}\bar{\mathbf{s}}'\|_F^2 = \|\mathbf{B}(\bar{\mathbf{s}}' - \bar{\mathbf{s}})\|_F^2 + \underbrace{2\Re\left\{\mathrm{Tr}\left\{\left(\mathbf{B}(\bar{\mathbf{s}}' - \bar{\mathbf{s}})\right)^\dagger \boldsymbol{\eta}\right\}\right\}}_{\xi}, \tag{3.6}$$

where $\sigma_\xi^2 = 2\sigma_\eta^2\|\mathbf{B}(\bar{\mathbf{s}}' - \bar{\mathbf{s}})\|_F^2$ (see Appendix A).

Substitution of (3.6) into (3.5) yields

$$\begin{aligned} P\left(\bar{\mathbf{s}}' \to \bar{\mathbf{s}} \mid \mathbf{B}, \bar{\mathbf{s}}'\right) &\leq P\left(\xi \leq -\|\mathbf{B}(\bar{\mathbf{s}}' - \bar{\mathbf{s}})\|_F^2\right), \\ &\leq \frac{1}{2}\exp\left(-\frac{\|\mathbf{B}(\bar{\mathbf{s}}' - \bar{\mathbf{s}})\|_F^2}{4\sigma_\eta^2}\right). \end{aligned} \tag{3.7}$$

Let us denote by $\varpi(\bar{\mathbf{s}}', \bar{\mathbf{s}}) = \|\mathbf{B}(\bar{\mathbf{s}}' - \bar{\mathbf{s}})\|_F^2$ the received squared Euclidean distance between $\bar{\mathbf{s}}$ and $\bar{\mathbf{s}}'$, and by $\bar{\mathbf{s}}_0$ the sequence mapped by the all-zero codeword. Let us denote by $\mathcal{S}$ the set of all possible $\bar{\mathbf{s}}$ and by $\mathcal{S}_l = \{\bar{\mathbf{s}} \in \mathcal{S} | \varpi(\bar{\mathbf{s}}_0, \bar{\mathbf{s}}) = l\}$ the set of the sequences $\bar{\mathbf{s}}$ that have the received squared Euclidean weights (SEWs) equal to l (or SEW-l, for brevity). The probability of fault detection for any $\bar{\mathbf{s}} \in \mathcal{S}$ (word-error probability) reads

$$\begin{aligned} P_e = P_e\left(\bar{\mathbf{s}}_o\right) &\leq \sum_{\bar{\mathbf{s}} \in \mathcal{S},\ \bar{\mathbf{s}} \neq \bar{\mathbf{s}}_o} P\left(\bar{\mathbf{s}}_o \to \bar{\mathbf{s}} \mid \mathbf{B}, \bar{\mathbf{s}}_o\right), \\ &\leq \sum_l A_l P\left(\bar{\mathbf{s}}_o \to \bar{\mathbf{s}}_l \mid \mathbf{B}, \bar{\mathbf{s}}_o\right), \end{aligned} \tag{3.8}$$

where the equality $(P_e = P_e(\bar{\mathbf{s}}_o))$ is obtained due to the linearity of the outer FEC code [65], $\bar{\mathbf{s}}_l \in \mathcal{S}_l$ and A_l is the cardinality of $\mathcal{S}_l$. We further denote by $A_{w,l}$ the number

of $\bar{\mathbf{s}}$ that has input Hamming weight-w and output SEW-l. Hence, A_l can be calculated in function of $A_{w,l}$ as

$$A_l = \sum_w A_{w,l}. \tag{3.9}$$

On the other hand, when $\bar{\mathbf{s}}_o$ is wrongly decoded into $\bar{\mathbf{s}}_l$, there are w binary errors among K components. Therefore, from (3.7), (3.8) and (3.9), the bit-error probability of the input binary message is upper bounded by

$$P_b \leq \sum_w \sum_l \frac{w A_{w,l}}{2K} \exp\left(-\frac{l}{4\sigma_\eta^2}\right). \tag{3.10}$$

The encoded binary sequence $\mathbf{c}$ has length N and Hamming weight-κ (so does its interleaved sequence $\bar{\mathbf{c}}$). We can deduce that, assuming a uniform interleaver, a Hamming weight-κ outer codeword $\mathbf{c}$ can be interleaved into $\binom{N}{\kappa}$ possible $\bar{\mathbf{c}}$. Therefore, $A_{w,l}$ can be calculated as follows

$$A_{w,l} = \sum_{\kappa=d_f}^{N} \frac{A_{w,\kappa}^{(\text{out})} A_{\kappa,l}^{(\text{in})}}{\binom{N}{\kappa}}, \tag{3.11}$$

where $A_{w,\kappa}^{(\text{out})}$ is the number of FEC codewords with weight-κ generated by data words with weight-w, $A_{\kappa,l}^{(\text{in})}$ is the number of modulated sequences with SEW-l generated by FEC codewords with weight-κ and d_f is the free distance of convolutional codes.

The conditional weight enumerating function (CWEF) of the inner mapper can be defined by

$$A^{(\text{in})}(\kappa, L) = \sum_l A_{\kappa,l}^{(\text{in})} L(l) \,|_{L(l)=\frac{1}{2}\exp\left(-\frac{l}{4\sigma_\eta^2}\right)}. \tag{3.12}$$

From (3.10), (3.11) and (3.12), we deduce that

$$P_b \leq \sum_w \frac{w}{W} \sum_{\kappa=d_f}^{N} \frac{A_{w,\kappa}^{(\text{out})}}{\binom{N}{\kappa}} A^{(\text{in})}(\kappa, L). \tag{3.13}$$

3.3.2 Theoretical analysis

The sequence MIMO symbol mapper (which maps the whole $\bar{\mathbf{c}}$ into $\bar{\mathbf{s}}$) is equal to applying N/m times the MIMO symbol mapper, which instantaneously maps every m binary bits of $\bar{c}$ into a MIMO symbol $\mathbf{s}$ of size $b \times 1$, *i.e.* $\bar{\mathbf{s}} = [\mathbf{s_1}\ \mathbf{s_2}\ \ldots\ \mathbf{s}_{N/m}]$. At the outer

FEC decoder, the most probable error event is the error sequence with minimum Hamming weight, *i.e.* $\kappa = d_f$. Assuming a uniform interleaver for the minimum weight error sequence, there is a very small chance that the d_f error bits, after they are interleaved, be mapped to the same symbol. Hence, with $N \gg d_f$, we assume a perfect interleaver such that the d_f bits are scattered throughout the sequence and grouped into d_f different Hamming weight-1 blocks before being mapped to d_f different symbol vectors. Therefore, the SEW of the Hamming weight-1 symbols is critical to the asymptotic bit-error-rate (BER) performance. In other words, a low asymptotic BER can be achieved by maximizing the corresponding SEW of Hamming weight-1 symbol vectors.

Indeed, with a given outer FEC code, we can see from (3.13) that the smaller $A^{(\text{in})}(\kappa, L)$, the lower the asymptotic BER. For a codeword with Hamming weight d_f, *i.e.* $\kappa = d_f$, the CWEF $A^{(\text{in})}(\kappa, L)$ becomes $A^{(\text{in})}(d_f, L) = [T(1, L)]^{d_f}$ [66], where $T(1, L)$ stands for the CWEF among the Hamming weight-1 symbols of the MIMO symbol mapper. Therefore, in order to achieve a lower BER, $T(1, L)$ must be minimized. The general expression of $T(1, L)$, which is averaged over all possible reference points on the constellation, reads

$$T(1, L) = \sum_i \alpha_i L(\ell_i) \,|_{L(\ell)=\frac{1}{2}\exp\left(-\frac{\ell}{4\sigma_\eta^2}\right)}, \tag{3.14}$$

where $\ell_1 \leq \cdots \leq \ell_i \leq \cdots \leq \ell_{\max}$. We can see that $L(\ell_i)$ is exponentially increased with the decrease of ℓ_i. Therefore, to minimize $T(1, H)$, the maximization of ℓ_1 is crucial.

In this subsection, we propose a novel precoder, which maximizes ℓ_1 for the conventional Gray-mapped scheme. Indeed, ℓ_1 can be maximized by maximizing the squared Euclidean distance ℓ between any pair of the received constellation points located at Hamming distance equal to 1 from each other. Note that the received constellation is fixed thanks to the precoder. Thus, the distance ℓ is defined by

$$\ell = \|\mathbf{H}_v \mathbf{F}_d(\check{\mathbf{s}} - \check{\mathbf{s}}')\|^2, \tag{3.15}$$

where $\check{\mathbf{s}}$ and $\check{\mathbf{s}}'$ represent any pair of symbol vectors with associated binary conversions differing by only one bit. Let us denote by $\check{\boldsymbol{\nu}} = [\check{\nu}_1 \quad \check{\nu}_2]^T$ the difference vector between

$\check{\mathbf{s}}$ and $\check{\mathbf{s}}'$. From (3.1) and (3.2), calculations yield

$$\|\mathbf{H}_v\mathbf{F}_d\check{\boldsymbol{\nu}}\|^2 = |\check{\nu}_1|^2 \underbrace{\left(\sigma_1^2\cos^2\psi\cos^2\theta + \sigma_2^2\sin^2\psi\sin^2\theta\right)}_{D_1} + |\check{\nu}_2|^2 \underbrace{\left(\sigma_1^2\cos^2\psi\sin^2\theta + \sigma_2^2\sin^2\psi\cos^2\theta\right)}_{D_2}$$
$$+ \underbrace{\sin\theta\cos\theta(\check{\nu}_1{}^*\check{\nu}_2 e^{i\phi} + \check{\nu}_1\check{\nu}_2{}^* e^{-i\phi})(\sigma_1^2\cos^2\psi - \sigma_2^2\sin^2\psi)}_{D_3}. \tag{3.16}$$

Lemma 1. In case $b = 2$, with the conventional Gray-M mapping (bit-interleaved codewords are mapped at the modulator before being converted to MIMO symbols) and $\check{\boldsymbol{\nu}} = [\check{\nu}_1 \quad \check{\nu}_2]^T$ is the difference vector between any pair of symbol vectors $\check{\mathbf{s}}$ and $\check{\mathbf{s}}'$, whose binary mapped sequences differ exactly one bit, we have $\check{\nu}_1\check{\nu}_2 = 0$.

Proof. With $b = 2$, let us rewrite $\check{\mathbf{s}} = \{\check{s}(1) \quad \check{s}(2)\}$ and $\check{\mathbf{s}}' = \{\check{s}'(1) \quad \check{s}'(2)\}$. Since the binary mapped sequences of $\check{\mathbf{s}}$ and $\check{\mathbf{s}}'$ differ by only one bit and since the elements of $\check{\mathbf{s}}$ and $\check{\mathbf{s}}'$ are modulated before being serial-to-parallel converted to $\check{\mathbf{s}}$ and $\check{\mathbf{s}}'$, $\check{s}(1) \neq \check{s}'(1)$ and $\check{s}(2) \neq \check{s}'(2)$ can not be satisfied at the same time. Note that this property holds for any modulation type applied at the modulator (before the MIMO symbols conversion). Therefore, with $\check{\nu}_1 = \check{s}(1) - \check{s}'(1)$ and $\check{\nu}_2 = \check{s}(2) - \check{s}'(2)$, we deduce that $\check{\nu}_1\check{\nu}_2 = 0$ (*i.e.* at least $\check{\nu}_1$ or $\check{\nu}_2$ equals zero for any pair of symbol vectors $\check{\mathbf{s}}$ and $\check{\mathbf{s}}'$, whose associated binary mappings differ by exactly one bit). For example, let us denote by s_0, s_1, s_2, s_3 the 4-QAM constellation symbols, whose binary mapped sequences are $00, 01, 10, 11$ respectively. We arbitrarily pick up two 4-bits binary sequences 0101 and 0111 differing by only one bit. After applying the 4-QAM symbol mapping followed by the serial-to-parallel conversion, the corresponding symbol vectors are $\check{\mathbf{s}}_1 = \{s_1 \quad s_1\}$ and $\check{\mathbf{s}}_2 = \{s_1 \quad s_3\}$ respectively. Hence, $\check{\nu}_1 = 0$ and $\check{\nu}_2 = s_1 - s_3$ in this case, *i.e.* $\check{\nu}_1\check{\nu}_2 = 0$. For any other selection of the two 4-bits binary sequences differing by only one bit, it always comes that at least $\check{\nu}_1$ or $\check{\nu}_2$ is equal to zero. □

Taking into account Lemma 1, we deduce that $D_3 = 0$. From (3.16), we deduce that in order to maximize $\|\mathbf{H}_v\mathbf{F}_d\check{\boldsymbol{\nu}}\|^2$ independently with $|\check{\nu}_1|^2$ and $|\check{\nu}_2|^2$, we need to jointly maximize D_1 and D_2, which depend only on ψ and θ (the first two parameters of the precoding matrix $\mathbf{F}_d$ as shown in (3.1)). The joint maximization of (D_1, D_2) is considered as a multi-objective optimization without any special expectation for the solutions. Therefore, a non-preference method [67] is applied. The problem of finding (ψ, θ) that

jointly maximizes (D_1, D_2) becomes

$$(\psi^\star, \theta^\star) = \arg\min_{\mathcal{F}} \left(\mho(\psi,\theta) = (D_1 - D_1^{\max})^2 + (D_2 - D_2^{\max})^2\right), \tag{3.17}$$

where $D_k^{\max}$ is the maximum value of D_k over the set $\mathcal{F}$, defined by $0^o \leq \psi \leq 90^o$ and $0^o \leq \theta \leq 90^o$.

Lemma 2. The maximum values of D_1 and D_2 over the set $\mathcal{F}$ are $D_1^{\max} = D_2^{\max} = \sigma_1^2$.

Proof. See Appendix B □

The optimization problem (3.17) can be solved by searching for (ψ, θ) that satisfy the first and second order conditions for a local minimum. Taking into account Lemma 2, the partial derivatives of $\mho$ with respect to θ and ψ yield

$$\begin{aligned}
\frac{\partial \mho}{\partial \theta}(\psi,\theta) &= -\sin(4\theta)\left(\sigma_1^2\cos^2\psi - \sigma_2^2\sin^2\psi\right)^2, \\
\frac{\partial \mho}{\partial \psi}(\psi,\theta) &= \frac{\sigma_s^4}{\sigma_\eta^4}\sin(2\psi)\left[2(\sigma_1^4+\sigma_2^4)\sin^2\psi - 2\sigma_1^2\sigma_2^2 + (\sigma_1^2+\sigma_2^2)\sin^2(2\theta)\left(\sigma_1^2 - (\sigma_1^2+\sigma_2^2)\sin^2\psi\right)\right].
\end{aligned} \tag{3.18}$$

The set of points (ψ, θ) of $\mathcal{F}$ that satisfy the first order conditions for a local minimum ($\nabla\mho = \mathbf{0}$), denoted by $\mathcal{F}^\star$, equals

$$\begin{aligned}
\mathcal{F}^\star = \Bigg\{ &(0,0), (0,45^o), (0,90^o), (90^o,0), (90^o,45^o), (90^o,90^o), \left(\arcsin\frac{\sigma_1\sigma_2}{\sqrt{(\sigma_1^4+\sigma_2^4)}}, 0\right), \\
&\left(\arcsin\frac{\sigma_1\sigma_2}{\sqrt{(\sigma_1^4+\sigma_2^4)}}, 90^o\right)\Bigg\}.
\end{aligned} \tag{3.19}$$

The only point of $\mathcal{F}^\star$ that satisfies the second order conditions for a local minimum (the Hessian matrix $\nabla^2\mho$ is positive-definite) is $(\psi, \theta) = (0, 45^o)$, where $\frac{\partial^2\mho}{\partial\psi^2}(0,45^o) = \sigma_1^2(\sigma_1^2 - \sigma_2^2)$, $\frac{\partial^2\mho}{\partial\theta^2}(0,45^o) = 4\sigma_1^4$ and $\frac{\partial^2\mho}{\partial\theta\partial\psi}(0,45^o) = 0$. As $\mho$ is convex, the minimum is global.

ℓ_1 is maximized by taking $(\psi = 0, \theta = 45^o)$. However, the detection performance can be further improved by forcing the minimum Euclidean distances between all symbols on the received constellation, which is denoted by $d_{\min}$, to a higher value. This helps to improve the other ℓ_i, with $i > 1$. Therefore, the last parameter ϕ will be selected so as to maximize $d_{\min}$ while assuming $(\psi = 0, \theta = 45^o)$. Let us denote by $\boldsymbol{\nu} = [\nu_1 \quad \nu_2]^T$ the difference vector between any pair of symbol vectors $\mathbf{s}$ and $\tilde{\mathbf{s}}$ ($\mathbf{s} \neq \tilde{\mathbf{s}}$). With $(\psi = 0, \theta = 45^o)$,

substitution yields $\|\mathbf{H}_v\mathbf{F}_d\boldsymbol{\nu}\|^2 = \frac{\sigma_1^2}{2}|\nu_1 + \nu_2 e^{i\phi}|^2$. Hence, the optimum value of ϕ, referred to as ϕ_{opt}, can be found by

$$\phi_{opt} = \underset{0^o\leq\phi\leq 90^o}{\arg\max}\left(\min_{\boldsymbol{\nu}} \underbrace{\frac{\sigma_1\sqrt{2}}{2}\left|\nu_1 + \nu_2 e^{i\phi}\right|}_{d(\phi)}\right). \tag{3.20}$$

Lemma 3. For any symmetric QAM modulation, the searching interval of ϕ can be restricted from $0^o \leq \phi \leq 90^o$ to $0^o \leq \phi \leq 45^o$.

Proof. See Appendix C. □

Up to this step, we resort to numerical optimization to look for ϕ. We consider 4-QAM modulation and the searching range is limited to $0^o \leq \phi \leq 45^o$ thanks to Lemma 3. Numerical search over all possible symbol vectors shows that $d_{\min}$ can be obtained by considering only the two following pairs of symbol vectors: $\left\{\left(\frac{1-i}{\sqrt{2}} \quad \frac{-1-i}{\sqrt{2}}\right)^T;\left(\frac{-1-i}{\sqrt{2}} \quad \frac{1-i}{\sqrt{2}}\right)^T\right\}$ and $\left\{\left(\frac{-1+i}{\sqrt{2}} \quad \frac{-1-i}{\sqrt{2}}\right)^T;\left(\frac{1-i}{\sqrt{2}} \quad \frac{-1+i}{\sqrt{2}}\right)^T\right\}$. The former couple has difference vector $\boldsymbol{\nu} = \left(\sqrt{2} \quad -\sqrt{2}\right)^T$, which yields $d(\phi) = \sigma_1\sqrt{2(1-\cos\phi)}$ referred to as d_1. The latter couple has difference vector $\boldsymbol{\nu} = \left(-\sqrt{2}+i\sqrt{2} \quad -i\sqrt{2}\right)^T$, which yields $d(\phi) = \sigma_1\sqrt{3-2(\cos\phi+\sin\phi)}$ referred to as d_2. Hence, $d_{\min}$ is equal to d_1 for $0 \leq \phi \leq 30^o$ and to d_2 for $30^o \leq \phi \leq 45^o$. The optimum value of ϕ, which maximizes $d_{\min}$, is obtained at the intersection between d_1 (increasing function of ϕ) and d_2 (decreasing function of ϕ), which yields $\phi_{opt} = 30^o$. Therefore, we propose to take $\phi = 30^o$. The proposed precoder with optimized defining parameters $(\psi = 0^o, \theta = 45^o, \phi = 30^o)$, is referred to as $\mathbf{F}_{\ell_1}$ precoder. It should also be noted that the parameters of this precoder are fixed, which makes its design and practical application easier.

3.3.3 Simulation results

We now provide examples to demonstrate the advantages of the proposed $\mathbf{F}_{\ell_1}$ precoder in terms of error-rate performance. Monte-Carlo simulations have been performed for a $n_T = 2$ transmit and $n_R = 2$ receive antennas MIMO configuration. The half-rate $(13, 15)_{octal}$-RSC code is used at the outer encoder. The frame length is 800 uncoded bits. The modulator uses Gray mapping rule (Gray-M) for the 4-QAM symbols before their conversion into symbol vectors.

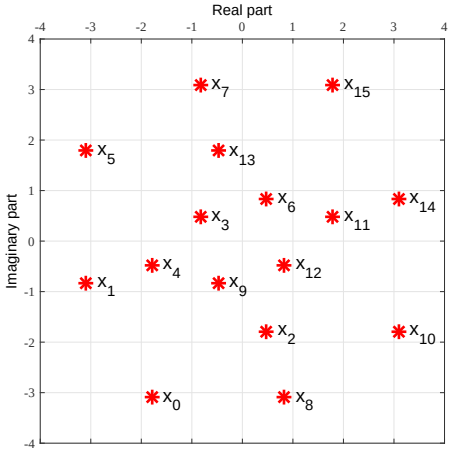


FIGURE 3.3: The received constellation on first subchannel of $\mathbf{F}_{\ell_1}$ precoder, Channel A, Gray-M mapping and 4-QAM modulation.

In the case of two data streams transmission ($b = 2$) and 4-QAM modulation, we compare the error-rate performance of $\mathbf{F}_{\ell_1}$ precoder with the max-$d_{\min}$ precoder ($\mathbf{F}_{r1}$ or $\mathbf{F}_{octa}$), which shows a better uncoded error-rate performance with maximum likelihood detection than the other precoders such as MMSE, water-filling, max-SNR and minimum BER, as mentioned in Section 2.3.2.2. The received constellations on the first and the second subchannels of $\mathbf{F}_{octa}$ are plotted in FIGURE 2.9(a) and FIGURE 2.9(b) respectively, while the received constellation on the first subchannel of $\mathbf{F}_{r1}$ is given in FIGURE 2.8. The received constellation of the $\mathbf{F}_{\ell_1}$ precoder is given in FIGURE 3.3.

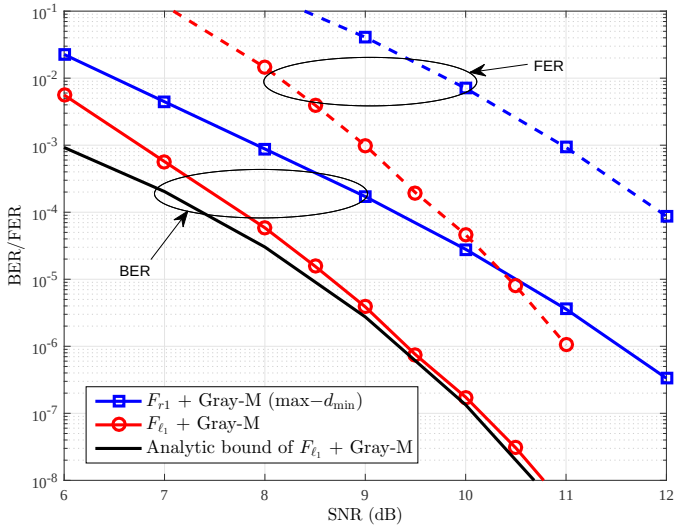


FIGURE 3.4: BER (solid lines) and FER (dashed lines) of Channel A, Gray-M mapping, (13,15)-RSC code and 4-QAM modulation.

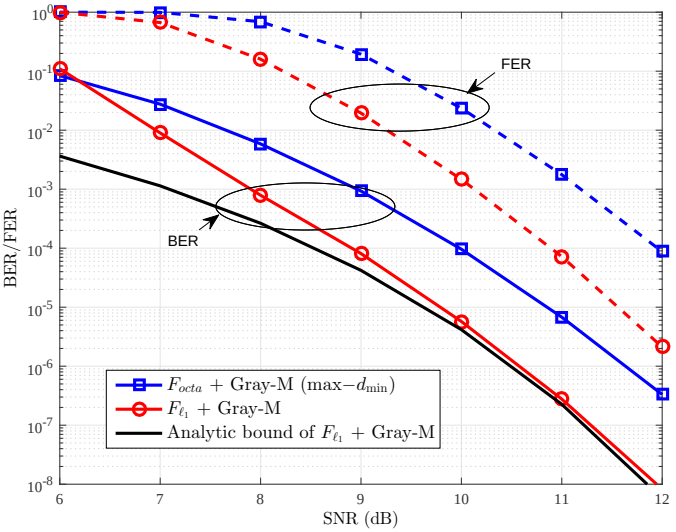


FIGURE 3.5: BER (solid lines) and FER (dashed lines) of Channel B, Gray-M mapping, (13,15)-RSC code and 4-QAM modulation.

The Channel A ($\gamma \simeq 8.3^o$) is considered to illustrate the case $\gamma \leq \gamma_o$. FIGURE 3.4 shows the BER performance of the $\mathbf{F}_{\ell_1}$ and max-$d_{\min}$ precoders on this channel. Note that the max-$d_{\min}$ precoder uses $\mathbf{F}_{r1}$ mode in this case ($\gamma \leq \gamma_o$). We observe that the $\mathbf{F}_{\ell_1}$ precoder achieves a gain of 2 dB at BER $= 10^{-6}$ and roughly 2.5 dB at BER $= 10^{-7}$ compared to the max-$d_{\min}$ precoder. In terms of frame-error-rate (FER), the gain of $\mathbf{F}_{\ell_1}$ compared to max-$d_{\min}$ precoder is also remarkable, namely, 2 dB at FER $= 10^{-3}$ and about 2.25 dB at FER $= 10^{-4}$. As shown in FIGURE 3.4, with the change of slope of the curves, we observe that the gain of $\mathbf{F}_{\ell_1}$ compared to max-$d_{\min}$ is more significant with the increase of SNR.

The Channel B ($\gamma = 30^o$) is also considered to illustrate the case $\gamma > \gamma_o$. The error-rate performances of $\mathbf{F}_{\ell_1}$ and max-$d_{\min}$ precoders on this channel are shown in FIGURE 3.5. Note that the max-$d_{\min}$ precoder uses $\mathbf{F}_{octa}$ mode in this case ($\gamma > \gamma_o$). We observe that the proposed $\mathbf{F}_{\ell_1}$ precoder achieves a gain of 1 dB at BER $= 10^{-6}$ compared to max-$d_{\min}$ precoder. In terms of FER, the gain is more than 1 dB at FER $= 10^{-4}$. The change of slope of the curves in this channel is slower than in Channel A. Hence, comparing Channel A and Channel B, it can be concluded that the advantage of $\mathbf{F}_{\ell_1}$ precoder is more significant over the low angle γ channels.

An analytic bound for the error-free feedback performance is also taken into account. The bound expressed by (3.13) is plotted by considering only codewords with free Hamming

distance (d_f). To compute this bound, we assume a perfect interleaver, which allows the inner mapper to map d_f error bits into different symbol vectors. As observed from both FIGURE 3.4 and FIGURE 3.5, the analytic bound of the $\mathbf{F}_{\ell_1}$ precoder matches the simulated curve at high SNR. The max-$d_{\min}$ precoder also reaches its analytic bound, which is not plotted here for brevity but will be shown later in Section 3.4.3.

3.4 Performance enhancement

In this section, we assume that the encoded codeword $\bar{\mathbf{c}}$ is directly mapped onto MIMO symbol vectors $\mathbf{s}$. Since $\mathbf{H}_v$ and $\mathbf{F}_d$ are known, this is equivalent to a direct mapping onto the received constellation, where the received symbol is denoted by $\mathbf{x} = \mathbf{H}_v\mathbf{F}_d\mathbf{s}$. Therefore, we are able to apply the maximum squared Euclidean weight (MSEW) criterion [66] (maximizing ℓ_1 and minimizing α_1 in (3.14)) to look for the optimized mapping at the received constellation. The $\mathbf{F}_{\ell_1}$ precoder proposed for the conventional Gray-M mapping in Section 3.3 is then modified to adapt to the MSEW mapping. The modified precoder is referred to as $\mathbf{F}_{\ell_1}$-mod precoder.

We firstly look for the optimized MSEW mappings for all of the considered precoding forms ($\mathbf{F}_{r1}, \mathbf{F}_{octa}, \mathbf{F}_{\ell_1}$-mod). We observe that, thanks to the MSEW mapping criterion (maximizing ℓ_1 and minimizing α_1), the error-floors achieved by the listed precoders are very low, which are out of the commonly used BER. Therefore, since ℓ_1 has been optimized thanks to the MSEW mapping, the precoding optimization criterion of maximizing ℓ_1 is no more interesting. The goal of this section is thus to propose a precoding strategy that improves the turbo-cliff region by switching among the available MSEW-mapped precoders. With that criterion in mind, we resort to EXIT chart to find the optimized solution. Thanks to EXIT chart analysis, we propose a second novel precoder (referred to as *EXIT-based* precoder), which uses the MSEW-mapped $\mathbf{F}_{\ell_1}$-mod and $\mathbf{F}_{octa}$ forms with the switching threshold $\gamma_1 = 22.5^o$. Simulation coincides with the analysis and is presented at the end of this section.

3.4.1 Direct mapping at the received constellation

From (3.1) and (3.2), it comes that, for $\mathbf{F}_{r1}$, $\mathbf{F}_{octa}$, $\mathbf{F}_{\ell_1}$ and $\mathbf{F}_{\ell_1}$-mod precoders, $\mathbf{B} = \mathbf{H}_v\mathbf{F}_d = \beta\check{\mathbf{B}}$, where $\check{\mathbf{B}}$ is a fixed matrix and β is a scalar, which depends on γ and ρ.

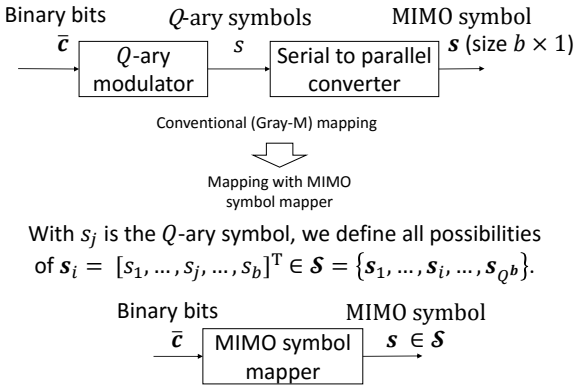


FIGURE 3.6: Conventional mapping versus mapping with MIMO symbol mapper.

Hence, for a given $\mathbf{F}_d$, the received constellation is unchanged and just scaled by a scalar factor β for all channel realizations. On the other hand, though the performance has been improved by using the $\mathbf{F}_{\ell_1}$ precoder, $T(1, H)$ can be further maximized if we can control the mapping at the received constellation. Since the received constellation is fixed, a direct mapping on it is possible. In this section, we propose a direct MIMO symbol mapping at the inner mapper as illustrated in FIGURE 3.6, which maps a block of m binary bits onto a vector of symbol $\mathbf{s}$. This is equivalent to a direct mapping onto the received symbol vector $\mathbf{x} = \mathbf{H}_v\mathbf{F}_d\mathbf{s}$ since $\mathbf{H}_v$ and $\mathbf{F}_d$ are known.

The maximum squared Euclidean weight (MSEW) mapping strategy is considered for the mapping at the received constellation. Pioneered in [66], the purpose of MSEW is to achieve a low error-rate by optimizing two mapping criteria. Firstly, it maximizes the minimum Euclidean distance between symbols with binary mapped sequences differing by one position (*i.e.* ℓ_1 in (3.14)). Secondly, it minimizes the number of pairs of symbols with binary mapped sequences differing by one position separated by the minimum Euclidean distance ℓ_1 (*i.e.* α_1 in (3.14)). With these two criteria, the best MSEW mapping is then obtained by computer search.

Since jointly finding the best mapping and optimizing all defining parameters of precoder is intractable, we propose to fix the first two parameters of the $\mathbf{F}_{\ell_1}$ precoder as found in Section 3.3, *i.e.* $(\psi = 0^o, \theta = 45^o)$, and to look for the parameter ϕ that yields the best MSEW mapping. With $(\psi = 0^o, \theta = 45^o)$, we firstly vary ϕ from 0 to 45^o. Secondly, we search the optimized MSEW mapping for each value of $\phi \in [0, 45^o]$. Finally, we look for the value of ϕ that yields the maximum ℓ_1 and minimum α_1. Computer search shows that the best MSEW mapping is achieved at $\phi = 45^o$. We refer to the new precoder, with the

defining parameters ($\psi = 0^o, \theta = 45^o, \phi = 45^o$), as $\mathbf{F}_{\ell_1}$-mod precoder. The constellation of this precoder is shown in FIGURE 3.7. For comparison, the MSEW mappings for the $\mathbf{F}_{r1}$ and $\mathbf{F}_{octa}$ precoders are also considered.

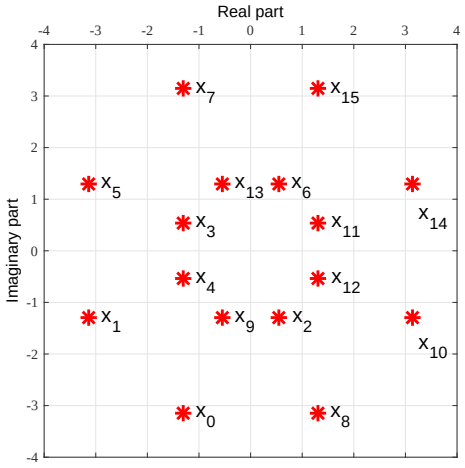


FIGURE 3.7: The received constellation on the first subchannel of $\mathbf{F}_{\ell_1}$-mod precoder, Channel A, Gray-M mapping and 4-QAM modulation.

Let us denote the 16 possible values of $\mathbf{s}$ by $\mathbf{s}_0 = [s_0\ s_0]^T$, $\mathbf{s}_1 = [s_0\ s_1]^T$, $\mathbf{s}_2 = [s_0\ s_2]^T$, $\mathbf{s}_3 = [s_0\ s_3]^T$, $\mathbf{s}_4 = [s_1\ s_0]^T$, $\mathbf{s}_5 = [s_1\ s_1]^T$, $\mathbf{s}_6 = [s_1\ s_2]^T$, $\mathbf{s}_7 = [s_1\ s_3]^T$, $\mathbf{s}_8 = [s_2\ s_0]^T$, $\mathbf{s}_9 = [s_2\ s_1]^T$, $\mathbf{s}_{10} = [s_2\ s_2]^T$, $\mathbf{s}_{11} = [s_2\ s_3]^T$, $\mathbf{s}_{12} = [s_3\ s_0]^T$, $\mathbf{s}_{13} = [s_3\ s_1]^T$, $\mathbf{s}_{14} = [s_3\ s_2]^T$ and $\mathbf{s}_{15} = [s_3\ s_3]^T$, where $s_0 = (-1-i)/\sqrt{2}$, $s_1 = (-1+i)/\sqrt{2}$, $s_2 = (1-i)/\sqrt{2}$, $s_3 = (1+i)/\sqrt{2}$. The received symbol $\mathbf{x}_i$, which is mapped onto the received constellation, reads $\mathbf{x}_i = \mathbf{H}_v\mathbf{F}_d\mathbf{s}_i$. With this definition, the mappings can be easily presented as shown in TABLE 3.1. This table provides the optimized MSEW mappings (but not unique) applied to the received constellations shown in FIGURE 2.9, FIGURE 2.8 and FIGURE 3.7. Each decimal value represents 4 binary bits that are mapped onto symbol vector $\mathbf{s}_i$, which corresponds to the received constellation point $\mathbf{x}_i$. For example, with MSEW mapping of the $\mathbf{F}_{r1}$ precoder, the symbol vector $\mathbf{s}_0$ is mapped to $(7)_{10}$ in decimal, which equals to $(0111)_2$ in binary. Note that the received constellation with $\mathbf{x}_i$ in FIGURE 2.9, FIGURE 3.3 and FIGURE 3.7 are plotted for Gray-M. Thus, for the Gray-M in TABLE. 3.1, the mapping is normally in the range from 0 to 15.

TABLE 3.1: The optimized binary representation in the constellation map of the precoders for two different mappings.

Mapping	Precoder	$[\mathbf{s}_0 \dots \mathbf{s}_i \dots \mathbf{s}_{15}]$
Gray-M	$\mathbf{F}_{r_1}/\mathbf{F}_{octa}/\mathbf{F}_{\ell_1}$	[0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15]
MSEW	$\mathbf{F}_{r_1}$	[7 2 1 11 13 4 8 14 12 6 10 15 5 3 0 9]
	$\mathbf{F}_{octa}/\mathbf{F}_{\ell_1}$-mod	[2 5 7 0 9 12 10 15 11 14 8 13 4 3 1 6]

In order to compare the precoders associated with the MSEW mapping shown in TABLE. 3.1, let us firstly compare their respective values ℓ_1. FIGURE 3.8 shows the plots of normalized ℓ_1 (ℓ_1 is normalized by ρ) versus γ (*i.e.* all possible channels). As observed, under the criterion of maximizing ℓ_1, it comes that the best precoding solution associated with MSEW mapping corresponds to the switch between the $\mathbf{F}_{r1}$ and $\mathbf{F}_{octa}$ precoders at the newfound threshold $\gamma_0' = 30.7^o$. Indeed, as will be later demonstrated in Section 3.4.3, the performance of max-$d_{\min}$ assuming MSEW mapping is improved by using the new threshold γ_0'. However, the criterion of maximizing ℓ_1 shows its advantages only at the error-floor, which is very low with MSEW mapping and out of the commonly-used SNR region. In this section, we are mostly interested in the turbo-cliff region. In Section 3.4.2, we resort to EXIT chart analysis to completely propose an appropriate precoding solution to be used with MSEW mapping at the received constellation.

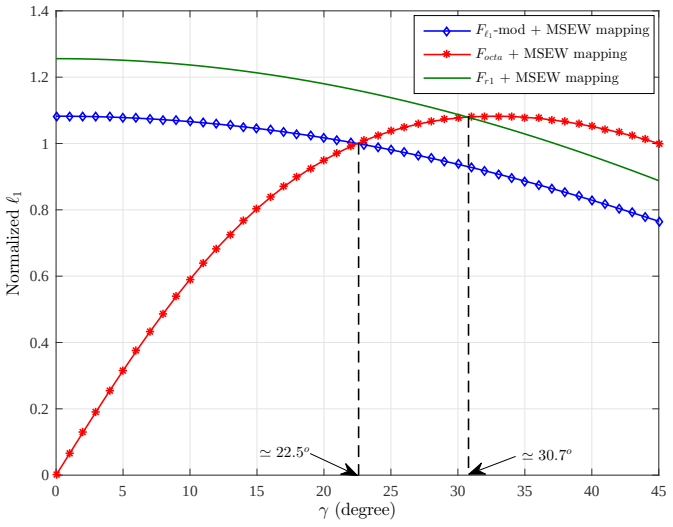


FIGURE 3.8: Normalized ℓ_1 versus γ for different MSEW mapped precoders.

3.4.2 EXIT chart analysis

As introduced in Section 1.3, EXIT chart is a useful tool to optimize the convergence of an iterative receiver based on extrinsic information exchanges between the two elementary component devices. For the turbo detection scheme, one device is the soft-input soft-output FEC decoder and the other one is the MIMO symbol-demapper. We use EXIT chart to analyze the influence of the precoders as well as the mappings on the evolution of the MI between the interleaved FEC-encoded binary sequence $\bar{\mathbf{c}}$ and its LLRs at input

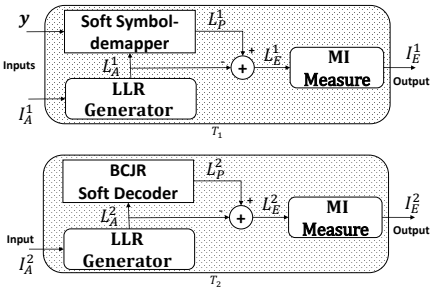


FIGURE 3.9: Block model for the EXIT chart measurement

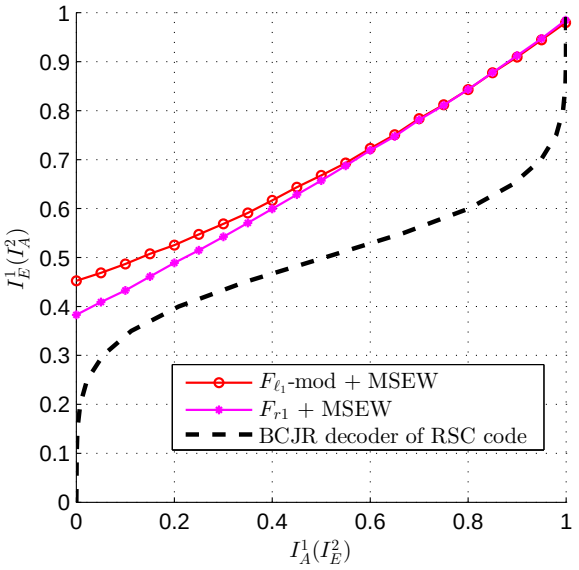


FIGURE 3.10: EXIT chart at SNR = 8.1 dB, Channel A, MSEW mapping given in TABLE 3.1.

and output of the demapper. The extrinsic MI at output of demapper is a function of the *a priori* knowledge I_A^1 and the SNR. We define $I_E^1 = T_1\left(I_A^1, \text{SNR}\right)$. Similarly, $I_E^2 = T_2(I_A^2)$ stands for the extrinsic MI at output of decoder. FIGURE 3.9 illustrates the way we compute the EXIT functions of both devices. The MI extraction as well as the LLRs generation have been introduced in Section 1.3. In this chapter, each EXIT chart measurement is averaged over 100 random realizations.

FIGURE 3.10 shows the EXIT charts of $\mathbf{F}_{\ell_1}$-mod and $\mathbf{F}_{r1}$ precoders for Channel A at SNR = 8.1 dB. Both precoders are associated with the MSEW mapping shown in TABLE. 3.1. We observe that both precoders converge towards a similar ending point. However, the starting point of $\mathbf{F}_{r1}$ is lower than the one of $\mathbf{F}_{\ell_1}$-mod precoder. This can lead to an intersection between the EXIT functions of $\mathbf{F}_{r1}$ precoder and the BCJR decoder. Hence, the EXIT chart predicts that the error-rate performance of $\mathbf{F}_{r1}$ is worse

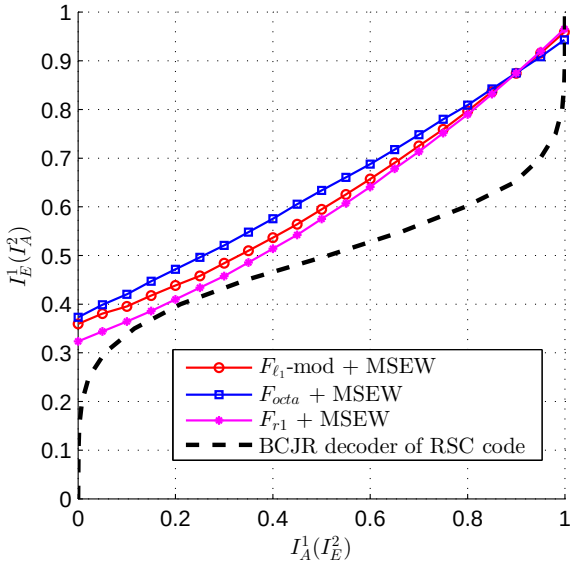


FIGURE 3.11: EXIT chart at SNR = 8.1 dB, Channel B, MSEW mapping given in TABLE 3.1.

compared to $\mathbf{F}_{\ell_1}$-mod at the turbo-cliff region.

Similar analysis is also applied for Channel B as shown in FIGURE 3.11. We observe that the EXIT function of $\mathbf{F}_{octa}$ predicts a better error-rate performance of this precoder compared to $\mathbf{F}_{\ell_1}$-mod precoder at the turbo-cliff region. Additionally, the EXIT chart shows that $\mathbf{F}_{r1}$ is the worst precoder at the turbo-cliff. Note that max-$d_{\min}$ precoder does not work in $\mathbf{F}_{r1}$ mode for Channel B. However, in the case of MSEW mapping, we also consider the $\mathbf{F}_{r1}$ precoder for Channel B. This is to emphasize that, even if maximizing ℓ_1 is a good criterion to design the precoders associated with the MSEW mapping as shown in Section 3.4.1, we need to also consider the EXIT chart analysis to define the best precoding solution in terms of mutual information exchange optimization. According to the EXIT chart analysis, $\mathbf{F}_{r1}$ should be discarded when designing a precoder optimized from the mutual information point of view. We thus keep only $\mathbf{F}_{octa}$ and $\mathbf{F}_{\ell_1}$-mod and resort to the ℓ_1 analysis reported in FIGURE 3.8 to set the switching threshold. The resulting optimized precoder is referred to as *EXIT-based* precoder. It has two precoding forms depending on a threshold $\gamma_1 = 22.5^o$ (see FIGURE 3.8). For the channels with $\gamma \leq \gamma_1$, the *EXIT-based* precoder uses $\mathbf{F}_{\ell_1}$-mod. For the channels with $\gamma > \gamma_1$, the *EXIT-based* precoder uses $\mathbf{F}_{octa}$.

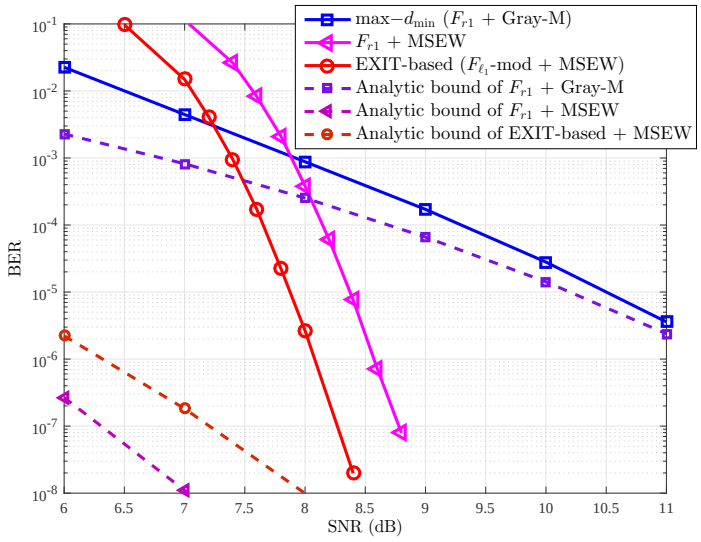


FIGURE 3.12: BER performance of the precoders associated with the corresponding mappings, 4-QAM, Channel A.

3.4.3 Simulation results

In this section, we present the simulation results in terms of error-rate performance of the precoders associated with the MSEW mapping to confirm the EXIT chart analysis done in Section 3.4.2. Performance of the max-$d_{\min}$ precoder (*i.e.* $\mathbf{F}_{r1}$ or $\mathbf{F}_{octa}$ used with Gray-M mapping), which has been shown in Section 3.3.3, is also put again for comparison purpose. The similar simulation setup as presented in Section 3.3.3 is considered.

FIGURE 3.12 shows the error-rate performance of the precoders used with the corresponding mappings for Channel A. As illustrated in this figure, the *EXIT-based* precoder achieves a gain of 2.6 dB at BER = 10^{-5} compared to the max-$d_{\min}$ precoder (with $\mathbf{F}_{r1}$ mode for Channel A and Gray-M mapping). The $\mathbf{F}_{r1}$ used with MSEW also achieves a significant gain compared to the conventional max-$d_{\min}$ precoder, which is about 2.1 dB at BER = 10^{-5}. On the other hand, we observe that the analytical bound of the $\mathbf{F}_{r1}$ used with MSEW mapping is lower than the one of *EXIT-based* precoder, which is in accordance with the ℓ_1 analysis done in Section 3.4.1. However, the convergence to the analytical bound is achieved for very low BER values out of the commonly targeted BER range. In contrast, the performance at the turbo-cliff region is more interesting. We see that, at the turbo-cliff, the *EXIT-based* precoder achieves a gain of 0.5 dB at BER = 10^{-6} compared to the $\mathbf{F}_{r1}$ used with MSEW. This is in accordance with the EXIT chart analysis done in Section 3.4.2.

Similar conclusions are also obtained for Channel B. As illustrated in FIGURE 3.13, the *EXIT-based* precoder achieves a gain of 2.3 dB at BER $= 10^{-5}$ compared to max-$d_{\min}$ precoder. At the turbo-cliff, the *EXIT-based* achieves a gain of 0.5 dB and 1 dB respectively at BER $= 10^{-6}$ compared to $\mathbf{F}_{\ell_1}$-mod and $\mathbf{F}_{r1}$ precoders, both being associated with MSEW mapping.

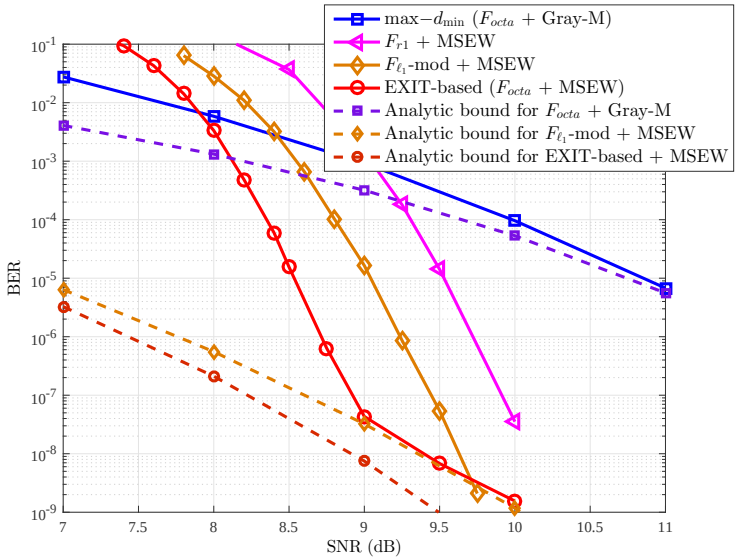


FIGURE 3.13: BER performance of the precoders associated with the corresponding mappings, 4-QAM, Channel B.

We finally illustrate the average error-rate performance over random channels, *i.e.* each

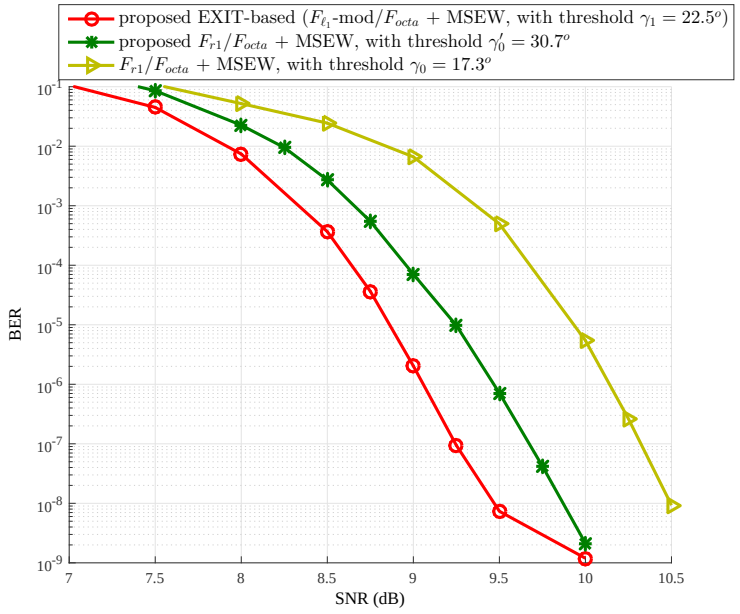


FIGURE 3.14: BER performance of the precoders associated with the MSEW mappings over random channels.

element of $\mathbf{H}$ is distributed as $\mathrm{H}_{i,j} \sim \mathcal{CN}(0,1)$. Thus, on average, each channel element has unit energy ($\mathrm{E}[|\mathrm{H}_{ij}|^2] = 1$). Note that, with the definition of instantaneous SNR in (3.3), the error-rate performance does not depend on ρ^2. Therefore, the system performance for different values of γ is obtained by taking the average of the randomly generated channels. Applying the optimized MSEW mappings for both $\mathbf{F}_{r1}$ and $\mathbf{F}_{octa}$ forms, as observed from FIGURE 3.14, the proposed precoding strategy that uses the switching threshold at $\gamma_0' = 30.7^o$ (the criterion is to maximize ℓ_1, see FIGURE 3.8) outperforms the one that uses the conventional switching threshold $\gamma_0 = 17.3^o$ (proposed in [16] for max-$d_{\min}$ precoder). The proposed *EXIT-based* precoder, which assumes the optimized MSEW mapping for $\mathbf{F}_{\ell_1}$-mod and $\mathbf{F}_{octa}$ and switches among them by using the threshold $\gamma_1 = 22.5^o$, outperforms the other solutions at the commonly used BER region. This is in accordance with the EXIT chart analysis done in Section 3.4.2. At the very low BER (BER $= 10^{-9}$), the *EXIT-based* precoder begins to saturate. This is in accordance with the ℓ_1 analysis done for MSEW mapping as shown in FIGURE 3.8. Since *EXIT-based* precoder yields the priority to optimize the turbo-cliff region (corresponds to the bottle-neck of EXIT chart) rather than the error-floor region (corresponds to ℓ_1, or the ending point of the EXIT function of the soft-demapper).

3.5 Conclusion

A first precoder referred to as $\mathbf{F}_{\ell_1}$ precoder, which focuses on the maximization of the minimum Euclidean distance between symbols with binary mapped sequences differing by one position, was proposed for the conventional Gray-M mapping (bit-interleaved codewords are mapped at the modulator before converted to MIMO symbols). In terms of error-rate performance, the $\mathbf{F}_{\ell_1}$ precoder significantly outperforms the max-$d_{\min}$ precoder, which shows the best performance for the Gray-M mapped, maximum likelihood detected, uncoded systems. Additionally, since the received constellation is fixed, we proposed a MIMO symbol mapper that directly maps the interleaved FEC-encoded binary sequence into MIMO symbols. Thanks to the MIMO symbol mapper, the MSEW mapping technique can be applied to the received constellation. By using the MSEW mapping criterion, ℓ_1 is maximized. Consequently, the error-floor of the MSEW-mapped precoded system is very low, which are out of the commonly used BER. Therefore, in the case of MSEW mapping, we focused on the precoding optimization at the turbo-cliff

region rather than at the error-floor (*i.e.* rather than maximizing ℓ_1). Taking benefit from the EXIT chart analysis, another precoder referred to as *EXIT-based* precoder was proposed in this chapter. The *EXIT-based* precoder is not only adapted to be used with the MSEW mapping at the received constellation but also takes into account the optimization at turbo-cliff region. We observed from the simulation that the *EXIT-based* precoder significantly outperforms the other precoders when used with an outer FEC code and a turbo detection.

As a result from this chapter, we deduce that optimizing only ℓ_1 (*e.g.* by using the MSEW mapping) results in a lower error-floor but it may lead to an intersection at the bottle-neck of EXIT chart, which scales the turbo-cliff region to a higher SNR. On the other hand, let us recall from Section 2.3.3 that the area beneath the EXIT function of the soft-demapper is proportional to the channel capacity. Therefore, it is interesting to firstly maximize the channel capacity (maximize the area beneath the EXIT function of the soft-demapper to avoid the intersection at bottle-neck) and consequently apply MSEW mapping to maximize ℓ_1 (maximize the ending point of the EXIT function of the soft-demapper) to achieve a lower error-floor. Therefore, the investigation of the GOPT precoder, which globally maximizes the channel capacity as introduced in Section 2.3.2.1, is interesting for this scheme. The main drawback of the GOPT algorithm is the high complexity. In the next chapter, we propose a low-complex suboptimal precoding algorithm to overcome this drawback. The new algorithm not only significantly reduces the complexity and assures a close channel capacity to GOPT, but also allows us to apply MSEW mapping, which is intractable for the conventional GOPT algorithm.

Appendices of chapter 3

A Calculation of σ_ξ^2

To simplify the notation, let us write $\mathbf{C} = \mathbf{B}(\bar{\mathbf{s}} - \bar{\mathbf{s}}_l)$ and $\mathbf{D} = \mathbf{C}^\dagger \boldsymbol{\eta}$. Since $\mathbf{D}_{i,i}$ are independent complex random variable, the variance of the complex random variable $\mathrm{Tr}\{\mathbf{D}\}$ reads

$$\begin{aligned}
\mathrm{var}\left(\mathrm{Tr}\{\mathbf{D}\}\right) &= \mathrm{var}\left(\sum_i \{\mathbf{D}_{i,i}\}\right), \\
&= \sum_i \mathrm{var}\left(\mathbf{D}_{i,i}\right), \\
&= \sum_i \mathrm{var}\left(\mathbf{C}_{:,i}^\dagger \boldsymbol{\eta}_{:,i}\right), \\
&= \sum_i \mathrm{E}\left[\left(\mathbf{C}_{:,i}^\dagger \boldsymbol{\eta}_{:,i} - \mathrm{E}[\mathbf{C}_{:,i}^\dagger \boldsymbol{\eta}_{:,i}]\right)\left(\mathbf{C}_{:,i}^\dagger \boldsymbol{\eta}_{:,i} - \mathrm{E}[\mathbf{C}_{:,i}^\dagger \boldsymbol{\eta}_{:,i}]\right)^\dagger\right], \\
&= \sum_i \mathbf{C}_{:,i}^\dagger \underbrace{\mathrm{E}\left[\left(\boldsymbol{\eta}_{:,i} - \mathrm{E}[\boldsymbol{\eta}_{:,i}]\right)\left(\boldsymbol{\eta}_{:,i} - \mathrm{E}[\boldsymbol{\eta}_{:,i}]\right)^\dagger\right]}_{\sigma_\eta^2 \mathbf{I}_b} \mathbf{C}_{:,i}, \\
&= \sigma_\eta^2 \underbrace{\sum_i \mathbf{C}_{:,i}^\dagger \mathbf{C}_{:,i}}_{\|\mathbf{C}\|_F^2}, \\
&= \sigma_\eta^2 \|\mathbf{B}(\bar{\mathbf{s}} - \bar{\mathbf{s}}_l)\|_F^2.
\end{aligned} \tag{5.21}$$

Therefore, we have $\mathrm{Tr}\left\{\left(\mathbf{B}(\bar{\mathbf{s}} - \bar{\mathbf{s}}_l)\right)^\dagger \boldsymbol{\eta}\right\} \sim \mathcal{CN}\left(0, \sigma_\eta^2 \|\mathbf{B}(\bar{\mathbf{s}} - \bar{\mathbf{s}}_l)\|_F^2\right)$. Hence, $\xi \sim \mathcal{N}\left(0, \sigma_\xi^2\right)$, where $\sigma_\xi^2 = 2\sigma_\eta^2 \|\mathbf{B}(\bar{\mathbf{s}} - \bar{\mathbf{s}}_l)\|_F^2$.

B Proof of Lemma 2

From (3.16), the problem of finding (ψ, θ) that maximize D_1 becomes

$$(\psi^\diamond, \theta^\diamond) = \arg\max_{\mathcal{F}} \left(\sigma_1^2 \cos^2\psi \cos^2\theta + \sigma_2^2 \sin^2\psi \sin^2\theta\right). \tag{5.22}$$

Recall that the set $\mathcal{F}$ is defined by $0^o \leq \psi \leq 90^o$ and $0^o \leq \theta \leq 90^o$. Partial derivatives of D_1 with respect to θ and ψ yield

$$\frac{\partial D_1}{\partial \psi} = \sin 2\psi \left(\sigma_2^2 \sin^2\theta - \sigma_1^2 \cos^2\theta\right), \tag{5.23}$$

$$\frac{\partial D_1}{\partial \theta} = \sin 2\theta \left(\sigma_2^2 \sin^2\psi - \sigma_1^2 \cos^2\psi\right). \tag{5.24}$$

The set of points (ψ, θ) of $\mathcal{F}$ that satisfy the first order conditions for a local maximum ($\nabla D_1 = 0$), denoted by $\mathcal{F}^\diamond$, equals

$$\mathcal{F}^\diamond = \left\{(0,0), (0,90^o), (90^o,0), (90^o,90^o), \left(\arctan\frac{\sigma_1}{\sigma_2}, \arctan\frac{\sigma_1}{\sigma_2}\right)\right\}. \tag{5.25}$$

The only point of $\mathcal{F}^\diamond$ that satisfies the second order conditions for a local maximum is $(\psi, \theta) = (0,0)$, where which yields $D_1^{\max} = \sigma_1^2$. The same optimization is applied for D_2, which yields $D_2^{\max} = \sigma_1^2$ at $(\psi = 0, \theta = 90^o)$.

C Proof of Lemma 3

For any difference vector $\boldsymbol{\nu} = (\nu_1 \quad \nu_2)^T$, let us define

$$\phi_1 = \arg\max_{0^o \leq \phi \leq 45^o} \left(\min \frac{\sigma_1\sqrt{2}}{2} \left|\nu_1 + \nu_2 e^{i\phi}\right|\right) \tag{5.26}$$

and

$$\phi_2 = \arg\max_{45^o \leq \check{\phi} \leq 90^o} \left(\min \frac{\sigma_1\sqrt{2}}{2} \left|\nu_1 + \nu_2 e^{i\check{\phi}}\right|\right). \tag{5.27}$$

By considering $\check{\phi} = \frac{\pi}{2} - \phi$, we obtain

$$\begin{aligned} \phi_2 &= \underset{0^o \le \phi \le 45^o}{\arg\max} \left(\min \frac{\sigma_1 \sqrt{2}}{2} \left| \nu_1 + \nu_2 e^{i(\frac{\pi}{2} - \phi)} \right| \right), \\ &= \underset{0^o \le \phi \le 45^o}{\arg\max} \left(\min \frac{\sigma_1 \sqrt{2}}{2} \left| \nu_1^* + e^{-i\frac{\pi}{2}} \nu_2^* e^{i\phi} \right| \right), \end{aligned} \tag{5.28}$$

where ν^* stands for the conjugation of ν. Further, for Q-ary QAM modulation, given any difference vector $\boldsymbol{\nu} = \begin{pmatrix} \nu_1 & \nu_2 \end{pmatrix}^T \in \mathcal{V}$, the vector $\begin{pmatrix} \nu_1^* & e^{-i\frac{\pi}{2}} \nu_2^* \end{pmatrix}^T$ is also a valid difference vector in $\mathcal{V}$ thanks to the symmetry of the constellation. Therefore, from (5.26) and (5.28), we obtain that $\phi_1 = \phi_2$, *i.e.* the lemma is proved.

Chapter 4

Complexity Reduction for the Optimization of Linear Precoders over Random MIMO Channels

The content of this chapter is mainly based on the following paper:

- Nhat-Quang Nhan, Philippe Rostaing, Karine Amis, Ludovic Collin, and Emanuel Radoi. "Complexity Reduction for the Optimization of Linear Precoders over Random MIMO Channel". Submitted to *IEEE Transaction on Wireless Communications*, pending for review.

4.1 Introduction

The optimal precoder for encoded systems over MIMO channels is the globally optimized (referred to as GOPT) precoder [22], which has been introduced in Section 2.3.2.1. The GOPT precoder aims to globally maximize the mutual information between the finite alphabet input and the corresponding channel output. In [22], the authors proposed an algorithm (referred to as GOPT algorithm) to find the optimal precoding matrix for each MIMO channel at a given SNR value. Thus, this precoder does not have a fixed closed-form and it is SNR-dependent. Unfortunately, since it was targeted for the optimal solution, the computational complexity of the GOPT algorithm is painfully high.

It is the high complexity that limits the application of GOPT over random channels and makes GOPT to be considered as a lower bound for fixed channels only.

In this chapter, we focus on the complexity reduction for the GOPT precoding algorithm. Our main contributions are twofold. On one hand, by using a lower bound of the channel mutual information, it is proved in [68] that maximizing the minimum Euclidean distance $d_{\min}$ is asymptotically equivalent to maximizing the channel mutual information. On the other hand, transformation allows us to present a precoding matrix as $\mathbf{F}_d = \mathbf{\Psi\Theta}$, where the diagonal matrix $\mathbf{\Psi}$ links to the power allocation on the subchannels and the unitary matrix $\mathbf{\Theta}$ links to the forms of the received constellations. Therefore, as the first contribution, we propose a novel low complexity optimized (referred to as LCOPT) precoding algorithm, which has a lower computational complexity compared to GOPT. The proposed algorithm uses fixed unitary matrix $\mathbf{\Theta}$ taken from the solution of maximizing $d_{\min}$ and optimizes only the power allocation matrix $\mathbf{\Psi}$ based on mutual information criterion. The selection of $\mathbf{\Theta}$ is discussed for two cases namely $b = 2$ and $b > 2$. With a fixed $\mathbf{\Theta}$, the received constellation form of LCOPT is fixed. Thus, as the second contribution, we propose to apply the maximum squared Euclidean weight (MSEW) mapping [66], which has been introduced in Chapter 3, on the received constellation of the proposed LCOPT precoder. Note that it is impractical to apply MSEW mapping on the received constellation of GOPT since its constellation is changed for each channel realization and SNR. To avoid the search of optimal GOPT precoding matrix for each channel realization and SNR in case $b = 2$, we also propose in this chapter a method to construct precoding codebooks for GOPT precoder.

The remainder of this chapter is organized as follows. Section 4.2 introduces the system model. In this section, the conventional GOPT precoding algorithm is recalled in a more general scheme, which takes into account the channel matrix dimension reduction as introduced in Chapter 2. This allows us to apply GOPT to the systems that desire a given number of data streams. Section 4.3 presents the proposed LCOPT algorithm. The selections of $\mathbf{\Theta}$ are also discussed in this section for two cases, namely $b = 2$ and $b > 2$. In addition, the association of LCOPT precoder and MIMO symbol mapper, which allows mapping on received constellation, is introduced. Section 4.4 begins with the proposed codebook construction method for GOPT in case $b = 2$. The simulation results are then presented. Finally, the EXIT chart analysis is used to validate the simulation results. Section 4.5 concludes the chapter.

4.2 Preliminaries

4.2.1 System model

We consider a MIMO system with n_R receive, n_T transmit antennas and b independent data streams to be transmitted. The transmitting binary sequence is grouped and mapped onto Q-ary modulated symbols s. These symbols are then converted into MIMO symbol vectors $\mathbf{s}$, which have size $b \times 1$. Each vector $\mathbf{s}$ is then precoded with a precoding matrix $\mathbf{F}$ and transmitted through the MIMO channel. According to (2.16), the channel output $\mathbf{y}$ after the channel transformation reads $\mathbf{y} = \mathbf{H}_v\mathbf{F}_d\mathbf{s} + \boldsymbol{\eta}$. With the equivalent model, the channel mutual information between the discrete input $\mathbf{s}$ and the channel output $\mathbf{y}$ ($\mathcal{I}(\mathbf{y},\mathbf{s})$) is given by (2.24). It is proved that the mutual information $\mathcal{I}(\mathbf{y},\mathbf{s})$ is a concave function with respect to $\mathbf{W}$ [22]. Note that $\zeta_{m,k}$ in (2.24) can be directly expressed in terms of $\mathbf{W}$ as shown in Appendix C. The gradient of $\mathcal{I}(\mathbf{y},\mathbf{s})$ with respect to $\mathbf{W}$, which is denoted by $\nabla_{\mathbf{W}}\mathcal{I}(\mathbf{y},\mathbf{s})$, reads

$$\nabla_{\mathbf{W}}\mathcal{I}(\mathbf{y},\mathbf{s}) = \frac{\boldsymbol{\Phi}}{\sigma_\eta^2}, \tag{4.1}$$

where $\boldsymbol{\Phi}$ is the minimum mean square error (MMSE) matrix, which is estimated by $\boldsymbol{\Phi} = \mathrm{E}\left\{(\mathbf{s}-\hat{\mathbf{s}})(\mathbf{s}-\hat{\mathbf{s}})^\dagger\right\}$, $\hat{\mathbf{s}} = \mathrm{E}\{\mathbf{s}|\mathbf{y}\}$. Note that, using the gradient ascent method, we may update

$$\mathbf{W}_{k+1} = \mathbf{W}_k + \delta\mathbf{W}, \tag{4.2}$$

where $\delta\mathbf{W} = \mu_{\mathbf{W}}\nabla_{\mathbf{W}}\mathcal{I}(\mathbf{y},\mathbf{s})$, in which $\mu_{\mathbf{W}}$ is a sufficient step size. Then, we have to also update

$$\mathbf{F}_d^{(k+1)} = \mathbf{F}_d^{(k)} + \delta\mathbf{F}_d, \tag{4.3}$$

where the updated precoder has to satisfy the power constraint $\|\mathbf{F}_d^{(k+1)}\|_F^2 = \|\mathbf{F}_d^{(k)}\|_F^2 = 1$. However, solving directly for the incremental update $\delta\mathbf{F}_d$ can be difficult. Additionally, in that case, the step size $\mu_{\mathbf{W}}$ must be very small to avoid divergence and, hence, it slows down the convergence. To tackle this challenge, the authors in [22] developed an algorithm that partially optimizes $\mathbf{F}_d$ in order to maximize $\mathcal{I}(\mathbf{y},\mathbf{s})$. The algorithm is briefly recalled in Section 4.2.2.

4.2.2 Globally Optimized (GOPT) precoders

In this subsection, we summary the GOPT [22] algorithm in a more accessible way to make a foundation for our proposed scheme in Section 4.3.1. The contribution in this part is that we recall the GOPT algorithm for the equivalent model in (2.16). The equivalent model is more general than the model considered in [22] since the channel dimension reduction has been taken into account in (2.16), which allows the transmission of any desired number of data streams. The SVD of $\mathbf{F}_d$ gives $\mathbf{F}_d = \mathbf{U}_\mathbf{F}\mathbf{\Sigma}_\mathbf{F}\mathbf{V}_\mathbf{F}^\dagger$.

Proposition 1. With $\mathbf{F}_d = \mathbf{U}_\mathbf{F}\mathbf{\Sigma}_\mathbf{F}\mathbf{V}_\mathbf{F}^\dagger$, the matrix $\mathbf{U}_\mathbf{F}$ must be equal to the identity matrix $\mathbf{I}_b$ to achieve the maximum channel mutual information.

Proof. See Appendix A. □

As we aim to maximize the mutual information according to Proposition 1, we have to set $\mathbf{U}_\mathbf{F} = \mathbf{I}_b$. Hence, by rewriting $\mathbf{\Psi} = \mathbf{\Sigma}_\mathbf{F}$, $\mathbf{\Theta} = \mathbf{V}_\mathbf{F}^\dagger$, we deduce

$$\mathbf{F}_d = \mathbf{\Psi}\mathbf{\Theta}. \tag{4.4}$$

The matrix $\mathbf{\Psi}$ controls the power allocation on each subchannel, while $\mathbf{\Theta}$ concerns itself with the rotation and scaling of the symbols on the received constellation. Note that the authors in [22] proposed to select the left singular matrix of $\mathbf{F}$ to be equal to the right singular matrix of $\mathbf{H}$. In our case, we consider the equivalent model in (2.16). Therefore, the conditions established in Proposition 1 must be taken into account.

It is proved that the mutual information $\mathcal{I}(\mathbf{y},\mathbf{s})$ is also a concave function with respect to $\mathbf{\Psi}^2$ [22]. The gradient of $\mathcal{I}(\mathbf{y},\mathbf{s})$ with respect to $\mathbf{\Psi}^2$, which is denoted by $\nabla_{\mathbf{\Psi}^2}\mathcal{I}(\mathbf{y},\mathbf{s})$, reads

$$\nabla_{\mathbf{\Psi}^2}\mathcal{I}(\mathbf{y},\mathbf{s}) = \operatorname{diag}\left(\mathbf{H}_v^2\mathbf{\Theta}\mathbf{\Phi}\mathbf{\Theta}^\dagger\right). \tag{4.5}$$

With $\mathbf{F}_d = \mathbf{\Psi}\mathbf{\Theta}$, the authors in [22] proposed an iterative algorithm that respectively updates $\mathbf{\Theta}$ and $\mathbf{\Psi}$ based on the gradient ascent method (instead of directly update $\mathbf{F}_d$, which seems to be infeasible). Though $\mathbf{\Psi}$ can be updated directly using its gradient in (4.5), we however must rely on the incremental of $\mathbf{W}$ to update $\mathbf{\Theta}$. The matrix $\mathbf{\Theta}$ can

be updated by $\delta\mathbf{W}$ as follows. Any $b \times b$ unitary matrix $\mathbf{\Theta}$ can be written as

$$\mathbf{\Theta} = \mathbf{D} \prod_{p=b-1}^{1} \prod_{q=p+1}^{b} \mathbf{U}_{pq}(\omega_{pq}, \nu_{pq}), \tag{4.6}$$

where $\mathbf{D}$ is an $b \times b$ unitary diagonal matrix, which does not affect $\mathcal{I}(\mathbf{y}, \mathbf{s})$ and can be chosen as an identity matrix in this case. The matrix $\mathbf{U}_{pq}(\omega_{pq}, \nu_{pq})$ is formed by replacing the $(p,p)^{\text{th}}$,$(q,q)^{\text{th}}$, $(p,q)^{\text{th}}$ and $(q,p)^{\text{th}}$ entries of the identity matrix $\mathbf{I}_b$ by $\cos\omega_{pq}$, $\cos\omega_{pq}$, $\sin\omega_{pq}e^{-j\nu_{pq}}$ and $-\sin\omega_{pq}e^{j\nu_{pq}}$ respectively, where $\omega_{pq} \in (-\pi, \pi]$ and $\nu_{pq} \in [-\frac{\pi}{2}, \frac{\pi}{2}]$. The incremental $\delta\omega_{pq}$ and $\delta\nu_{pq}$, which are used to update $\mathbf{\Theta}$, can be found by solving

$$\delta\mathbf{\Theta} = \sum_{p=b-1}^{1} \sum_{q=p+1}^{b} \frac{\partial\mathbf{\Theta}}{\partial\omega_{pq}}\delta\omega_{pq} + \sum_{p=b-1}^{1} \sum_{q=p+1}^{b} \frac{\partial\mathbf{\Theta}}{\partial\nu_{pq}}\delta\nu_{pq}, \tag{4.7}$$

where $\delta\mathbf{\Theta}$ is calculated from $\delta\mathbf{W}$ by the following first-order approximation

$$\delta\mathbf{W} \simeq [\delta\mathbf{\Theta}]^{\dagger}\mathbf{H}_v^2\mathbf{\Psi}^2\mathbf{\Theta} + \mathbf{\Theta}^{\dagger}\mathbf{H}_v^2\mathbf{\Psi}^2[\delta\mathbf{\Theta}]. \tag{4.8}$$

An example, which solves for $\delta\omega_{pq}$ and $\delta\nu_{pq}$ from $\delta\mathbf{W}$ in case $b = 2$, is given in Appendix B. The associated algorithm is briefly recalled hereinafter as Algorithm 1. The precoding matrices, which are found by using this algorithm, are referred to as GOPT precoders.

4.3 Analysis

4.3.1 Low Complexity Optimized (LCOPT) precoders

Although the GOPT precoder globally maximizes the channel capacity thanks to Algorithm 1, the computational complexity of this algorithm is painfully expensive. Indeed, the complexity cost is caused by two main issues listed below.

As for the first issue, in order to find the step sizes $\mu_{\mathbf{W}}$ and $\mu_{\mathbf{\Psi}^2}$, Algorithm 1 has to compute the channel mutual information $\mathcal{I}(\mathbf{y}, \mathbf{s})$ many times during the backtracking line search (BTLS) process. However, as shown in (2.24), the computation cost of $\mathcal{I}(\mathbf{y}, \mathbf{s})$ is high, since it needs a high number of trials for the accuracy of expectation. To tackle the challenge, in [68], the lower bound of channel mutual information was proposed to avoid the estimation of $\mathcal{I}(\mathbf{y}, \mathbf{s})$. The lower bound of channel mutual information, which

Algorithm 1 Globally optimization algorithm for linear precoders [22].

Inputs: Maximum number of iterations l_{max}, equivalent channel matrix $\mathbf{H}_v$ and signal-to-noise ratio (SNR).
Initialization: Define $f(\mathbf{W})$ as a function of mutual information $\mathcal{I}(\mathbf{y},\mathbf{s})$ with input $\mathbf{W}$ (see Appendix C). Define $g(\mathbf{\Psi}^2)$ as a function of mutual information $\mathcal{I}(\mathbf{y},\mathbf{s})$ with input $\mathbf{\Psi}^2$ (using (2.24), where $\mathbf{F}_d = \mathbf{\Psi\Theta}$). Select the initial values for ω_{pq} and ν_{pq} and the initial diagonal matrix for $\mathbf{\Psi}$, which satisfies $\mathrm{Tr}\{\mathbf{\Psi}^2\} = 1$. Normally, we select $\mathbf{\Psi}_{\text{init}} = \mathbf{I}_b/b$. Select $\vartheta \in [0 \quad 0.5]$ and $\zeta \in [0 \quad 1]$ for the backtracking line search (BTLS). Calculate the MMSE matrix $\mathbf{\Phi}$.
while $k \leq k_{max}$ or non-convergence **do**
    From $\mathbf{\Phi}$, find $\nabla_{\mathbf{W}}\mathcal{I}(\mathbf{y},\mathbf{s})$ by (4.1)
    Determine the ascent direction $\Delta_{\mathbf{W}} = \nabla_{\mathbf{W}}\mathcal{I}(\mathbf{y},\mathbf{s})$
    Find the step size $\mu_{\mathbf{W}}$ using BTLS: set $\mu_{\mathbf{W}} = 1$
    while $f(\mathbf{W} + \mu_{\mathbf{W}}\Delta_{\mathbf{W}}) < f(\mathbf{W}) + \vartheta\mu_{\mathbf{W}}[\nabla_{\mathbf{W}}\mathcal{I}(\mathbf{y},\mathbf{s})]^T\Delta_{\mathbf{W}}$ **do**
        $\mu_{\mathbf{W}} = \zeta\mu_{\mathbf{W}}$
    end while
    Update $\delta\mathbf{W} = \mu_{\mathbf{W}}\nabla_{\mathbf{W}}\mathcal{I}(\mathbf{y},\mathbf{s})$. From the updated $\delta\mathbf{W}$, solve for $\delta\mathbf{\Theta}$ by (4.8). From $\delta\mathbf{\Theta}$, solve for $\delta\omega_{pq}$ and $\delta\nu_{pq}$ by (4.7). Update $\omega_{pq} = \omega_{pq} + \delta\omega_{pq}$ and $\nu_{pq} = \nu_{pq} + \delta\nu_{pq}$
    Update $\mathbf{\Theta}$ using the new ω_{pq} and ν_{pq}
    From the new value of $\mathbf{\Theta}$, recompute the MMSE matrix $\mathbf{\Phi}$ (for the next calculation of $\nabla_{\mathbf{\Psi}^2}\mathcal{I}(\mathbf{y},\mathbf{s})$)
    Find $\nabla_{\mathbf{\Psi}^2}\mathcal{I}(\mathbf{y},\mathbf{s}) = \mathrm{diag}\left(\mathbf{H}_v^2\mathbf{\Theta\Phi\Theta}^\dagger\right) - \lambda\mathbf{I}_b$, where $\lambda = \mathrm{Tr}\{\mathrm{diag}\left(\mathbf{H}_v^2\mathbf{\Theta\Phi\Theta}^\dagger\right)\}/b$
    Determine the ascent direction $\Delta_{\mathbf{\Psi}^2} = \nabla_{\mathbf{\Psi}^2}\mathcal{I}(\mathbf{y},\mathbf{s})$
    Find the step size $\mu_{\mathbf{\Psi}^2}$ using BTLS: set $\mu_{\mathbf{\Psi}^2} = 1$
    while $g(\mathbf{\Psi}^2 + \mu_{\mathbf{\Psi}^2}\Delta_{\mathbf{\Psi}^2}) < g(\mathbf{\Psi}^2) + \vartheta\mu_{\mathbf{\Psi}^2}[\nabla_{\mathbf{\Psi}^2}\mathcal{I}(\mathbf{y},\mathbf{s})]^T\Delta_{\mathbf{\Psi}^2}$ **do**
        $\mu_{\mathbf{\Psi}^2} = \zeta\mu_{\mathbf{\Psi}^2}$
    end while
    if $\mu_{\mathbf{\Psi}^2} \approx 0$ **then**
        Update $\mathbf{\Psi}^2 = \mathbf{\Psi}^2 - \mu_{\mathbf{\Psi}^2}\mathbf{\Psi}^2$
    else
        Update $\mathbf{\Psi}^2 = \mathbf{\Psi}^2 + \mu_{\mathbf{\Psi}^2}\mathbf{\Psi}^2$
    end if
    Set any negative diagonal entry of $\mathbf{\Psi}^2$ to zero before normalizing the updated $\mathbf{\Psi}^2$ to satisfy $\mathrm{Tr}\{\mathbf{\Psi}^2\} = 1$
    From the new value of $\mathbf{\Psi}$, update the MMSE matrix $\mathbf{\Phi}$ (for the next calculation of $\nabla_{\mathbf{W}}\mathcal{I}(\mathbf{y},\mathbf{s})$)
    $k = k + 1$
end while
Output: $\mathbf{F}_d = \mathbf{\Psi\Theta}$.

is denoted by $\mathcal{I}_{\mathrm{LB}}(\mathbf{y},\mathbf{s})$, reads

$$\begin{aligned}\mathcal{I}_{\mathrm{LB}}(\mathbf{y},\mathbf{s}) &= b\log_2 Q - (1/\ln 2 - 1)b \\ &\quad - \frac{1}{Q^b}\sum_{m=1}^{Q^b}\log_2\sum_{k=1}^{Q^b}\exp\left(-\frac{\|\mathbf{H}_v\mathbf{F}_d(\mathbf{s}_m-\mathbf{s}_k)\|^2}{2\sigma_\eta^2}\right).\end{aligned} \tag{4.9}$$

The channel mutual information is then approximated by [68]

$$\mathcal{I}(\mathbf{y},\mathbf{s}) \approx \mathcal{I}_{\mathrm{LB}}(\mathbf{y},\mathbf{s}) + (1/\ln 2 - 1)b. \tag{4.10}$$

Since there is no expectation in (4.9), the complexity of Algorithm 1 is reduced by using the approximated mutual information in (4.10) to replace (2.24).

As for the second factor that penalizes the complexity, the computation required for updating $\mathbf{\Theta}$ (*i.e.* to calculate (4.6), (4.7) and (4.8)) is high, especially with high data stream number b. In addition, the convergence of Algorithm 1 is sensible to the initial values (ω_{pq},ν_{pq}) of $\mathbf{\Theta}$, which must be carefully selected for fast convergence. In this chapter, we propose a suboptimal solution to overcome these drawbacks. On one hand, it can be obviously derived from (4.10) that maximizing $\mathcal{I}_{\mathrm{LB}}(\mathbf{y},\mathbf{s})$ helps maximizing the channel mutual information $\mathcal{I}(\mathbf{y},\mathbf{s})$. According to [68], the asymptotic optimality at high SNR region shows that maximizing $\mathcal{I}_{\mathrm{LB}}(\mathbf{y},\mathbf{s})$ at high SNR leads to the following optimization problem

$$\begin{aligned}&\max\left\{\min_{m\neq k}\left\{\|\mathbf{H}_v\mathbf{F}_d(\mathbf{s}_m-\mathbf{s}_k)\|^2\right\}\right\}, \\ &\text{subject to } \|\mathbf{F}_d\|_F^2 = 1.\end{aligned} \tag{4.11}$$

Therefore, at high SNR region, maximizing $\mathcal{I}_{\mathrm{LB}}(\mathbf{y},\mathbf{s})$ is equivalent to maximizing the minimum Euclidean distance $d_{\min} = \min\limits_{m\neq k}\left\{\|\mathbf{x}_m-\mathbf{x}_k\|^2\right\}$ among the received constellation symbols, which are defined by $\mathbf{x}_i = \mathbf{H}_v\mathbf{F}_d\mathbf{s}_i$. On the other hand, let us recall from (4.4) that the matrix $\mathbf{\Psi}$ controls the power allocation on each subchannel, while $\mathbf{\Theta}$ concerns itself with the rotation and scaling of the symbols on the received constellation. Hence, with a fixed matrix $\mathbf{\Theta}$, the received constellation form is fixed. Therefore, by assuming that the matrix $\mathbf{\Theta}$ can be chosen thanks to the solution of maximizing $d_{\min}$, we propose a novel low-complexity sub-optimal algorithm (see Algorithm 2 below) that only updates the power allocation matrix $\mathbf{\Psi}$. We refer to the precoding matrices found by using the new low-complexity algorithm as low-complexity optimized (LCOPT) precoders. We would like to point out that, with high-order modulation, the computational complexity

for the estimation of the MMSE matrix $\mathbf{\Phi}$ is high. By using the new method, we only estimate the MMSE matrix $\mathbf{\Phi}$ once per iteration, instead of twice per iteration as by using Algorithm 1 [22]. Therefore, the complexity is further reduced. Discussions about the selection of $\mathbf{\Theta}$ in cases $b = 2$ and $b > 2$ are presented in Section 4.3.2 and Section 4.3.3 respectively.

Algorithm 2 Low complexity optimization algorithm for linear precoders.

1: **Inputs:** Maximum number of iterations l_{max}, equivalent channel matrix $\mathbf{H}_v$, SNR and $\mathbf{\Theta}$.
2: **Initialization:** Define $g(\mathbf{\Psi}^2)$ as a function of mutual information $\mathcal{I}(\mathbf{y}, \mathbf{s})$ with input $\mathbf{\Psi}^2$ (using (2.24) or (4.10), where $\mathbf{F}_d = \mathbf{\Psi\Theta}$). Select $\mathbf{\Psi}_{\text{init}} = \mathbf{I}_b/b$. Calculate the MMSE matrix $\mathbf{\Phi}$.
3: **while** $k \leq k_{max}$ or non-convergence **do**
4: Find $\nabla_{\mathbf{\Psi}^2}\mathcal{I}(\mathbf{y}, \mathbf{s}) = \text{diag}\left(\mathbf{H}_v^2\mathbf{\Theta\Phi\Theta}^\dagger\right) - \lambda\mathbf{I}_b$, where $\lambda = \text{Tr}\{\text{diag}\left(\mathbf{H}_v^2\mathbf{\Theta\Phi\Theta}^\dagger\right)\}/b$ (note that $\mathbf{\Theta}$ is fixed and reads value from the input)
5: Determine the ascent direction $\Delta_{\mathbf{\Psi}^2} = \nabla_{\mathbf{\Psi}^2}\mathcal{I}(\mathbf{y}, \mathbf{s})$
6: Find the step size $\mu_{\mathbf{\Psi}^2}$ by using BTLS and update $\mathbf{\Psi}^2$ in a similar way as presented in Algorithm 1
7: From the new value of $\mathbf{\Psi}$, update the MMSE matrix $\mathbf{\Phi}$ (for the next calculation of $\nabla_{\mathbf{\Psi}^2}\mathcal{I}(\mathbf{y}, \mathbf{s})$)
8: $k = k + 1$
9: **end while**
10: **Output:** $\mathbf{F}_d = \mathbf{\Psi\Theta}$

4.3.2 Selection of input $\mathbf{\Theta}$ for Algorithm 2 in case $b = 2$

The precoder that maximizes $d_{\min}$, which is referred to as max-$d_{\min}$ precoder, has different solutions categorized by the modulation order [16, 17]. For the commonly used 4-QAM modulation, the optimal solution is the one introduced in Section 2.3.2.2. Let us respectively rewrite (2.27) and (2.28) in Section 2.3.2.2 as follows

- if $0 \leq \gamma \leq \gamma_0$

$$\mathbf{F}_d = \mathbf{F}_{r_1} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \underbrace{\begin{pmatrix} \sqrt{3+\sqrt{3}} & \sqrt{3-\sqrt{3}}e^{i\frac{\pi}{12}} \\ -\sqrt{3-\sqrt{3}} & \sqrt{3+\sqrt{3}}e^{i\frac{\pi}{12}} \end{pmatrix}\frac{1}{\sqrt{6}}}_{\mathbf{\Theta}_{\mathbf{F}_{r1}}}, \tag{4.12}$$

- if $\gamma_0 < \gamma \leq \pi/4$

$$\mathbf{F}_d = \mathbf{F}_{octa} = \begin{pmatrix} \cos\psi & 0 \\ 0 & \sin\psi \end{pmatrix} \underbrace{\begin{pmatrix} 1 & e^{i\frac{\pi}{4}} \\ -1 & e^{i\frac{\pi}{4}} \end{pmatrix}\frac{1}{\sqrt{2}}}_{\mathbf{\Theta}_{\mathbf{F}_{octa}}}, \tag{4.13}$$

where $\psi = \arctan\frac{\sqrt{2}-1}{\tan\gamma}$ and $\gamma_0 = \arctan\sqrt{\frac{\sqrt{2}-1}{2\sqrt{2}+\sqrt{6}-1}}$ ($\approx 17,28^0$). Note that this solution is SNR-independent and it depends on γ (defined in Section 3.2.2) to switch between the two precoding forms.

Since the forms of the received constellations are fixed and since the max-$d_{\min}$ precoder has been designed such that $d_{\min}$ is maximum, we can define the LCOPT precoder by taking the optimized matrix $\mathbf{\Theta}$ of the max-$d_{\min}$ precoder and by searching for the power allocation matrix $\mathbf{\Psi}$ that maximizes the channel mutual information. Note that it is not necessary to use both $\mathbf{\Theta}_{\mathbf{F}_{r1}}$ and $\mathbf{\Theta}_{\mathbf{F}_{octa}}$. In fact, with $\gamma < \gamma_0$ ($\sigma_1 \gg \sigma_2$), the suitable power allocation is $\psi = 0$, *i.e* spreading power only on the first subchannel. Therefore, we just need to use the Algorithm 2 with the input $\mathbf{\Theta} = \mathbf{\Theta}_{\mathbf{F}_{octa}}$ fixed for any value of γ. The found precoder is then compared to $\mathbf{F}_{r1}$ in terms of channel mutual information to find the best precoding matrix.

Let us denote by RPA_i the received power allocated on the i^{th} subchannel, which reads $\text{RPA}_i = \sigma_i\sqrt{P_i}$, and by Γ the ratio between RPA_1 and RPA_2 ($\Gamma = \text{RPA}_1/\text{RPA}_2$). FIGURE 4.1 illustrates Γ for all values of γ. We observe that Γ of $\mathbf{F}_{octa}$ keeps constant for all $\gamma > \gamma_0$ and SNRs. Indeed, we can deduce from the expression of ψ in (4.13) that $\tan\psi\tan\gamma = \sqrt{2}-1$, which implies $\Gamma = 1/(\sqrt{2}-1)$. For LCOPT, the received power allocated on the first subchannel is always greater than on the second one ($\text{RPA}_1 \geq \text{RPA}_2$, the equality is reached at the limit for $\gamma = 45^o$). Considering LCOPT precoder for a fixed SNR, we observe that the lower γ, the greater RPA_1 compared to RPA_2. Considering LCOPT precoder for a fixed channel with low γ ($\gamma \lessapprox 30^o$), we observe that the lower SNR, the higher RPA_1. The reverse is observed for high values of γ. In conclusion, the power allocation strategy of LCOPT precoder is completely different from max-$d_{\min}$.

The proposed LCOPT precoding algorithm not only reduces the computational complexity of the conventional GOPT algorithm (as apparently observed from Algorithms 1 and 2) but also achieves a fast convergence while avoiding the initial values (ω_{pq}, ν_{pq}). Indeed, without loss of generality, we select a channel with $\gamma = 17.5^o$, which is close to the expectation of γ for 2×2 MIMO systems. Hence, $\mathbf{H}_v = [\cos(17.5^o) \quad 0\ ;\ 0 \quad \sin(17.5^o)]$. FIGURE 4.2 shows the convergence trajectories of GOPT (Algorithm 1) with different initial values of (ω_{pq}, ν_{pq}) and LCOPT (Algorithm 2, initial values are not required) over this channel at SNR = 12 dB, 4-QAM modulation. With $b = 2$ and 4-QAM, LCOPT

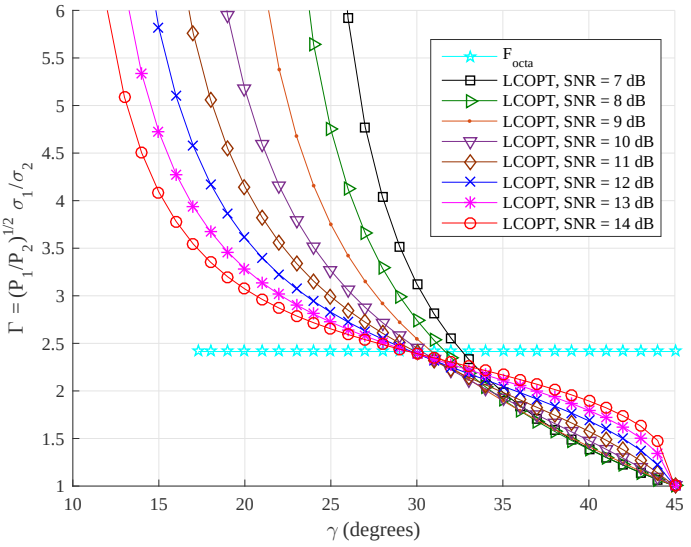


FIGURE 4.1: Scaling factor between the received constellations of the first substream over the second substream.

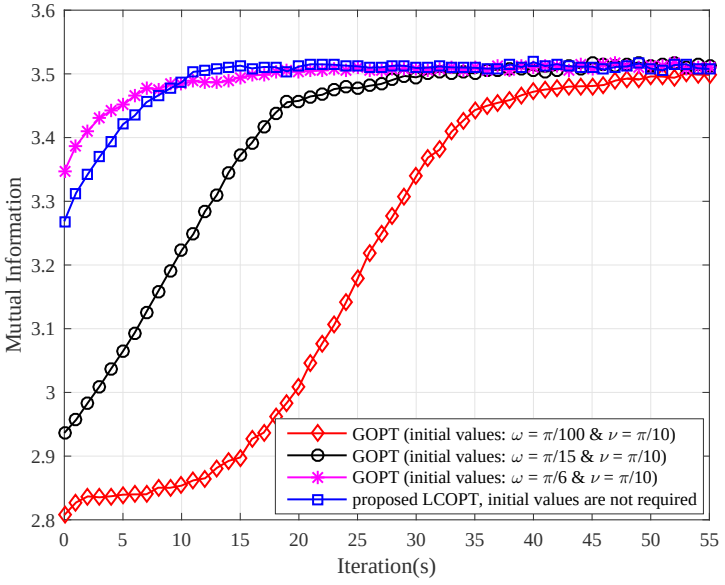


FIGURE 4.2: Convergence trajectories for mutual information, $b = 2$, 4-QAM, channel $\gamma = 17.5^o$ and SNR = 12 dB.

uses $\mathbf{\Theta} = \mathbf{\Theta}_{\mathbf{F}_{octa}}$ (see (4.13)). It is observed from FIGURE 4.2 that, for each channel and SNR value, the conventional GOPT precoding algorithm requires suitable initial values (ω_{pq}, ν_{pq}) for fast convergence. In contrast, the proposed LCOPT algorithm achieves a fast convergence speed without initial values requirement.

4.3.3 Selection of input $\boldsymbol{\Theta}$ for Algorithm 2 in case $b > 2$

In this subsection, we propose a suboptimal solution for $\boldsymbol{\Theta}$ in case $b > 2$. The idea is to find $\boldsymbol{\Theta}$ that improves the channel mutual information. The selection of $\boldsymbol{\Theta}$ is independent from $\boldsymbol{\Psi}$, *i.e.* in this subsection we suppose that $\boldsymbol{\Psi}$ is fixed. From (4.9), we deduce that the mutual information can be improved by increasing the squared distance $d^2 = \|\mathbf{H}_v\mathbf{F}_d(\mathbf{s}_m - \mathbf{s}_k)\|^2$ for a large number of difference vectors $\mathbf{e} = \mathbf{s}_m - \mathbf{s}_k$. The squared distance d^2 can be expressed by

$$d^2 = \mathbf{e}^\dagger \mathbf{W}\mathbf{e}, \tag{4.14}$$

$$= \sum_{i=1}^{b} \delta_i |e_i|^2 + 2 \sum_{i=1, j>i}^{b} \Re\{e_i^* e_j w_{ij}\}, \tag{4.15}$$

where w_{ij} stands for the (i,j) entry of $\mathbf{W} = \boldsymbol{\Theta}\boldsymbol{\Psi}^2\mathbf{H}_v^2\boldsymbol{\Theta}^\dagger$. The diagonal entries of $\mathbf{W}$ (w_{ii}) are denoted by δ_i for $i = 1, \ldots, b$.

On one hand, d^2 tends to minimum as soon as, for at least one index i, e_i attains its minimum value (*i.e.* when $|e_i| = |e_{\min}|$). Therefore, we need to focus on the case when, for at least one index i, $|e_i| = |e_{\min}|$ in order to maximize d^2. On the other hand, it comes from (4.15) that the dominant term of d^2 is $\sum_{i=1}^{b} \delta_i |e_i|^2$. This term can be maximized thanks to $\boldsymbol{\Theta}$ since $\delta_i > 0$ ($\mathbf{W}$ is an hermitian matrix). By considering only the dominant terms (*i.e.* first term in (4.15)), the derivation of the inner summation over k in (4.9) leads to the optimization problem for finding $\boldsymbol{\delta} = [\delta_1, \ldots, \delta_b]$, which can be expressed as follows

$$\begin{aligned} &\boldsymbol{\delta} = \arg\min_{\boldsymbol{\delta}} \left\{ \sum_{i=1}^{b} n_i e^{-\delta_i |e_{\min}|^2 / 2\sigma_\eta^2} \right\}, \\ &\text{subject to } \sum_{i=1}^{b} \delta_i = \mathrm{Tr}\{\mathbf{W}\} = P_w, \end{aligned} \tag{4.16}$$

where n_i is the number of e such that $|e_i| = |e_{\min}|$ on the i^{th} entry of $\mathbf{e} = [e_1, \ldots, e_i, \ldots, e_b]^T$. As all the streams use the same modulation, it comes that $n_i = n_j, \forall i, j$. Therefore, $\boldsymbol{\delta} = \arg\min_{\boldsymbol{\delta}} \left\{ \sum_{i=1}^{b} e^{-\delta_i |e_{\min}|^2 / 2\sigma_\eta^2} \right\}$.

Lemma 4. With $c = |e_{\min}|^2 / 2\sigma_\eta^2$ and $\sum_{\ell=1}^{b} \delta_\ell = P_w$, the following inequality holds

$$\frac{1}{b} \sum_{i=1}^{b} e^{-c\delta_i} \geq e^{-c\bar{\delta}}, \tag{4.17}$$

with $\bar{\delta} = \frac{P_w}{b}$. The equality occurs if $\delta_i = \bar{\delta}, \forall i$.

Proof. The convexity of exponential function yields $e^{\sum_i \kappa_i x_i} \leq \sum_i \kappa_i e^{x_i}$ with $0 < \kappa_i < 1$ and $\sum_i \kappa_i = 1$. Therefore, by applying this property with $\kappa_i = \frac{1}{b}$ and $x_i = -c\delta_i$, the lemma is proved. □

Let us denote by β_ℓ the diagonal elements of $\mathbf{\Psi}^2\mathbf{H}_v^2$. Then $P_w = \mathrm{Tr}\{\mathbf{W}\} = \sum_{\ell=1}^{b} \beta_\ell$ as $\mathbf{\Theta}$ is unitary. According to Lemma 4, the solution to (4.16) is $\delta_i = \bar{\delta} = \frac{1}{b}\sum_{\ell=1}^{b} \beta_\ell, \forall i$. On the other hand $\delta_i = \sum_{j=1}^{b} \beta_j |\Theta_{i,j}|^2$. We thus deduce that $|\Theta_{i,j}|^2 = \frac{1}{b}, \forall i, j$. The discrete Fourier transform (DFT) matrix defined by $\mathbf{\Theta}_{DFT} = \frac{1}{\sqrt{b}}\left[e^{-i2\pi kl/b}\right]_{k,l=0,\ldots,b-1}$ is a unitary matrix that satisfies that condition. Therefore in case $b > 2$, we propose to use $\mathbf{\Theta}_{DFT}$ as input of Algorithm 2. Note that, by using a different criterion (maximizing $d_{\min}$ instead of maximizing the channel mutual information), the authors in [21] also proposed $\mathbf{\Theta}_{DFT}$ for the generalization of max-$d_{\min}$ as presented in Section 2.3.2.2.

Let us consider a 3×3 channel, which is characterized by $(\gamma_1 = 40^o, \gamma_2 = 27^o)$ as shown in (2.29), to illustrate the convergence of the suboptimal LCOPT precoder compared to the optimal GOPT precoder in case $b > 2$. The convergence trajectories of the mutual information of LCOPT and GOPT over this channel, at SNR = 9.77 dB (equivalent to 5 dB in [22][1]), are presented in FIGURE 4.3. We observe that the GOPT precoder needs good initial pair (ω, ν) for fast convergence. For LCOPT, beside of $\mathbf{\Theta}_{DFT}$, we also take into account the $\mathbf{\Theta}$ of max-$d_{\min}$ solution for $b = 3$ ($\mathbf{\Theta}_{\text{max-}d_{\min}}$), which has been presented in Section 2.3.2.2. It is observed that the convergence speeds of LCOPT using both of the $\mathbf{\Theta}_{DFT}$ and $\mathbf{\Theta}_{\text{max-}d_{\min}}$ are faster than GOPT.

4.3.4 LCOPT precoder and MIMO symbol mapper association

In this subsection, we exploit the MIMO symbol mapper introduced in Section 3.4 (see FIGURE 3.6) and apply the MSEW mapping on the received constellation. Let us recall that the channel matrix $\mathbf{H}_v$ and the power allocation matrix $\mathbf{\Psi}$ only scale the amplitude of the constellations between subchannels. However, the matrix $\mathbf{\Theta}$ completely changes the form of the received constellation. Therefore, in order to apply MSEW mapping, the search must be done for each matrix $\mathbf{\Theta}$. For example, in Section 3.4, we presented two different mapping strategies for two precoding forms $\mathbf{F}_{r1}$ and $\mathbf{F}_{octa}$ of max-$d_{\min}$ precoder,

[1] in [22], the SNR is normalized by n_T. Therefore, the corresponding SNR reads SNR $= 5 + 10\log_{10}(3) \simeq 9.77$ dB.

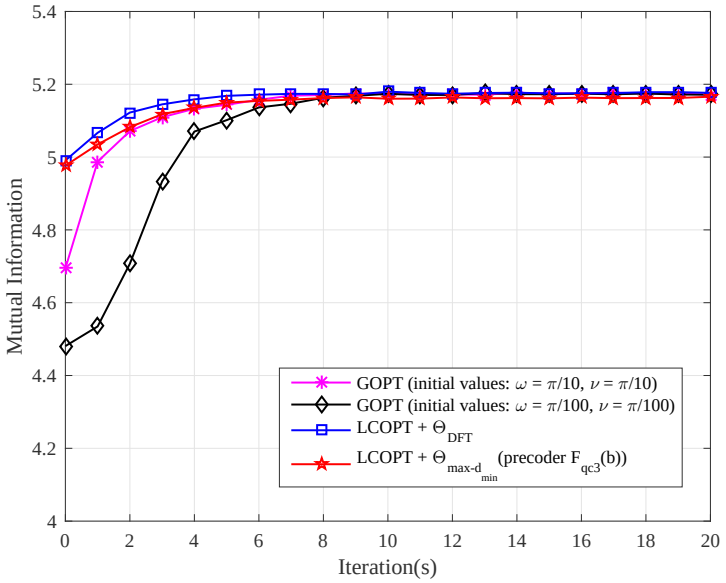


FIGURE 4.3: Convergence trajectories for mutual information, $b = 3$, 4-QAM and SNR = 9.77 dB.

which respectively have $\mathbf{\Theta}_{\mathbf{F}_{r1}}$ and $\mathbf{\Theta}_{\mathbf{F}_{octa}}$. Unfortunately, the GOPT precoder [22] has different $\mathbf{\Theta}$ for each channel and SNR. Hence, in order to use the MSEW mapping with the GOPT precoder, we must search for the optimized mapping for each channel realization at each SNR. Thus, it is impractical to apply the MSEW mapping on the received constellation of the GOPT precoder. However, thanks to the proposed LCOPT precoder, we can fix the matrix $\mathbf{\Theta}$ and then find the best MSEW mapping for the corresponding received constellation form. The simulation results in Section 4.4 show significant error-rate performance improvement of LCOPT used with MSEW mapping.

4.4 Simulations

4.4.1 Codebook construction for GOPT

The main drawback of the GOPT [22] precoder is that its optimum definition depends on the channel realization and on the SNR. Therefore, applying this precoder over random channels involves high computation time. In order to compare the proposed LCOPT precoder with the GOPT precoder over random channels, we propose in this subsection a simple method to construct a precoding codebook based on GOPT precoder in case $b = 2$ data streams.

Let us recall from (3.2) that, by defining the instantaneous received SNR as SNR = $\rho^2\frac{\sigma_s^2}{\sigma_\eta^2}$, the virtual channel $\mathbf{H}_v$ can be only characterized by angle γ. We then rely on γ and SNR to build a precoding codebook for each MIMO configuration. Without loss of generality, let us pick up an example to demonstrate the codebook construction method. We consider $(n_T = 2, n_R = 2)$ MIMO systems, which satisfy the condition $b \leq \min(n_T, n_R)$. Thanks to the cumulative distribution function (CDF), we firstly split γ into N ranges, which are equal in probability. For example, we split γ into $N = 10$ ranges, which are separated by the thresholds from γ_1 to γ_9, and each range corresponds to 10% of distribution. Secondly, we select $\gamma = \xi_i$, which is the mean of the distribution on each range $[\gamma_{i-1} \quad \gamma_i]_{i\in\{1,\ldots,N\}}$ (*i.e.* $Pr(\gamma \in [\gamma_{i-1}, \xi_i]) = Pr(\gamma \in [\xi_i, \gamma_i])$, $\gamma_0 = 0$ and $\gamma_N = 45^o$), to represent any γ falling in the range. The Algorithm 1 is then applied to find the optimal precoding matrix $\mathbf{F}_{\xi_i}$ for the channel with $\gamma = \xi_i$. Finally, any channel that has angle $\gamma_{i-1} \leq \gamma < \gamma_i$ will be associated with the precoding matrix $\mathbf{F}_d = \mathbf{F}_{\xi_i}$. The process is repeated for all SNRs of interest to construct the codebook.

4.4.2 Simulation scheme

To illustrate the benefit of LCOPT precoder, we test the MIMO system with the same system model considered in Chapter 3. The transceiver structure is depicted in FIGURE 3.1. In order to compare with the max-$d_{\min}$ precoder presented in Section 4.3.2 as well as to avoid complexity for the codebook construction of GOPT [22] precoder, we consider $b = 2$ data streams and 4-QAM modulation ($M = 4$) for the simulation. The optimized MSEW mappings for $\mathbf{F}_{r1}$ and $\mathbf{F}_{octa}$ provided in TABLE 3.1 are considered. Note that, in general, the LCOPT precoder may use input $\boldsymbol{\Theta} = \boldsymbol{\Theta}_{\mathbf{F}_{r1}}$ or $\boldsymbol{\Theta} = \boldsymbol{\Theta}_{\mathbf{F}_{octa}}$. However, for the SNR range considered in this chapter, the proposed LCOPT precoder uses input $\boldsymbol{\Theta} = \boldsymbol{\Theta}_{\mathbf{F}_{octa}}$ only (*i.e.* the constellation forms of LCOPT and $\mathbf{F}_{octa}$ are similar in this case). Therefore, the optimized MSEW mapping for LCOPT precoder is the same as the one of $\mathbf{F}_{octa}$. Thus LCOPT and max-$d_{\min}$ precoders only differ by their power allocation matrix $\boldsymbol{\Psi}$.

4.4.3 Simulation results

In this subsection, we illustrate results for the case $b = 2$. A randomly generated 2×2 MIMO channel is considered for the Monte-Carlo simulation, *i.e.* each element of $\mathbf{H}$ is

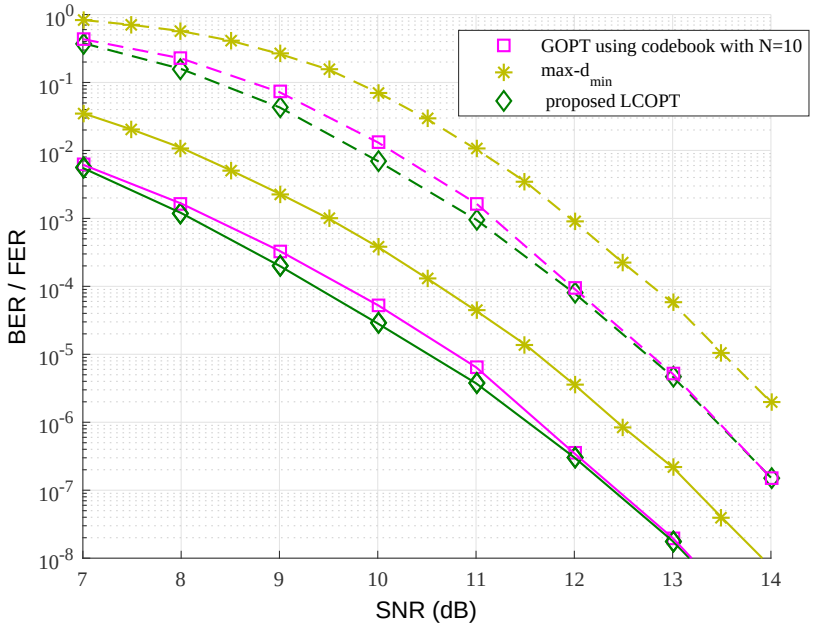


FIGURE 4.4: FER (dashed-lines) and BER (solid-lines) performances, all precoders are used with Gray-M mapping.

distributed as $\mathrm{H}_{i,j} \sim \mathcal{CN}(0,1)$. The equivalent channel $\mathbf{H}_v$ is then computed from $\mathbf{H}$ by the transformation presented in Section 2.2. The frame length is set to 800 uncoded bits and its components are interleaved by a random interleaver. The half-rate $(13,15)_{octal}$-RSC code is used as FEC encoder.

We firstly consider the conventional Gray-M mapping. FIGURE 4.4 shows frame-error rate (FER) and bit-error rate (BER) performances of the system in FIGURE 3.1 when the precoders are used with Gray-M mapping. The GOPT precoder is simulated by using a predefined codebook, which is constructed by using the method proposed in Section 4.4.1 with the codebook resolution $N = 10$. We observe that the proposed low complexity sub-optimal LCOPT precoder has a similar error-rate performance compared to the GOPT precoder. Note that the LCOPT slightly outperforms the GOPT in this case because GOPT relies on the codebook. The LCOPT precoder also outperforms max-$d_{\min}$ precoder in terms of both FER and BER performances. For example, at BER $= 10^{-6}$ and FER $= 10^{-4}$, the LCOPT precoder respectively achieves the gains of roughly 1 dB and 0.8 dB compared to max-$d_{\min}$ precoder.

Secondly, we apply, at the MIMO symbol mapper, the optimized MSEW mappings of max-$d_{\min}$ and LCOPT precoders. The simulation results are shown in FIGURE 4.5. It is observed that, assuming MSEW mappings, the LCOPT precoder achieves a gain of more

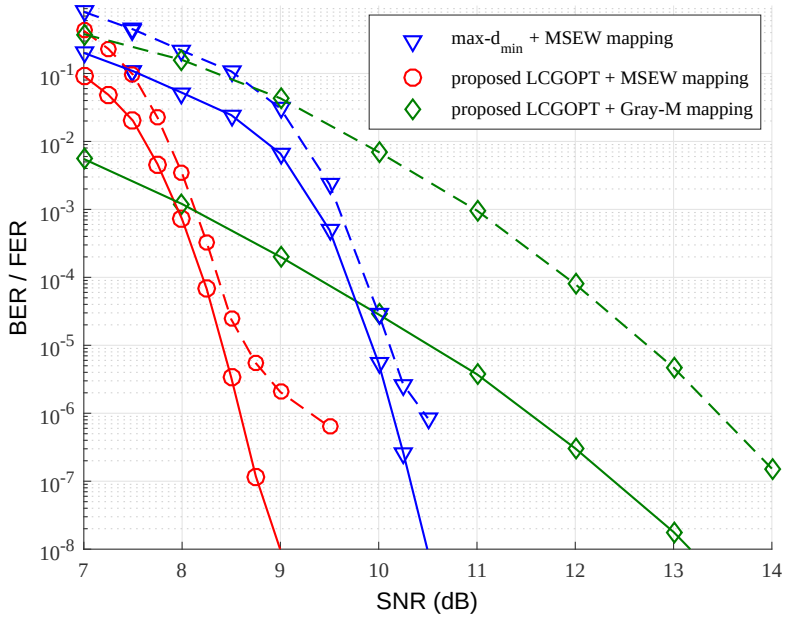


FIGURE 4.5: FER (dashed-lines) and BER (solid-lines) performances, max-$d_{\min}$ and the proposed LCOPT precoders are used with optimized MSEW mappings, GOPT precoder is used with Gray-M mapping.

than 1.5 dB at BER = 10^{-6} and roughly 1.5 dB at FER = 10^{-5} compared to max-$d_{\min}$ precoder. In FIGURE 4.5, we present again the result of Gray-M-mapped LCOPT precoder from FIGURE 4.4 to show the important role of MSEW mapping in error-rate performance. Indeed, by using MSEW mapping, we observe that the proposed LCOPT precoder significantly improves its error-rate performance compared to the case when Gray-M mapping is used. The LCOPT precoder used with MSEW mapping achieves a gain of roughly 4.1 dB at BER = 10^{-8} and more than 4 dB at FER = 10^{-6} compared to the LCOPT precoder used with Gray-M mapping. We would like to point out that, as mentioned in Section 4.4.2, searching for the optimized MSEW mappings of GOPT precoder is impractical since the MSEW search must be done for each channel realization and SNR value. However, with a fixed constellation form (fixed $\boldsymbol{\Theta}$), the MSEW mapping for LCOPT precoder is fixed for all channel realizations and SNRs. This shows the advantage of the proposed LCOPT precoder compared to the GOPT precoder.

We resort to extrinsic information (EXIT) chart [57] to account for the saturation in terms of FER in the high SNR region in FIGURE 4.5. The EXIT chart also helps to explain the gain of LCOPT used with Gray-M mapping compared to LCOPT used with MSEW mapping at the SNR region before the turbo-cliff. Introduction to EXIT chart can be found in Section 1.3.

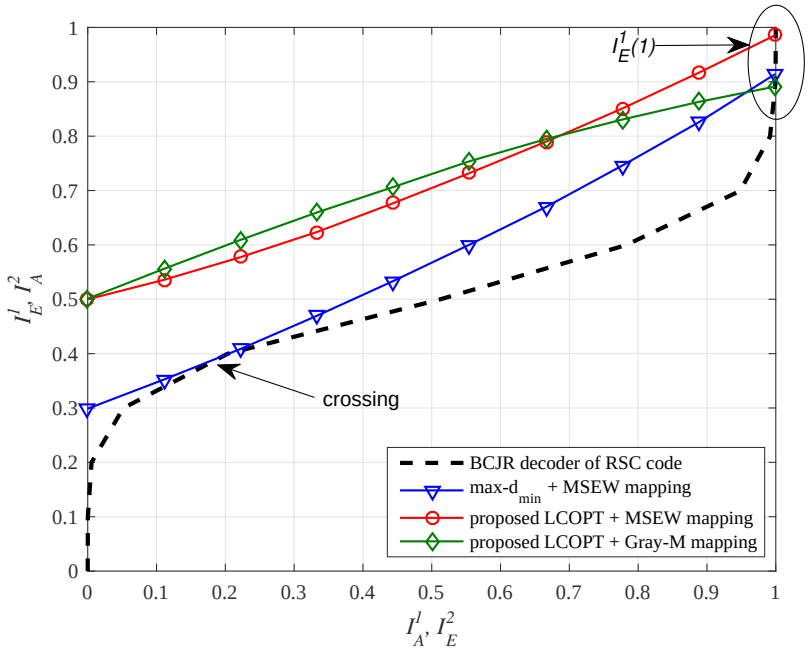


FIGURE 4.6: EXIT chart for channel $\gamma = 17.5^o$ at SNR = 9 dB.

We exploit the channel $\gamma = 17.5^o$, which has been used in FIGURE 4.2, to plot EXIT chart at SNR = 9 dB. The EXIT chart is shown in FIGURE 4.6. The dashed line represents the EXIT function of decoder, while the solid lines are the EXIT functions of demapper when different precoders and mappings are used. On one hand, we observe that, at SNR = 9 dB, the EXIT function of LCOPT used with MSEW mapping has a higher ending point $I_E^1(1)$ compared to the other solutions. This validates the performance gain of LCOPT used with MSEW at this SNR (see FIGURE 4.5). Additionally, in case LCOPT used with MSEW, since there is no intersection between the two EXIT functions, the iterative receiver converges easily. Hence, in this case, the erroneous frames can result from a few of erroneous bits. This accounts for the saturation in terms of FER (or, in other words, the FER curve has reached its lower bound, which increases the gap between FER and BER) as shown in FIGURE 4.5. On the other hand, at SNR = 9 dB, though the EXIT function of max-$d_{\min}$ is used with MSEW has a higher $I_E^1(1)$ compared to LCOPT used with Gray-M, the EXIT tunnel is however closed (early-crossing between two EXIT functions). Thus, the error-rate performance of LCOPT used with Gray-M is better than max-$d_{\min}$ used with MSEW at this SNR. Note that, at higher SNR, the EXIT tunnel in case max-$d_{\min}$ used with MSEW is wider. Hence, it avoids the early crossing and converges to a higher $I_E^1(1)$ (*i.e.* achieves a better error-rate performance) compared to LCOPT used with Gray-M. This is in accordance with the simulation shown

in FIGURE 4.5.

We conclude from previous simulation results that among the precoders that we proposed so far, whose designs took into account the outer FEC encoder and a turbo detection at the receiver, the LCOPT performs the best in terms of error-rate. To complete our study, we illustrate in FIGURE 4.7 the influence of the FEC error-correction capability on the performance. The encoded sequence length is 1600 bits. We consider three RSC codes differing by their memory order and thus free distance (d_f) and the corresponding number of error events (e_{ev}) with d_f. The d_f and e_{ev} are ($d_f = 5, e_{ev} = 1$), ($d_f = 6, e_{ev} = 2$) and ($d_f = 6, e_{ev} = 1$) for the memory order-2 $(7,5)_{octal}$, the memory order-3 $(13,15)_{octal}$ and the memory order-4 $(23,37)_{octal}$ RSC codes respectively. We observe that the error-floor is all the lower as the memory order is high. This is accounted for by the error probability whose dominant term at high SNR corresponds to error events defined by a decided sequence located at free distance from the original sequence. On the other hand we notice that the water-fall happens in a SNR range all the more shifted to the left as the memory-order is low. The EXIT chart accounts for it. Indeed the higher the free distance, the flatter the EXIT function of the RSC code. Given an SNR value, the EXIT function of the soft demapper is fixed and if the SNR value is low, the intersection of both curves is all the more probable as the RSC EXIT function is flat. By using the LCOPT, the error-floor happens at BER values lower than 10^{-6}, beyond the BER range of interest with respect to the targeted applications. As a conclusion, for rather short frames, we recommend to use the LCOPT precoder with low memory-order RSC codes (yielding an additional advantage in terms of computational complexity).

4.5 Conclusion

In this chapter, we have proposed a novel low complexity optimized (LCOPT) precoding algorithm to overcome the complexity of the globally optimized (GOPT) precoding algorithm introduced in [22]. The proposed LCOPT algorithm uses a fixed unitary matrix $\mathbf{\Theta}$ taken from the solution of maximizing $d_{\min}$. The optimization of the power allocation $\mathbf{\Psi}$ is then carried out by using the BTLS algorithm. The LCOPT algorithm has a fast convergence and avoids the initial selection of (ω_{pq}, ν_{pq}), which has vital impact on the convergence speed of GOPT algorithm. Assuming mappings on the received constellation, searching for the optimized MSEW mapping of GOPT precoder is intractable,

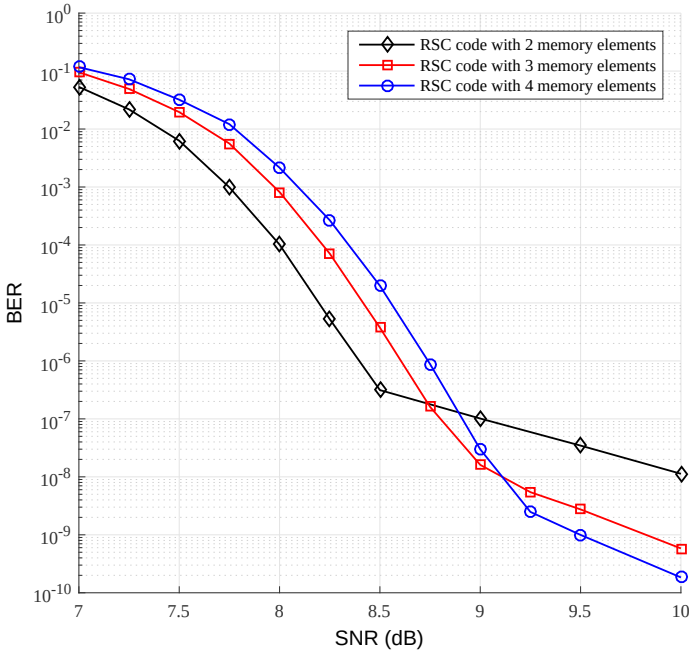


FIGURE 4.7: BER performances, the proposed LCOPT precoders are used with optimized MSEW mappings, $b = 2$, 4-QAM modulation.

since the constellation form is changed by each channel realization and SNR. However, this drawback is easily solved by the proposed LCOPT, whose received constellation form is fixed with $\boldsymbol{\Theta}$. In order to compare with the computational time-consuming algorithm GOPT over random channels, we performed simulations for the case $b = 2$ data streams and 4-QAM modulation. We have proposed a method to construct precoding codebook for GOPT precoder to avoid the search of optimal GOPT precoding matrix for each channel realization and SNR. Simulations show similar performances of the LCOPT and GOPT. Both outperform the max-$d_{\min}$ in case Gray-M mapping is used. In addition, the results show significant error-rate performance improvement of LCOPT precoder thanks to MSEW mapping. It is also shown that, assuming MSEW mapping, the LCOPT precoder significantly outperforms max-$d_{\min}$ precoder. Finally, EXIT chart analysis was provided to validate the simulation. As future work, the error-rate comparison between LCOPT and the generalized max-$d_{\min}$ (DFT-max-$d_{\min}$) precoder [21] is needed. In addition, we propose to test LCOPT precoder with different encoders and iterative receivers. It would be also interesting to optimize the choice of $\boldsymbol{\Theta}$ for $b > 2$ and extend the constellation size per data stream.

In the next chapter, we focus our study on precoder optimization assuming turbo equalization at the receiver, which has lower complexity compared to turbo detection. With turbo equalization, the received symbols are initially processed by an interference canceller, where they are decomposed into parallel sub-streams, before being converted into soft-messages and entering the decoder. Therefore, the mapping on received constellation is not essential for the turbo equalization scheme.

Appendices of chapter 4

A Proof of Proposition 1

Let us consider $\mathbf{F}_d = \mathbf{U}_\mathbf{F}\mathbf{\Sigma}_\mathbf{F}\mathbf{V}_\mathbf{F}^\dagger$ ($\|\mathbf{F}_d\|_F^2 = 1$). The channel mutual information directly depends on the hermitian matrix $\mathbf{W} = \mathbf{F}_d^\dagger\mathbf{H}_v^\dagger\mathbf{H}_v\mathbf{F}_d$. The eigen-decomposition of $\mathbf{W}$ yields

$$\mathbf{W} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\dagger, \tag{5.18}$$

where $\mathbf{Q}$ is a unitary matrix and $\mathbf{\Lambda}$ is a diagonal matrix with real positive values. One can always find a diagonal matrix $\tilde{\mathbf{\Sigma}}_\mathbf{F}$ with real positive values such that $\mathbf{\Lambda} = \mathbf{H}_v^2\tilde{\mathbf{\Sigma}}_\mathbf{F}^2$. Then, the hermitian matrix $\mathbf{W}$ reads

$$\mathbf{W} = \mathbf{Q}\tilde{\mathbf{\Sigma}}_\mathbf{F}\mathbf{H}_v^2\tilde{\mathbf{\Sigma}}_\mathbf{F}\mathbf{Q}^\dagger. \tag{5.19}$$

Thus, the new precoding matrix $\mathbf{F}'_d = \tilde{\mathbf{\Sigma}}_\mathbf{F}\mathbf{Q}^\dagger$ yields the same mutual information as the one obtained when the precoding matrix $\mathbf{F}_d = \mathbf{U}_\mathbf{F}\mathbf{\Sigma}_\mathbf{F}\mathbf{V}_\mathbf{F}^\dagger$ is applied. The matrix trace of $\mathbf{W}$ reads

$$\mathrm{Tr}\{\mathbf{W}\} = \mathrm{Tr}\{\mathbf{H}_v^2\tilde{\mathbf{\Sigma}}_\mathbf{F}^2\} = \mathrm{Tr}\{\mathbf{\Sigma}_\mathbf{F}\mathbf{U}_\mathbf{F}^\dagger\mathbf{H}_v^2\mathbf{U}_\mathbf{F}\mathbf{\Sigma}_\mathbf{F}\} \leq \mathrm{Tr}\{\mathbf{H}_v^2{\mathbf{\Sigma}_\mathbf{F}}^2\}. \tag{5.20}$$

Note that the equality holds only for a unitary matrix $\mathbf{U}_\mathbf{F}$ equal to the identity matrix $\mathbf{I}_b$. The proof of the inequality in (5.20) can be found in [69]. Let us denote by σ_i and $\tilde{\sigma}_i$ the diagonal elements of $\mathbf{\Sigma}_\mathbf{F}$ and $\tilde{\mathbf{\Sigma}}_\mathbf{F}$ respectively. From (5.20), we deduce that $\tilde{\sigma}_i^2 \leq \sigma_i^2$ for $i = 1, \ldots, b$. Hence, the power constraint of $\mathbf{F}'_d$ satisfies the following inequality

$$\|\mathbf{F}'_d\|_F^2 = \mathrm{Tr}\{\tilde{\mathbf{\Sigma}}_\mathbf{F}^2\} \leq \mathrm{Tr}\{{\mathbf{\Sigma}_\mathbf{F}}^2\} = \|\mathbf{F}_d\|_F^2 = 1. \tag{5.21}$$

In order to keep the unity power constraint, the new precoding matrix must be normalized as $\mathbf{F}''_d = \alpha \mathbf{F}'_d$ where $\alpha = 1/\|\mathbf{F}'_d\|_F^2$ with $\alpha \geq 1$ thanks to (5.21). Let us denote by $I(\mathbf{H}_v\mathbf{F}''_d)$, $I(\mathbf{H}_v\mathbf{F}'_d)$ and $I(\mathbf{H}_v\mathbf{F}_d)$ the mutual information obtained when $\mathbf{F}''_d$, $\mathbf{F}'_d$ and $\mathbf{F}_d$ are respectively applied. Note that $I(\mathbf{H}_v\mathbf{F}'_d) = I(\mathbf{H}_v\mathbf{F}_d)$. As $\alpha \geq 1$, the following expression holds

$$I(\mathbf{H}_v\mathbf{F}''_d) \geq I(\mathbf{H}_v\mathbf{F}'_d) = I(\mathbf{H}_v\mathbf{F}_d) \tag{5.22}$$

The equality holds only for $\mathbf{U}_\mathbf{F} = \mathbf{I}_b$, which yields $\mathbf{F}''_d = \mathbf{F}'_d = \mathbf{F}_d$ (*i.e.* $\alpha = 1$, $\mathbf{Q} = \mathbf{V}_\mathbf{F}$ and $\tilde{\mathbf{\Sigma}}_\mathbf{F} = \mathbf{\Sigma}_\mathbf{F}$).

B An example of solving $\delta\omega_{pq}$ and $\delta\nu_{pq}$ from $\delta_\mathbf{W}$ in case $b = 2$

With $b = 2$, from (4.6), the $b \times b$ unitary matrix $\mathbf{\Theta}$ reads

$$\mathbf{\Theta} = \begin{pmatrix} \cos\omega & \sin\omega\,(\cos\nu - j\sin\nu) \\ -\sin\omega\,(\cos\nu - j\sin\nu) & \cos\omega \end{pmatrix}. \tag{5.23}$$

From (4.7), calculation yields

$$\begin{aligned}[\delta_\mathbf{\Theta}] = &\begin{pmatrix} -\sin\omega & \cos\omega\,(\cos\nu - j\sin\nu) \\ -\cos\omega\,(\cos\nu + j\sin\nu) & -\sin\omega \end{pmatrix}\delta_\omega \\ &+\begin{pmatrix} 0 & -\sin\omega\,(\sin\nu + j\cos\nu) \\ \sin\omega\,(\sin\nu - j\cos\nu) & 0 \end{pmatrix}\delta_\nu.\end{aligned} \tag{5.24}$$

Let us define $\mathbf{H}_v^2\mathbf{\Psi}^2 = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}$ and denote $K = A_1 - A_2$. From (4.8), calculation yields:

$$\begin{aligned}\delta_\mathbf{W} = &\begin{pmatrix} -K\sin(2\omega) & K\cos(2\omega)\,(\cos\nu - j\sin\nu) \\ K\cos(2\omega)\,(\cos\nu + j\sin\nu) & K\sin(2\omega) \end{pmatrix}\delta_\omega \\ &+\begin{pmatrix} 0 & -\frac{K}{2}\sin(2\omega)\,(\sin\nu + j\cos\nu) \\ -\frac{K}{2}\sin(2\omega)\,(\sin\nu - j\cos\nu) & 0 \end{pmatrix}\delta_\nu,\end{aligned} \tag{5.25}$$

where $\delta_{\mathbf{W}} = \underbrace{\mu_{\mathbf{W}}\mathbf{\Phi}}_{\mathbf{B}}$. Let us define $\mathbf{B} = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$. In this case, B_{21} and B_{12} form a complex conjugate pair. Hence,

$$\begin{aligned} B_{21} + B_{12} &= 2\Re\{B_{21}\} \\ &= 2K\cos(2\omega)\cos\nu\delta\omega - K\sin(2\omega)\sin\nu\delta_\nu, \end{aligned} \tag{5.26}$$

and

$$\begin{aligned} |B_{21} - B_{12}| &= |j2\Im\{B_{21}\}| \\ &= |j\left(2K\cos(2\omega)\sin\nu\delta_\omega + K\sin(2\omega)\cos\nu\delta_\nu\right)|, \end{aligned} \tag{5.27}$$

where $\Re\{a\}$ and $\Im\{a\}$ are the real and the imaginary parts of a respectively. Finally, the four equations to solve for $(\delta_\omega, \delta_\nu)$ can be formulated as

$$\begin{pmatrix} -K\sin(2\omega) & 0 \\ K\sin(2\omega) & 0 \\ K\cos(2\omega)\cos\nu & -\frac{K}{2}\sin(2\omega)\sin\nu \\ K\cos(2\omega)\sin\nu & \frac{K}{2}\sin(2\omega)\cos\nu \end{pmatrix} \times \begin{pmatrix} \delta_\omega \\ \delta_\nu \end{pmatrix} = \begin{pmatrix} B_{11} \\ B_{22} \\ \Re\{B_{21}\} \\ \Im\{B_{21}\} \end{pmatrix}. \tag{5.28}$$

C Find mutual information in function of W

Let us rewrite

$$\begin{aligned} \zeta_{m,k} &= (\|\mathbf{H}_v\mathbf{F}_d\mathbf{e}_{m,k} + \boldsymbol{\eta}\|^2 - \|\boldsymbol{\eta}\|^2)/\sigma_\eta^2 \\ &= \left(\mathbf{e}_{m,k}^\dagger \underbrace{\mathbf{F}_d^\dagger\mathbf{H}_v^\dagger\mathbf{H}_v\mathbf{F}_d}_{\mathbf{W}}\mathbf{e}_{m,k} + \boldsymbol{\eta}^\dagger\mathbf{H}_v\mathbf{F}_d\mathbf{e}_{m,k} + \mathbf{e}_{m,k}^\dagger\mathbf{F}_d^\dagger\mathbf{H}_v^\dagger\boldsymbol{\eta}\right)/\sigma_\eta^2, \\ &= \frac{1}{\sigma_\eta^2}\mathbf{e}_{m,k}^\dagger\mathbf{W}\mathbf{e}_{m,k} + \Re\left\{\frac{2}{\sigma_\eta^2}\mathbf{e}_{m,k}^\dagger\underbrace{\mathbf{F}_d^\dagger\mathbf{H}_v^\dagger}_{R}\boldsymbol{\eta}\right\}. \end{aligned} \tag{5.29}$$

The singular value decomposition of $\mathbf{W}$ yields

$$\mathbf{W} = \mathbf{U}_\mathbf{W}\mathbf{\Sigma}_\mathbf{W}\mathbf{V}_\mathbf{W} = \underbrace{\mathbf{U}_\mathbf{W}\mathbf{\Sigma}_\mathbf{W}^{\frac{1}{2}}}_{R}\mathbf{\Sigma}_\mathbf{W}^{\frac{1}{2}}\mathbf{V}_\mathbf{W}. \tag{5.30}$$

Thus $R = \mathbf{U}_\mathbf{W}\mathbf{\Sigma}_\mathbf{W}^{\frac{1}{2}}$.

Chapter 5

Optimization of linear MIMO precoding assuming MMSE-based turbo equalization

The content of this chapter is mainly based on the following papers:

- Nhat-Quang Nhan, Philippe Rostaing, Karine Amis, Ludovic Collin, and Emanuel Radoi. "Optimization of linear MIMO precoding assuming MMSE-based turbo equalization". Submitted to *IEEE Transaction on Wireless Communications*, pending for review.
- Nhat-Quang Nhan, Philippe Rostaing, Karine Amis, Ludovic Collin, and Emanuel Radoi. "Association and joint optimization of max-dmin precoder with MIMO turbo equalization". In *IEEE Global Communications Conference (GLOBECOM)*, pages 1 - 6, 2015.

5.1 Introduction

The optimization of linear precoders for frequency domain minimum mean square error (MMSE) turbo equalization has been investigated in several papers [70–72]. However, a closed-form precoding matrix was not proposed. These designs relied on convex programming algorithms to find power allocation strategies that minimize transmission

power while maintaining acceptable error-rate performance. For time domain MMSE turbo equalization, the authors in [73] applied a Hadamard precoding matrix and optimized the equalization coefficients by assuming full channel state information (CSI) at the receiver only, instead of optimizing the precoding matrix by assuming full CSI at both transmitter and receiver.

In this chapter, non-frequency selective channels are still considered and the association of multiple-input multiple-output (MIMO) linear precoder and MMSE interference cancellation (IC) turbo equalizer in time domain duplex closed-loop schemes is investigated. We firstly propose a threshold adaptation for the max-$d_{\min}$ precoder in order to improve the system error-rate performance when it is used with time domain MMSE IC turbo equalization. The new precoder is referred to as max-$d_{\min}$-mod precoder. Secondly, we propose a novel MIMO precoder, which is targeted for low-complex outer forward error correction codes assuming time domain MMSE IC turbo equalization at the receiver. The proposed precoder is referred to as *Genie-optimized precoder*. Simulation results show that the Genie-optimized precoded scheme outperforms the other selected reference precoded schemes in terms of error-rate. The results in this chapter can be applied to communication systems that require low complexity, where low-complex FEC codes are used.

The remainder of this chapter is organized as follows. Section 5.2 introduces the system model along with main expressions of the low-complexity interference canceller, which takes into account the associated MIMO precoder. A short introduction to extrinsic information transfer (EXIT) chart for turbo equalization is also presented. In Section 5.3, by assuming that the two precoding forms ($\mathbf{F}_{r1}$ and $\mathbf{F}_{octa}$) of max-$d_{\min}$ precoder are used at the transmitter, we propose a dynamic switching threshold adaptation to replace for the conventional fixed switching threshold of max-$d_{\min}$ precoder. The precoder, which uses the proposed SNR-dependent dynamic threshold to switch between $\mathbf{F}_{r1}$ and $\mathbf{F}_{octa}$ precoding forms, is referred to as max-$d_{\min}$-mod precoder. Simulation shows that the max-$d_{\min}$-mod precoder significantly enhances the system error-rate performance over random channels. In Section 5.4, we propose a new precoder that is particularly well designed when turbo equalization is used at the receiver. Firstly, the optimization problem at the initial state and at the optimum convergence state of turbo equalization is defined, which solves the judicious selection of the three parameters involved in the parameterized form of precoding matrix. Consequently, the optimized parameters define

the novel Genie-optimized precoder introduced above. In Section 5.5, EXIT chart is used to support the analytical results. The EXIT chart comparison between existing precoders and the new one is also illustrated. In Section 5.6, simulated error rates are presented to validate the theoretical analysis. Section 5.7 concludes the chapter and gives some perspectives.

5.2 System model and preliminaries

5.2.1 System model

Let us consider a MIMO system with n_R receive, n_T transmit antennas and b independent data streams to be transmitted. We assume full-CSI at both the transmitter and the receiver. A binary recursive-systematic convolutional (RSC) code is used to encode the information data bits. The FEC codeword is then interleaved before being mapped to Q-ary quadrature amplitude modulation (QAM) symbols. The modulated symbols are converted into a b-dimensional symbol vector $\mathbf{s}$. The vector $\mathbf{s}$ is then precoded by a matrix $\mathbf{F}$ and transmitted through the MIMO channel. At the receiver side, the MMSE IC turbo equalization, which has been introduced in Section 1.2.3, is investigated.

According to (2.16), the channel output $\mathbf{y}$ reads $\mathbf{y} = \mathbf{H}_v\mathbf{F}_d\mathbf{s} + \boldsymbol{\eta}$. For ease of reading, let us recall in this chapter the parameterized form of $\mathbf{F}_d$, which has been introduced in (3.1), for $b = 2$ as follows

$$\mathbf{F}_d = \begin{pmatrix} \cos\psi & 0 \\ 0 & \sin\psi \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & e^{i\phi} \end{pmatrix}, \tag{5.1}$$

where ψ ($0^o \leq \psi \leq 90^o$) is linked to the power allocation on the eigen-subchannels, θ ($0^o \leq \theta \leq 90^o$) and ϕ ($0^o \leq \phi \leq 90^o$) allow us to respectively mix and rotate the symbols onto the two eigen-subchannels.

A complete system including the equivalent model and the turbo equalization structure is shown in FIGURE 5.1, where $L^1_{\mathrm{A}}, L^1_{\mathrm{P}}$ and L^1_{E} respectively stand for the *a priori*, the *a posteriori* and the extrinsic log likelihood ratios (LLRs) of the SBC, while the equivalent notations for the BCJR soft decoder are $L^2_{\mathrm{A}}, L^2_{\mathrm{P}}$ and L^2_{E}.

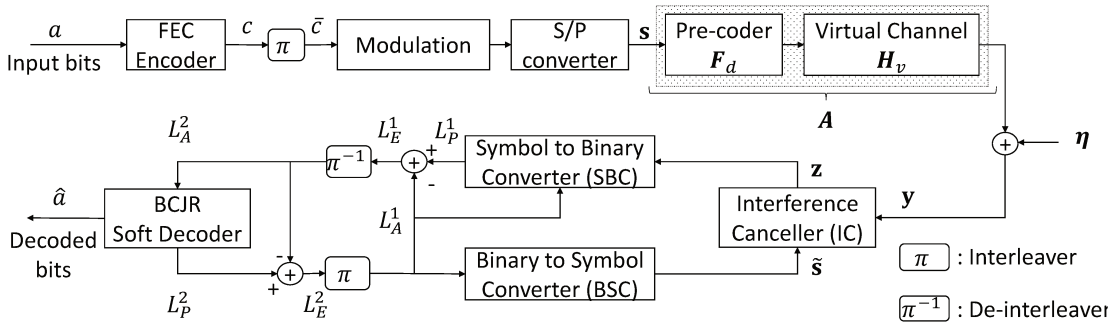


FIGURE 5.1: Precoded turbo equalization system.

5.2.2 MIMO precoder for turbo equalization

The turbo equalization principle as well as its structure are detailed in Section 1.2.3. We apply herein the derivation results from Section 1.2.3 while taking into account the precoder at the transmitter. Let us denote $\mathbf{A} = \mathbf{H}_v\mathbf{F}_d$. We then deduce from (1.30) that $\mathbf{A}$ is equivalent to a new channel matrix (see FIGURE 5.1). Hence, by applying the equivalent channel matrix $\mathbf{A}$ to (1.42) and (1.42), we respectively obtain

$$\mathbf{W}_{k,:} = \sigma_s^2 \mathbf{A}_{:,k}^{\dagger} \left(\mathbf{B} + \sigma_{\tilde{s}}^2 \mathbf{A}_{:,k} \mathbf{A}_{:,k}^{\dagger} \right)^{-1}, \tag{5.2}$$

and

$$\mathbf{Q}_{k,:} = \mathbf{W}_{k,:}\mathbf{A} - \mathbf{W}_{k,:}\mathbf{A}_{:,k} e_k, \tag{5.3}$$

where e_k is the k^{th} row of $\mathbf{I}_b$. $\mathbf{A}_{:,k}$ and $\mathbf{A}_{k,:}$ respectively denote the k^{th} column and k^{th} row of $\mathbf{A}$. We also recall from (1.48) that the IC output can be modeled as follows

$$z_k = \mu_k s_k + \xi_k, \tag{5.4}$$

where ξ_k is independent from s_k, has zero mean and variance $\sigma_{\xi_k}^2 = \sigma_s^2 \mu_k (1 - \mu_k)$. Note that, throughout this chapter, we refer to $\mathbf{z}$ as the IC output while $\mathbf{y}$ denotes the virtual channel output.

5.2.3 EXIT function of turbo equalizer

In this chapter, we use extrinsic information transfer (EXIT) chart, which has been introduced in Section 1.3, to analyze the convergence behavior of the turbo equalizer. In the particular case when turbo equalizer is used at the receiver, one component in

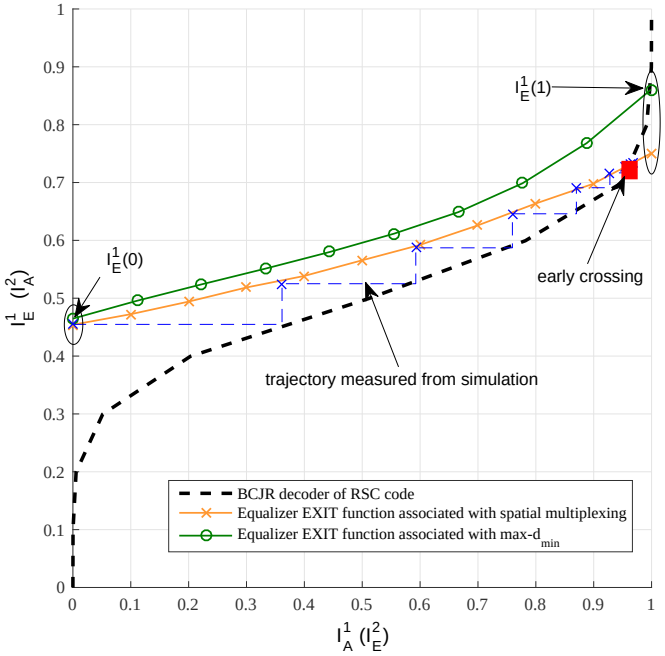


FIGURE 5.2: EXIT chart of turbo equalization vs. the real trajectory (obtained from simulation) for Channel A, SNR = 10 dB, 4-QAM and $(13, 15)$-RSC code.

the EXIT chart is the soft-input soft-output FEC decoder (with input L_A^1 and output L_E^1) and the other one contains the BSC and SBC revolved around the MMSE-based IC (with input L_A^2 and output L_E^2) as shown in FIGURE 5.1. The mutual information (MI) extraction as well as the log-likelihood ratio (LLR) generation are carried out according to the method described in Section 1.3. Parameters of the J-function are taken from [74]. We consider the half-rate RSC defined by its generator polynomials in octal form $(13, 15)$.

FIGURE 5.2 shows the EXIT charts of turbo equalization with spatial multiplexing and max-$d_{\min}$ precoder used at the transmitter. The EXIT charts are computed for Channel A (defined in Chapter 3 with $\gamma = 8.3^o$) at SNR = 10 dB. The trajectory is matching with the EXIT functions.

5.3 Association and Joint Optimization of max-$d_{\min}$ Precoder with MIMO Turbo Equalization

5.3.1 Analysis

Let us denote by $I_E^1(1)$ the extrinsic MI at output of the SBC at the convergence when $I_A^1 = 1$. As shown in FIGURE 5.2, the $I_E^1(1)$ of precoder max-$d_{\min}$ is higher than the

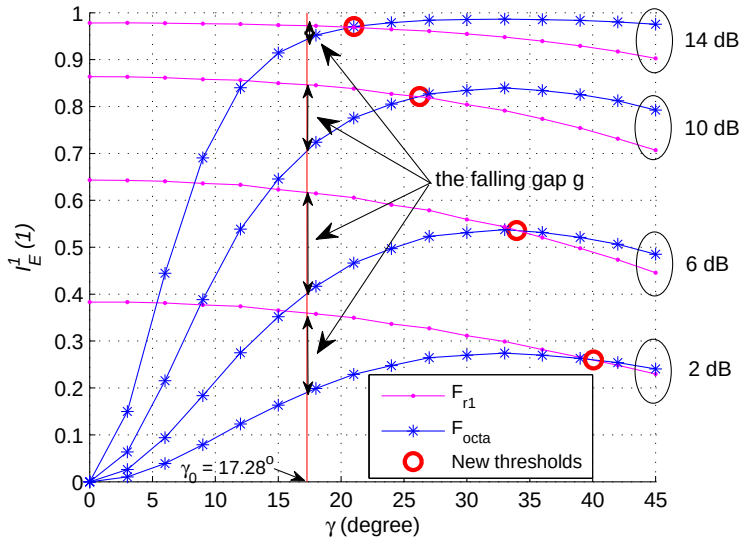


FIGURE 5.3: $I_E^1(1)$ versus γ at different SNR.

spatial multiplexing one. It predicts that the max-$d_{\min}$ precoded system has a lower error-floor [75] . In addition, the opening of tunnel between the two EXIT functions corresponding to max-$d_{\min}$ is wider, which means that the turbo equalizer with max-$d_{\min}$ converges faster than with the spatial multiplexing. Moreover, the SNR lower-bound, which avoids early crossing of EXIT chart, is lower with the opening of tunnel. These properties show the advantages of using precoder in association with the outer code and a turbo equalization applied at the receiver side.

Recall from Section 2.3.2.2 that the threshold γ_0, which is used to switch between $\mathbf{F}_{r1}$ and $\mathbf{F}_{octa}$ of the precoder max-$d_{\min}$, was selected so as to maximize the $d_{\min}$ of the received constellation points for the uncoded systems. In this subsection, we focus on the convergence state of the turbo equalization to show that performance corresponding to the max-$d_{\min}$ precoder can be further optimized by selecting a new threshold for each SNR such that $I_E^1(1)$ is maximized.

FIGURE 5.3 shows the plots in terms of $I_E^1(1)$ of both $\mathbf{F}_{r1}$ and $\mathbf{F}_{octa}$, for all values of γ, at each SNR. Note that $\rho^2 = 1$ in this case. The interest of this figure is threefold. First, it shows that, using the original threshold γ_0, there is a falling gap g between the $I_E^1(1)$ of $\mathbf{F}_{r1}$ and the one of $\mathbf{F}_{octa}$, *i.e.* the $I_E^1(1)$ at the $\gamma > \gamma_0$ is smaller than the counterpart, which reduces the performance. Therefore, we need to select a new threshold γ_{th} (see the bold circles in FIGURE 5.3) that takes into account the outer FEC code as well as turbo equalization assumption. Second, we found that γ_{th} is a function of SNR satisfying $\gamma_{th} > \gamma_0$. Third, we obtain that the falling gap g is very small at the very high SNR, *i.e.*

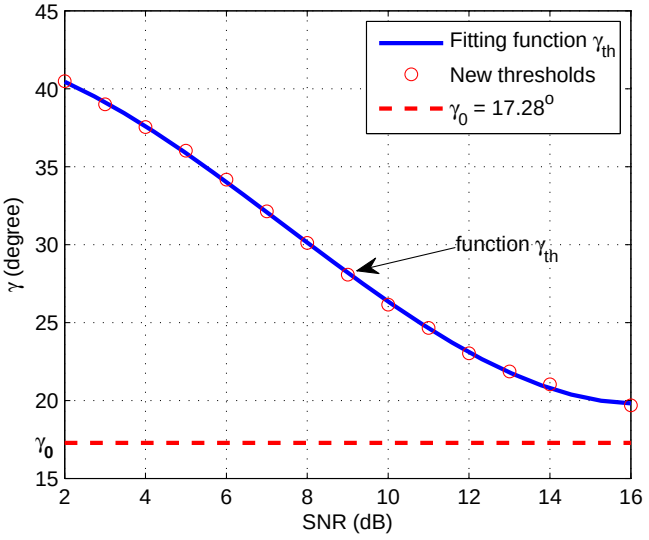


FIGURE 5.4: The new threshold γ_{th} in function of SNR.

the difference in terms of $I_E^1(1)$ between the original threshold γ_0 and the new one is not significant.

FIGURE 5.4 shows the fitting curve as a function of SNR, which is obtained by plotting the new thresholds defined in FIGURE 5.3 for many different SNRs and fitting the obtained values with the least-squares method [76, Chapter 6]. Similar to the falling gap g, the difference between γ_0 and the new γ_{th} is inversely proportional to the SNR. This is in accordance with the simulation results given in next subsection. The obtained fitting function $\gamma_{th}(x)$ is a cubic polynomial, which reads

$$\gamma_{th}(x) = \alpha_4 + \alpha_3 x + \alpha_2 x^2 + \alpha_1 x^3, \tag{5.5}$$

where x is SNR in dB, $\alpha_1 = 8.66524\times10^{-3}$, $\alpha_2 = -0.19457$, $\alpha_3 = -0.50131$ and $\alpha_4 = 42.15576$. The $\gamma_{th}(x)$ is measured in degree and the fitting is obtained for $x \in \{2,\ldots,16\}$. For the region $x > 16$, we fix $\gamma_{th} = 20^o$. The region $x < 2$ is not interesting due to the early crossing in the EXIT chart of the turbo equalization.

In summary, instead of using the static threshold γ_0 to switch between $\mathbf{F}_{r1}$ and $\mathbf{F}_{octa}$ as max-$d_{\min}$ does, we propose the new threshold γ_{th} for each SNR as presented in (5.5). The new precoder, that uses γ_{th}, is now referred to as max-$d_{\min}$-mod precoder.

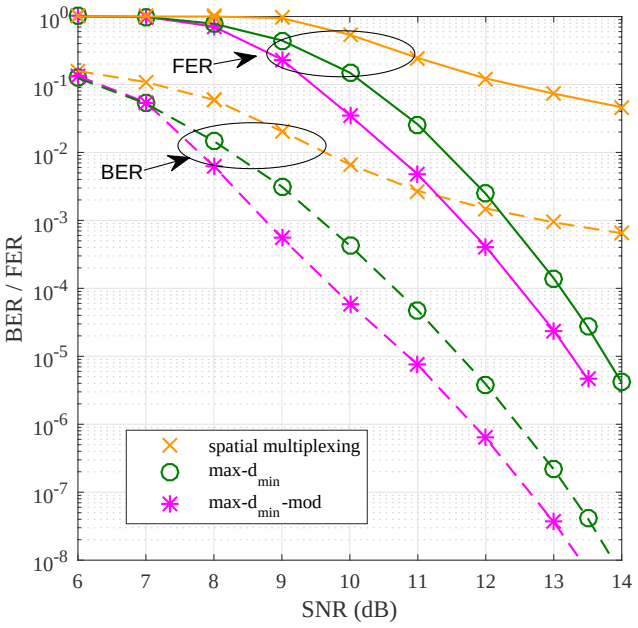


FIGURE 5.5: BER (dashed lines) and FER (solid lines) performance of the max-$d_{\min}$ and max-$d_{\min}$-mod precoded turbo equalization in a 2×2 MIMO system.

5.3.2 Results

In this subsection, we provide simulation results to illustrate the advantage of max-$d_{\min}$-mod compared to max-$d_{\min}$ over random channels. A randomly generated MIMO channel is considered for the Monte-Carlo simulation, *i.e.* each element of $\mathbf{H}$ is distributed as $\mathrm{H}_{i,j} \sim \mathcal{CN}(0,1)$. The half-rate $(13,15)_{octal}$-RSC code is used as FEC encoder. The frame length is set to 2000 uncoded bits and its components are interleaved by a random interleaver. We consider the instantaneous received SNR as defined in Chapter 3.

Let us consider a MIMO system with $n_T = 2$ transmit and $n_R = 2$ receive antennas (MIMO 2×2) configuration. FIGURE 5.5 shows the bit-error-rate (BER) performance of the turbo equalization when the spatial multiplexing, max-$d_{\min}$ and max-$d_{\min}$-mod precoders are used at the transmitter side. We observe that, in case of an outer FEC with turbo equalization at the receiver, the MIMO precoder improves the system performance. More precisely, the max-$d_{\min}$ precoder achieves a gain of 1.5 dB at BER = 10^{-2} and of roughly 3.5 dB at BER = 10^{-3} compared to the spatial multiplexing. The gain is even larger in the high SNR region, which is in accordance with the analysis in FIGURE 5.2, since the MI at the convergence state, $I_E^1(1)$, of max-$d_{\min}$ is higher than the one of spatial multiplexing.

Moreover, it is shown from FIGURE 5.5 that, by using the proposed threshold γ_{th}, the max-d_{min}-mod precoder respectively achieves a gain of roughly 0.8 dB and of 1 dB at BER = 10^{-3} and BER = 10^{-4} compared to max-d_{min} precoder. In addition, the performance of max-d_{min} and max-d_{min}-mod precoders are close to each other at the very high SNR (*e.g.* max-d_{min}-mod achieves a gain of roughly 0.5 dB at BER = 10^{-7} compared to max-d_{min}). This confirms the conclusion drawn from Section 5.3.1 that, at the very high SNR, γ_{th} is close to γ_0. Similar observations are obtained in terms of frame-error-rate (FER).

We have proposed in this section the new threshold γ_{th} to maximize $I_E^1(1)$ by using the two fixed forms of max-d_{min}. In the next subsection, we will propose a completely new form of the precoding matrix $\mathbf{F}_d$. The new precoder aims to maximize the mutual information between channel input $\mathbf{s}$ and the symbol $\mathbf{y}$ at output of the equalizer. To this end, analytical results from the next section show that we need to maximize $I_E^1(1)$ and $I_E^1(0)$ respectively. As the maximization of $I_E^1(1)$ is substantial, it further accounts for the significant advantage of the max-d_{min}-mod compared to the max-d_{min} precoder as shown in this section.

5.4 Optimization of the defining precoder parameters: Genie-optimized precoder

5.4.1 Problem statement

As shown in FIGURE 5.1, the input LLRs fed to the FEC decoder are calculated from the output $\mathbf{z}$ of the interference canceller. Therefore, the MI $I(\mathbf{z},\mathbf{s})$ plays an essential role on the system error-rate performance. In addition, we obtain directly from (1.40) that the MI between channel output $\mathbf{y}$ and the corresponding input $\mathbf{s}$ (channel capacity) is different from $I(\mathbf{z},\mathbf{s})$, *i.e.* $I(\mathbf{y},\mathbf{s}) \neq I(\mathbf{z},\mathbf{s})$. This relationship motivates our quest for a precoder that maximizes $I(\mathbf{z},\mathbf{s})$, since GOPT precoder (see Section 2.3.2.1), which is the globally optimal precoder in the literature, maximizes the channel capacity $I(\mathbf{y},\mathbf{s})$.

Since $\mathbf{z}$ takes into account the *a priori* information ($\tilde{\mathbf{s}}$) from FEC decoder, it is infeasible to find a precoder that globally optimizes $I(\mathbf{z},\mathbf{s})$. However, the chain rule of mutual information enables to decompose the symbol-wise MI into a sum of M bitwise MIs, which

allows us to write $I(\mathbf{z},\mathbf{s}) = \sum_{L=0}^{M-1} I(\mathbf{z},\mathbf{s}|L \text{ other bits known})$ [54, 77], where $M = (\log_2 Q)^b$ is the number of bits per mapped symbol. On one hand, ten-Brink has also shown in [77] that $I(\mathbf{z},\mathbf{s}|\text{no other bit known})(\approx I(\mathbf{z},\mathbf{s}|\tilde{\mathbf{s}} = \mathbf{0}))$ and $I(\mathbf{z},\mathbf{s}|\text{all other bits known})(\approx I(\mathbf{z},\mathbf{s}|\tilde{\mathbf{s}} = \mathbf{s}))$ respectively correspond to the starting ($I_E^1(0)$) and ending ($I_E^1(1)$) points in the EXIT function. On the other hand, the results in [75] show that maximizing the ending point of the equalizer EXIT function results in a low error-floor of turbo equalization, while maximizing the starting point of the equalizer EXIT function not only leads to a fast convergence but also shifts the turbo-cliff region to a lower SNR. Indeed, it is apparently shown in FIGURE 5.2 that maximizing $I_E^1(1)$ avoids the early intersection of the two EXIT functions at the convergence (*i.e.* it forces the early-crossing point in Fig. 5.2 to a higher position) and maximizing $I_E^1(0)$ gives a good initial point for the trajectory, which helps to avoid the intersection at bottleneck.

In summary, we propose in this section a precoder that maximizes $I(\mathbf{z},\mathbf{s}|\tilde{\mathbf{s}} = \mathbf{s})$ ($I(\mathbf{z},\mathbf{s})$ at the optimum convergence state) and $I(\mathbf{z},\mathbf{s}|\tilde{\mathbf{s}} = \mathbf{0})$ ($I(\mathbf{z},\mathbf{s})$ at the initial state). The optimization is split into two steps. First, the priority is to maximize $I(\mathbf{z},\mathbf{s}|\tilde{\mathbf{s}} = \mathbf{s})$ in order to minimize the error-floor. Thanks to this step, the first two parameters (ψ,θ) in (5.1) of $\mathbf{F}_d$ are found. Second, with the values of (ψ,θ) found in the first step, we optimize the last parameter ϕ in (5.1) of $\mathbf{F}_d$ to further improve $I(\mathbf{z},\mathbf{s}|\tilde{\mathbf{s}} = \mathbf{0})$.

5.4.2 Optimizing $I(\mathbf{z},\mathbf{s}|\tilde{\mathbf{s}} = \mathbf{s})$

In this subsection, we firstly prove that maximizing $I(\mathbf{z},\mathbf{s})$ leads to jointly maximizing the SNRs on all sub-streams of the IC output. Subsequently, we solve the thus defined optimization problem by considering the optimum convergence state.

Let us assume that the sub-streams at the output of IC are independent, then $I(\mathbf{z},\mathbf{s}) = \sum_{k=1}^{b} I(z_k,s_k)$. Hence, maximizing $I(\mathbf{z},\mathbf{s})$ is equivalent to jointly maximizing $I(z_k,s_k)$, where $k \in [1,\ldots,b]$ and s_k is the Q-ary modulated symbol on the k^{th} sub-stream. By applying the same reasoning as in [68] to the equivalent single-input single-output sub-stream expressed in (5.4), a lower-bound of $I(z_k,s_k)$ can be formulated as

$$I_{LB}(z_k,s_k) = \log_2 Q - \left(\frac{1}{\ln 2} - 1\right) - \frac{1}{Q}\sum_{m=1}^{Q} \log_2 \left[\sum_{n=1}^{Q} \exp\left(-\frac{\mu_k^2 |s_m - s_n|^2}{2\sigma_{\xi_k}^2}\right)\right]. \tag{5.6}$$

Thus, in order to maximize $I(z_k, s_k)$, we maximize the lower bound $I_{LB}(z_k, s_k)$, which connects to maximize $\frac{\mu_k^2}{\sigma_{\xi_k}^2}$, where $\sigma_{\xi_k}^2 = \sigma_s^2 \mu_k (1-\mu_k)$. From (1.49), this is equivalent to maximize the SNR p_k on the sub-stream. Note that the same conclusion is obtained when we maximize the analytical EXIT function on each sub-stream. Indeed, from [74], the EXIT function on each sub-stream is given by $I_{E,k} \approx \left(1 - 2^{-H_1 (4p_k)^{H_2}}\right)^{H_3}$ in case of complex-valued modulation. With 4-QAM modulation, $H_1 = 0.3073, H_2 = 0.8935$ and $H_3 = 1.1064$. Therefore, maximizing $I_{E,k}$ is also equivalent to maximize p_k.

Let us step forward to jointly optimize p_k, $k \in [1, \ldots, b]$. Recall from (1.54) that, with $\mathbf{s} = \tilde{\mathbf{s}}$, $\mathbf{B} = \sigma_\eta^2 \mathbf{I}_b$, $\mu_k = \frac{\frac{\sigma_s^2}{\sigma_\eta^2} \mathbf{A}_{:,k}^\dagger \mathbf{A}_{:,k}}{1 + \frac{\sigma_s^2}{\sigma_\eta^2} \mathbf{A}_{:,k}^\dagger \mathbf{A}_{:,k}}$, and ξ_k depends only on $\boldsymbol{\eta}$. The SNR thus equals

$$p_k = \frac{\sigma_s^2}{\sigma_\eta^2} \mathbf{A}_{:,k}^\dagger \mathbf{A}_{:,k}. \tag{5.7}$$

From (5.1), with $\mathbf{H}_v = \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix}$, we have $\mathbf{A}\mathbf{A}^\dagger = \begin{pmatrix} \sigma_1^2 \cos^2\psi & 0 \\ 0 & \sigma_2^2 \sin^2\psi \end{pmatrix}$. Expansion of (5.7) yields

$$p_1 = \frac{\sigma_s^2}{\sigma_\eta^2} \Big(\left(\sigma_1^2 \cos^2\psi - \sigma_2^2 \sin^2\psi\right) \cos^2\theta + \sigma_2^2 \sin^2\psi \Big), \tag{5.8}$$

$$p_2 = \frac{\sigma_s^2}{\sigma_\eta^2} \Big(\left(\sigma_1^2 \cos^2\psi - \sigma_2^2 \sin^2\psi\right) \sin^2\theta + \sigma_2^2 \sin^2\psi \Big). \tag{5.9}$$

The joint maximization of (p_1, p_2) is considered as a multi-objective optimization without any special expectation for the solutions. Therefore, a non-preference method [67] is applied. The problem of finding (ψ, θ) that jointly maximizes (p_1, p_2) becomes

$$(\psi^\star, \theta^\star) = \arg\min_{\mathcal{F}} \left(\mho(\psi, \theta) = (p_1 - p_1^{\max})^2 + (p_2 - p_2^{\max})^2 \right), \tag{5.10}$$

where $p_k^{\max}$ is the maximum value of p_k over the set $\mathcal{F}$, defined by $0^o \le \psi \le 90^o$ and $0^o \le \theta \le 90^o$.

Lemma 5. The maximum value of p_1 and p_2 over the set $\mathcal{F}$ are $p_1^{\max} = p_2^{\max} = \frac{\sigma_s^2}{\sigma_\eta^2} \sigma_1^2$.

Proof. See Appendix A. □

Exploiting (5.8), (5.9), Lemma 5 and trigonometry results, the expression of $\mho$ reads

$$\mho(\psi,\theta) = \frac{\sigma_s^4}{\sigma_\eta^4}\left[\left(\cos^4\theta+\sin^4\theta\right)\left(\sigma_1^2\cos^2\psi-\sigma_2^2\sin^2\psi\right)+2\sigma_1^2\sin^2\psi(\sigma_1^2-\sigma_2^2\sin^2\psi)\right]. \tag{5.11}$$

The optimization problem (5.10) can be solved by searching for (ψ,θ) that satisfy the first and second order conditions for a local minimum. Partial derivatives of $\mho$ with respect to θ and ψ yield

$$\begin{aligned}
\frac{\partial\mho}{\partial\theta}(\psi,\theta) &= -\frac{\sigma_s^4}{\sigma_\eta^4}\sin(4\theta)\left(\sigma_1^2\cos^2\psi-\sigma_2^2\sin^2\psi\right)^2,\\
\frac{\partial\mho}{\partial\psi}(\psi,\theta) &= \frac{\sigma_s^4}{\sigma_\eta^4}\sin(2\psi)\left[2(\sigma_1^4+\sigma_2^4)\sin^2\psi-2\sigma_1^2\sigma_2^2(\sigma_1^2+\sigma_2^2)\sin^2(2\theta)\left(\sigma_1^2-(\sigma_1^2+\sigma_2^2)\sin^2\psi\right)\right].
\end{aligned} \tag{5.12}$$

The set of points (ψ,θ) of $\mathcal{F}$ that satisfy the first order conditions for a local minimum ($\nabla\mho=\mathbf{0}$), denoted by $\mathcal{F}^\star$, equals

$$\begin{aligned}
\mathcal{F}^\star = \Bigg\{&(0,0),(0,45^o),(0,90^o),(90^o,0),(90^o,45^o),(90^o,90^o),\left(\arcsin\frac{\sigma_1\sigma_2}{\sqrt{(\sigma_1^4+\sigma_2^4)}},0\right),\\
&\left(\arcsin\frac{\sigma_1\sigma_2}{\sqrt{(\sigma_1^4+\sigma_2^4)}},90^o\right)\Bigg\}.
\end{aligned} \tag{5.13}$$

Second order partial derivatives of $\mho$ with respect to θ and ψ yield

$$\begin{aligned}
\frac{\partial^2\mho}{\partial\theta^2}(\psi,\theta) &= -4\frac{\sigma_s^4}{\sigma_\eta^4}\cos(4\theta)\left(\sigma_1^2\cos^2\psi-\sigma_2^2\sin^2\psi\right)^2,\\
\frac{\partial^2\mho}{\partial\psi^2}(\psi,\theta) &= 2\frac{\sigma_s^4}{\sigma_\eta^4}\cos(2\psi)[(\sigma_1^2+\sigma_2^2)\sin^2(2\theta)\left(\sigma_1^2-(\sigma_1^2+\sigma_2^2)\sin^2\psi\right)-2\sigma_1^2\sigma_2^2+2(\sigma_1^4+\sigma_2^4)\sin^2\psi]\\
&\quad-\frac{\sigma_s^4}{\sigma_\eta^4}\sin^2(2\psi)\left[(\sigma_1^2+\sigma_2^2)\sin^2(2\theta)-2(\sigma_1^4+\sigma_2^4)\right],\\
\frac{\partial^2\mho}{\partial\theta\partial\psi}(\psi,\theta) &= -\frac{\sigma_s^4}{\sigma_\eta^4}(\sigma_1^2+\sigma_2^2)\sin(4\theta)\sin(2\psi)\left(\sigma_1^2\cos^2\psi-\sigma_2^2\sin^2\psi\right).
\end{aligned} \tag{5.14}$$

The only point of $\mathcal{F}^\star$ that satisfies the second order conditions for a local minimum (the Hessian matrix $\nabla^2\mho$ is positive-definite) is $(\psi,\theta)=(0^o,45^o)$. As $\mho$ is convex, the minimum is global.

5.4.3 Improving $I(\mathbf{z}, \mathbf{s}|\tilde{\mathbf{s}} = \mathbf{0})$

At the initial state, where $\tilde{\mathbf{s}} = \mathbf{0}$, the equalizer coincides with a MMSE linear equalizer. The output of IC reads

$$\mathbf{z}_0 = \mathbf{W}\mathbf{y} = \mathbf{W}\mathbf{A}\mathbf{s} + \underbrace{\mathbf{W}\boldsymbol{\eta}}_{\boldsymbol{\xi}}. \tag{5.15}$$

Let us assume that $\boldsymbol{\xi} \sim \mathcal{CN}(0, \sigma_\xi^2 \mathbf{I}_b)$. We denote by $\mathcal{S} = \{\mathbf{s}^{(1)}, \ldots, \mathbf{s}^{(i)}, \ldots, \mathbf{s}^{(N)}\}$ the set of all possible values of the symbol vector $\mathbf{s}$, where $N = Q^b$, in which Q is the size of Q-ary QAM alphabet and $b = 2$ is the number of data streams. By applying directly [68], a lower bound of $I(\mathbf{z}, \mathbf{s}|\tilde{\mathbf{s}} = \mathbf{0})$ for the MIMO equivalent channel is given by

$$I_{LB}(\mathbf{z}_0, \mathbf{s}) = b \log_2 Q - \left(\frac{1}{\ln 2} - 1\right) b - \frac{1}{Q^b} \sum_{k=1}^{Q^b} \log_2 \left[\sum_{\ell=1}^{Q^b} \exp\left(- \frac{\|\mathbf{W}\mathbf{A}(\mathbf{s}^{(k)} - \mathbf{s}^{(\ell)})\|^2}{2\sigma_\xi^2} \right) \right]. \tag{5.16}$$

Thus, to maximize $I(\mathbf{z}, \mathbf{s}|\tilde{\mathbf{s}} = \mathbf{0})$, we maximize the lower bound $I_{LB}(\mathbf{z}_0, \mathbf{s})$ by maximizing $\|\mathbf{W}\mathbf{A}\boldsymbol{\nu}^{(k,\ell)}\|^2$, where $\boldsymbol{\nu}^{(k,\ell)} = \begin{pmatrix} \nu_1^{(k,\ell)} & \nu_2^{(k,\ell)} \end{pmatrix}^T = \mathbf{s}^{(k)} - \mathbf{s}^{(\ell)}$ is the difference vector between the vector symbols. Up to this step, we fix (ψ, θ) to the optimum values obtained in Section 5.4.2, *i.e.* $(\psi = 0^o, \theta = 45^o)$, and we aim to find the best value for ϕ. With $\tilde{\mathbf{s}} = \mathbf{0}$, we deduce $\mathbf{W} = \sigma_s^2 \mathbf{A}^\dagger \left(\sigma_s^2 \mathbf{A}\mathbf{A}^\dagger + \sigma_\eta^2 \mathbf{I}_b \right)^{-1}$. Hence, with $(\psi = 0^o, \theta = 45^o)$, calculation yields

$$\|\mathbf{W}\mathbf{A}\boldsymbol{\nu}^{(k,\ell)}\|^2 = \left(\frac{\frac{\sigma_1^2}{2}(\frac{\sigma_1^2}{2} + \sigma_\xi^2)}{(\frac{\sigma_1^2}{2} + \sigma_\xi^2)^2 - \frac{\sigma_1^4}{4}} \right)^2 \left\| \begin{pmatrix} \nu_1^{(k,\ell)} + \nu_2^{(k,\ell)} e^{i\phi} \\ \nu_1^{(k,\ell)} e^{-i\phi} + \nu_2^{(k,\ell)} \end{pmatrix} \right\|^2. \tag{5.17}$$

Since $|(\nu_1^{(k,\ell)} + \nu_2^{(k,\ell)} e^{i\phi})|^2 = |(\nu_1^{(k,\ell)} e^{-i\phi} + \nu_2^{(k,\ell)})|^2$, the optimum value of ϕ, referred to as ϕ_{opt}, can be found by

$$\phi_{opt} = \underset{0^o \le \phi \le 90^o}{\arg\max} \left(d_{\min}^{IC} \right), \tag{5.18}$$

where $d_{\min}^{IC} = \min_{k \neq \ell} \underbrace{\left| \nu_1^{(k,\ell)} + \nu_2^{(k,\ell)} e^{i\phi} \right|}_{d^{(k,\ell)}(\phi)}$.

Lemma 6. For any symmetric QAM modulation, the searching interval of ϕ can be restricted from $0^o \le \phi \le 90^o$ to $0^o \le \phi \le 45^o$.

Proof. See Appendix C of Chapter 3. □

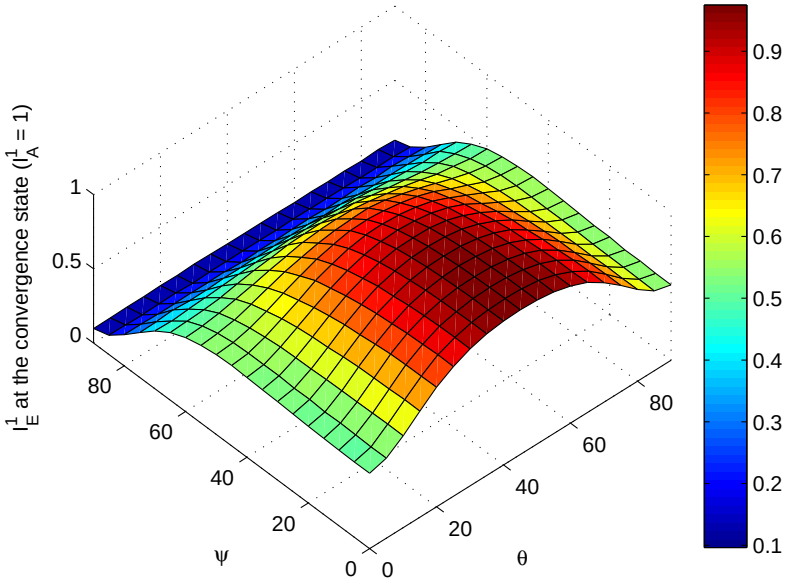


FIGURE 5.6: The extrinsic MI of IC at optimum convergence state ($I_A^1 = 1$), *i.e.* $I_E^1(1)$, $\phi = 30^o$, Channel A, SNR = 11 dB, 4-QAM, ψ and θ are in degree.

Since finding a complete analytical solution for ϕ is intractable, up to this step, we resort to numerical optimization to look for ϕ. We consider 4-QAM modulation and the searching range is limited to $0^o \leq \phi \leq 45^o$ thanks to Lemma 6. Numerical search over all possible symbol vectors shows that $d_{\min}^{IC}$ can be obtained by considering only the two following pairs of symbol vectors: $\left\{\left(\frac{1-i}{\sqrt{2}} \quad \frac{-1-i}{\sqrt{2}}\right)^T; \left(\frac{-1-i}{\sqrt{2}} \quad \frac{1-i}{\sqrt{2}}\right)^T\right\}$ and $\left\{\left(\frac{-1+i}{\sqrt{2}} \quad \frac{-1-i}{\sqrt{2}}\right)^T; \left(\frac{1-i}{\sqrt{2}} \quad \frac{-1+i}{\sqrt{2}}\right)^T\right\}$. The former couple has difference vector $\boldsymbol{\nu} = \left(\sqrt{2} \quad -\sqrt{2}\right)^T$, which yields $d^{(k,\ell)}(\phi) = 2\sqrt{1-\cos\phi}$ referred to as d_1. The latter couple has difference vector $\boldsymbol{\nu} = \left(-\sqrt{2}+i\sqrt{2} \quad -i\sqrt{2}\right)^T$, which yields $d^{(k,\ell)}(\phi) = \sqrt{6-4(\cos\phi+\sin\phi)}$ referred to as d_2. The $d_{\min}^{IC}$ is equal to d_1 for $0^o \leq \phi \leq 30^o$ and to d_2 for $30^o \leq \phi \leq 45^o$. The optimum value of ϕ, which maximizes $d_{\min}^{IC}$, is obtained at the intersection between d_1 (increasing function in ϕ) and d_2 (decreasing function in ϕ), which yields $\phi_{opt} = 30^o$. Therefore, we propose to take $\phi = 30^o$. The proposed precoder with optimized defining parameters $(\psi = 0^o, \theta = 45^o, \phi = 30^o)$, is referred to as Genie-optimized precoder. It should be noted that the parameters of this precoder are fixed, which makes its design and practical application easier.

5.5 Validation through EXIT chart

5.5.1 Comparison with the theoretical results

To check the consistency of the proposed solution, we fix $\phi = 30^o$ and make (ψ, θ) vary. We analyze the mutual information at the optimum convergence state in FIGURE 5.6 for SNR = 11 dB. The maximum value of I_E^1 is obtained at $(\psi = 0^o, \theta = 45^o)$, which confirms the theoretical calculations. We also would like to point out that there is no significant difference in terms of $I_E^1(1)$ when other values of ϕ are considered, since ϕ only shows its influence at the starting point $I_E^1(0)$.

5.5.2 Comparison with the existing precoders

In this subsection, we compare the proposed precoder to other selected existing precoders. The analysis is carried out through EXIT chart comparison. We consider the Channel A ($\gamma = 8.3^o$) and Channel B ($\gamma = 30^o$) as already introduced in Chapter 3.

The two best precoders that give the high channel capacity $I(\mathbf{y}, \mathbf{s})$, which are GOPT (see Section 2.3.2.1) and max-$d_{\min}$(see Section 2.3.2.2) precoders, are considered for comparison. Let us recall that the GOPT precoder gives the best capacity over the complex-valued MIMO channels since it aims to globally maximize the mutual information between the finite alphabet input and the corresponding channel output. The max-$d_{\min}$ precoder can also achieve a channel capacity close to the GOPT precoder one at high SNR. In the case of two data streams transmission ($b = 2$) and 4-QAM modulation, the max-$d_{\min}$ precoder shows a better uncoded error-rate performance with maximum likelihood detection than the other precoders such as MMSE, WF, MWF, max-SNR and minimum BER. The parameters of max-$d_{\min}$ applied to (5.1) are reminded in TABLE 5.1.

Precoder		ψ	θ	ϕ
Genie-optimized		0^o	45^o	30^o
max-$d_{\min}$	$\mathbf{F}_{\text{octa}}(\gamma > \gamma_0)$	$\arctan\left(\frac{\sqrt{2}-1}{\tan\gamma}\right)$	45^o	45^o
	$\mathbf{F}_{\text{r1}}$ $(\gamma \leq \gamma_0 \simeq 17.28^o)$	0^o	27.37^o	15^o

TABLE 5.1: The defining parameters of max-$d_{\min}$ precoder and the proposed Genie-optimized precoder.

The EXIT charts for the considered precoders over Channel A and Channel B are given in FIGURE 5.7 and FIGURE 5.8 respectively. We notice that the EXIT functions of

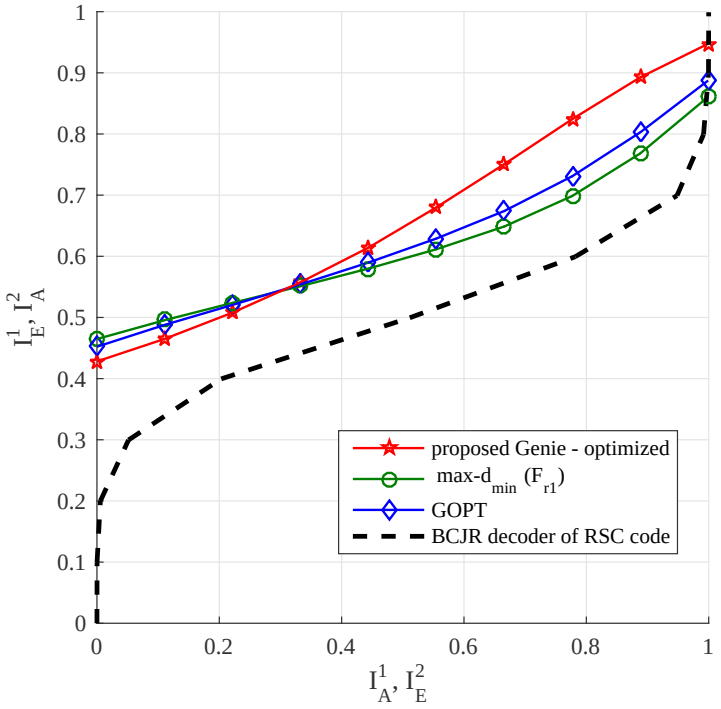


FIGURE 5.7: EXIT charts of IC using the considered precoders at SNR = 10 dB, Channel A, 4-QAM and $(13, 15)$-RSC code.

GOPT and max-$d_{\min}$ precoders are close to each other. This is in accordance with the conclusion in [68] about the similarity between GOPT and max-$d_{\min}$ at high SNR. More importantly, it is observed that the EXIT function of IC corresponding to the Genie-optimized precoder converges to a much higher I_E^1 than the others, which predicts a better error-rate performance at the error floor. This is also demonstrated in FIGURE 5.9. In this figure, the extrinsic MI at the ending points of the EXIT chart, *i.e.* I_E^1 at $I_A^1 = 1$ or $I_E^1(1)$, are plotted for all of possible channels, which are represented by the parameter γ. Recall that, for max-$d_{\min}$ precoder, the threshold to switch between $\mathbf{F}_{r1}$ and $\mathbf{F}_{octa}$ is $\gamma_0 \simeq 17.28^o$. As observed, the EXIT charts always predict significant gains (in terms of $I_E^1(1)$) of the Genie-optimized precoder compared to the others. This confirms the theoretical calculations done in Section 5.4. In addition, it is apparently observed from Fig. 5.9 that the Genie-optimized precoder (which has only one precoding form) is represented by one stable curve, and the max-$d_{\min}$ precoder (which has two precoding forms) is represented by a curve combined from two different stable parts separated by γ_0 (see Table 5.1). While the GOPT precoder, which is optimized by running an algorithm for each channel realization and SNR, does not have fixed forms. It has different precoding forms for different gammas and SNRs. Therefore, the plot of GOPT in Fig. 5.9 shows a fluctuated

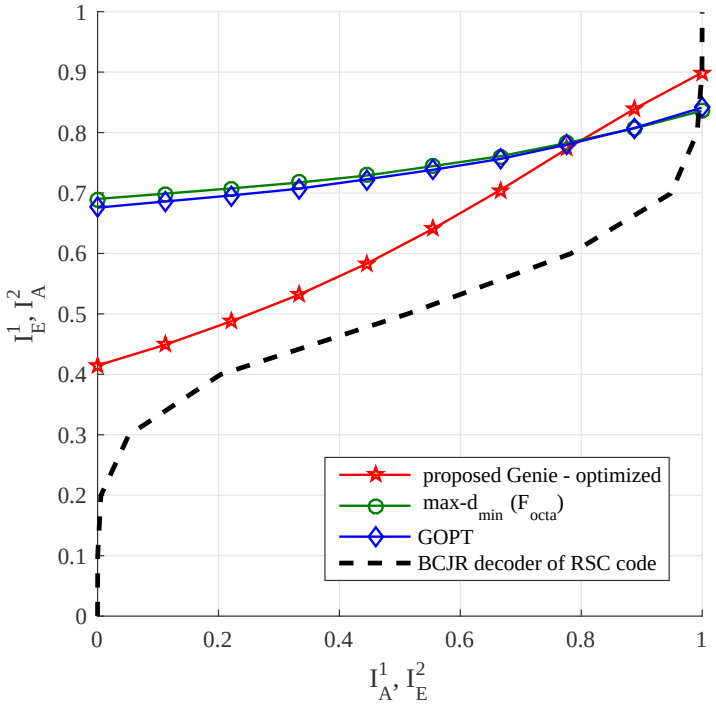


FIGURE 5.8: EXIT charts of IC using the considered precoders at SNR = 10 dB, Channel B, 4-QAM and $(13, 15)$-RSC code.

curve. This shows a high impact of the precoding form represented by the triple (ψ, θ, ϕ) on $I_E^1(1)$ (*i.e.* on the error-rate performance at the error-floor), which is optimized by the Genie-optimized precoder.

In Fig. 5.9, we also present the plot for max-$d_{\min}$-mod precoder proposed in Section 5.3. It is observed that, thanks to the new threshold γ_{th}, the values of $I_E^1(1)$ in the interval $\gamma_0 \leq \gamma \leq \gamma_{th}$ is improved by using the max-$d_{\min}$-mod precoder compared to the max-$d_{\min}$ precoder. One should note that, in Section 5.3, we proposed the max-$d_{\min}$-mod precoder by keeping the two precoding forms ($\mathbf{F}_{r1}$ and $\mathbf{F}_{octa}$) of max-$d_{\min}$ and introduced the threshold γ_{th} to improve $I_E^1(1)$. In this section, we completely optimize the parameterized precoding form of $\mathbf{F}_d$ (see (5.1)) to maximize $I_E^1(1)$ and $I_E^1(0)$ respectively.

We would like to point out that, though EXIT charts in this subsection are plotted at SNR=10 dB, similar observations are obtained at other SNR values. Thanks to the EXIT chart analysis, it is predicted that the proposed Genie-optimized precoder achieves a better error-rate performance compared to the selected reference precoders at the SNR values that are high enough to avoid intersections at bottleneck of the EXIT charts. Our

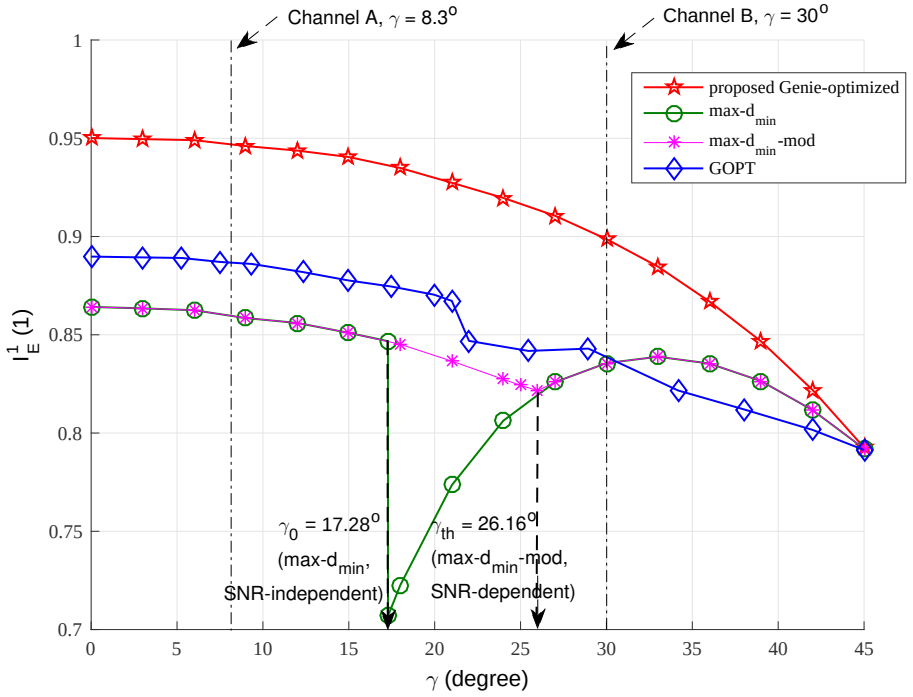


FIGURE 5.9: I_E^1 at $I_A^1 = 1$ versus the angle γ at SNR $= 10$ dB and 4-QAM.

simulation results are in excellent agreement with these analytical results and will be presented in Section 5.6.

We conclude this section by mentioning that maximizing $I_E^1(1)$ only results in significant error-rate performance gains when low-complex FEC codes are used. Because the slope of the EXIT function of FEC decoder is flatter with strong FEC codes than weak (low-complex) FEC codes, as a result, the decoder is able to reach a low error-rate (*i.e.* to reach $I_E^2 = 1$) even at a lower I_A^2. In contrast, since the slope of the EXIT function of FEC decoder is steep with low-complex FEC codes, the decoder needs high *a priori* information from equalizer to reach the low error-rate. In other words, it requires a high $I_E^1(1)$ to avoid early intersection, which leads to high error-rate. Therefore, the proposed Genie-optimized precoder is suitable to be used in the applications that need a low-complex FEC code, which could be crucial in the future machine-to-machine communications.

5.6 Simulation results and discussion

We now provide examples to demonstrate the advantages of the proposed precoder in terms of error-rate performance. Recall from Section 5.4 that we characterize a channel

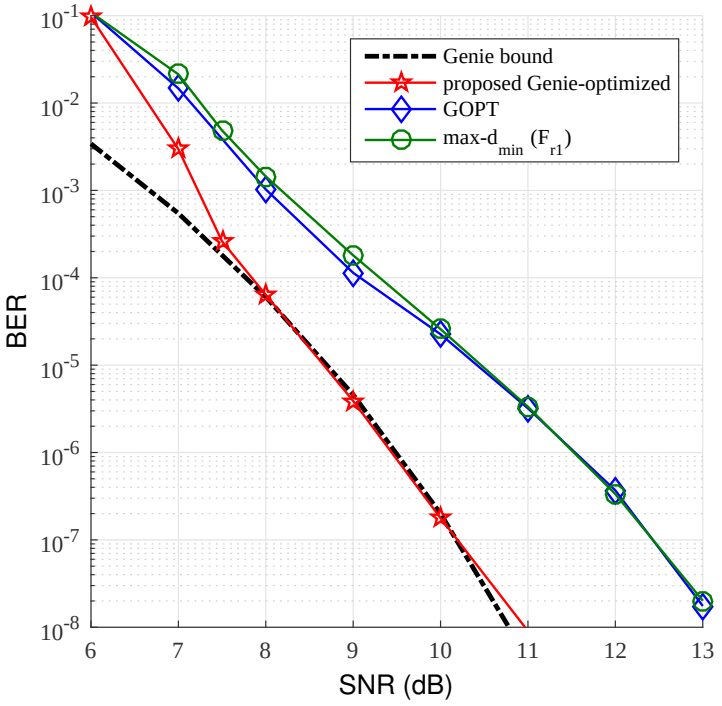


FIGURE 5.10: BER of Channel A, 4-QAM and $(13, 15)$-RSC code.

by the angle γ, where $0^o \leq \gamma \leq 45^o$. As in Section 5.5, we consider Channel A and Channel B to illustrate the two precoding forms of the max-$d_{\min}$ precoder. We complete the analysis by simulations over randomly generated channels to cover all possible values of γ. The half-rate $(13, 15)$-RSC code is used in a $n_T = 2$ transmit and $n_R = 2$ receive antennas MIMO configuration. The frame length is 2000 uncoded bits. The Monte-Carlo simulations stop when 100 error frames have occurred with the condition that at least 5000 frames have been generated. The number of iterations is fixed to 10.

Channel A with $\gamma = 8.3^o$ is considered to illustrate the case $\gamma \leq \gamma_0$. Note that with $\gamma = 8.3^o$ max-$d_{\min}$ uses only the strongest subchannel σ_1. The GOPT algorithm also yields the optimal precoder that uses only the strongest subchannel. FIGURE 5.10 shows the BER performance of the different precoders on this channel. The performances of max-$d_{\min}$ and GOPT precoders are close to each other over this channel. The Genie-optimized precoder achieves a gain of almost 2 dB compared to max-$d_{\min}$ at BER $\simeq 10^{-6}$. Note that, in case $\gamma \leq \gamma_0$, max-$d_{\min}$ uses the $\mathbf{F}_{r1}$ form. In addition, the optimum convergence state of turbo equalization referred to as the genie bound, with the Genie-optimized precoder used at transmitter, is also simulated. The bound is obtained by providing both SBC and BSC blocks with the maximum *a priori* LLRs, *i.e.* $L_A^1 = (2\bar{c} - 1)K$, where

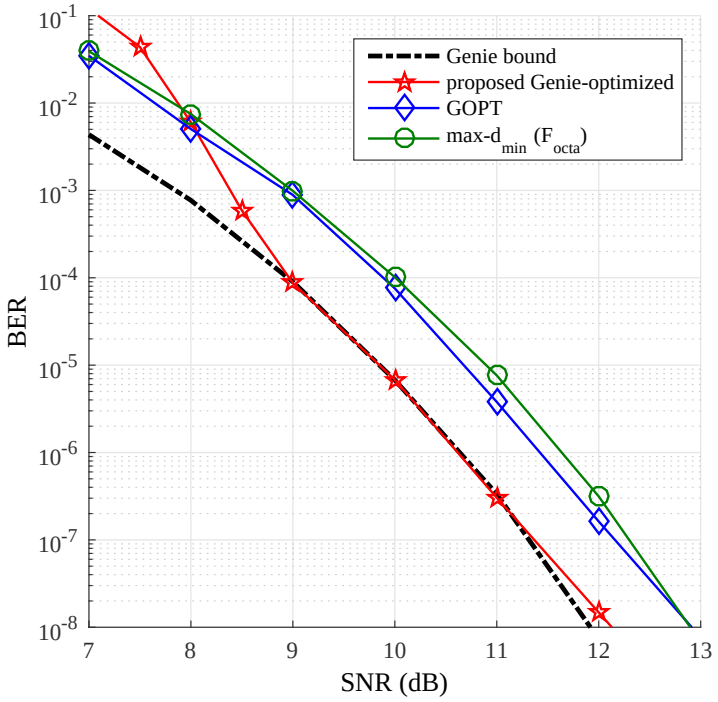


FIGURE 5.11: BER of Channel B, 4-QAM and $(13, 15)$-RSC code.

$\bar{c}$ is the bit-interleaved coded binary sequence and K is a large-enough positive constant. The error-rate performance is directly measured at the first iteration. We can see from FIGURE 5.10 that the Genie-optimized precoder converges close to its lower bound.

Channel B with $\gamma = 30^o$ is considered to illustrate the case $\gamma > \gamma_0$, where max-$d_{\min}$ uses $\mathbf{F}_{octa}$ form. With the fixed precoding form, the Genie-optimized precoder continues to use only the strongest subchannel σ_1. While, in contrary to the solutions for Channel A, the max-$d_{\min}$ and GOPT (in the considered SNR range) precoders use both subchannels σ_1 and σ_2. This shows the difference in terms of precoding strategies between the proposed Genie-optimized precoder compared to the max-$d_{\min}$ precoder (asymptotically maximizes the channel capacity at high SNRs) and the GOPT precoder (maximizes the channel capacity). FIGURE 5.11 shows the BER performance of the considered precoders. The Genie-optimized precoder achieves a gain of roughly 1 dB at BER = 10^{-6} compared to max-$d_{\min}$ and GOPT. At low SNR, the GOPT and max-$d_{\min}$ perform slightly better than Genie-optimized. This is in accordance with the EXIT chart analysis done for Channel B (see FIGURE 5.8), which predicts the early intersection at the bottleneck of the chart for Genie-optimized precoder at low SNR. By comparing Channel A and Channel B, it can be concluded that the advantage of Genie-optimized precoder is more significant over

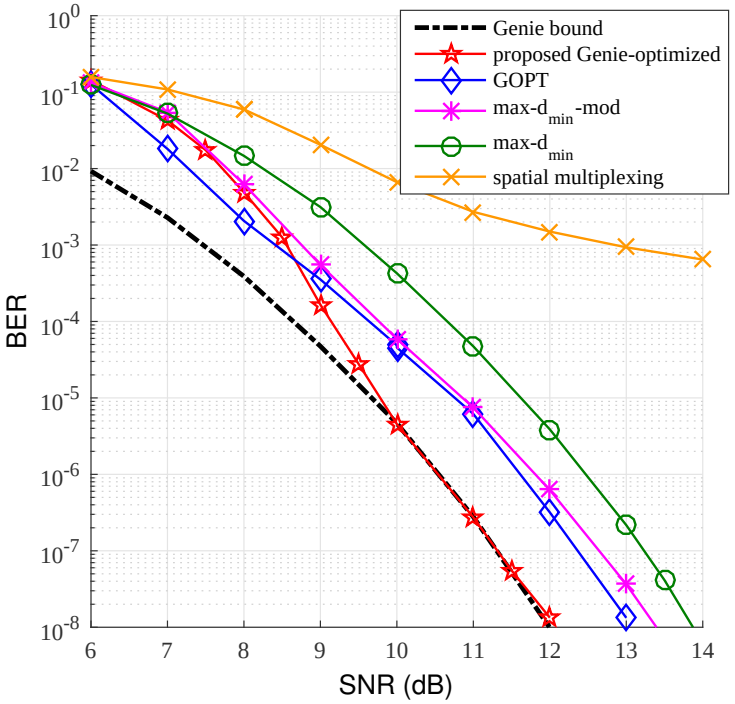


FIGURE 5.12: Average BER over random channels, 4-QAM and $(13, 15)$-RSC code.

the low angle γ channels (*i.e.* $\sigma_2 \ll \sigma_1$).

Randomly generated channels (with $h_{ij} \sim \mathcal{CN}(0, 1)$, where h_{ij} is element of the 2×2 channel matrix $\mathbf{H}$) are considered in this subsection to affirm the advantage of Genie-optimized precoder for any value of γ. By taking the average over randomly generated channels, the performance curve in Fig. 5.12 shows the average error-rate over γ. As observed, GOPT precoder performs slightly better than Genie-optimized precoder at low SNR. The explanation for this very little gain is similar to the one done for Channel B, since the results for random channels are averaged over different values of γ including the high γ channels such as Channel B. More importantly, the proposed solution achieves a gain of roughly 4.5 dB at BER $= 10^{-3}$ and 2 dB at BER = 10^{-6} compared to spatial multiplexing and max-$d_{\min}$ respectively. Compared to GOPT, the proposed Genie-optimized precoder achieves a gain of roughly 1 dB at BER $= 10^{-7}$. Note that the Genie-optimized precoder has significant advantage in terms of computational complexity compared to GOPT precoder. Because Genie-optimized uses a fixed SNR-independent closed-form, instead of using the complex SNR-dependent algorithm to find precoding matrices as in the case of GOPT.

We observe that, over random channels, GOPT precoder achieves a significant gain

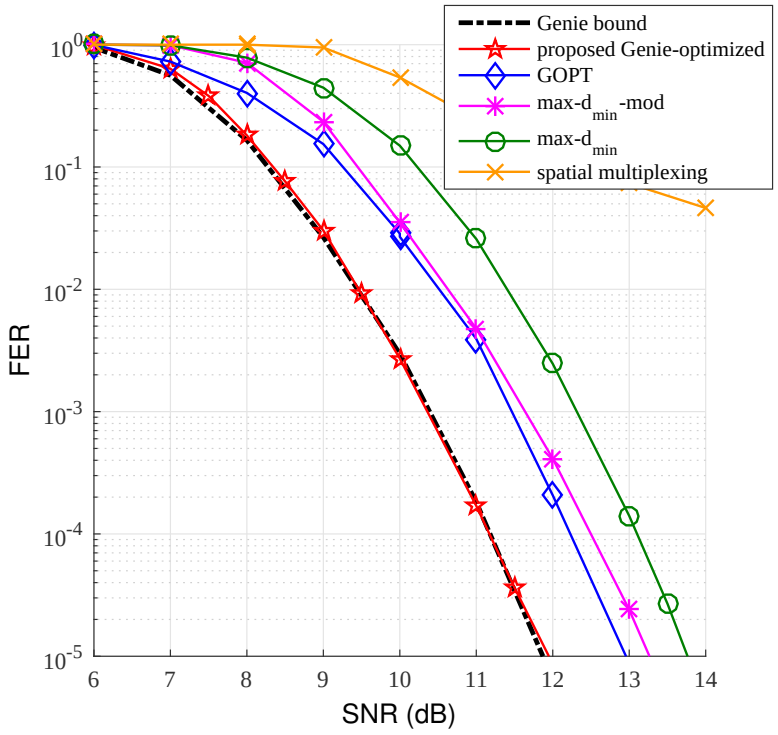


FIGURE 5.13: Average FER over random channels, 4-QAM and $(13, 15)_{octal}$-RSC code.

compared to max-$d_{\min}$ precoder, which is not observed over the fixed channels A and B. This can be explained by the significant differences in terms of $I_E^1(1)$ between GOPT and max-$d_{\min}$ precoders in the interval $\gamma_0 \leq \gamma \leq \gamma_{th}$ (see Fig. 5.9), where, over random channels, the random angle γ can be distributed in. In contrast, the differences are not significant at the γs of Channel A and Channel B. The performance of max-$d_{\min}$-mod precoder, which achieves a gain compared to the conventional max-$d_{\min}$ precoder, is plotted again for comparison. Similar observations are obtained in terms of FER as illustrated in Fig. 5.13. Note that, in terms of FER, the proposed Genie-optimized precoder always outperforms the other reference precoders (even at low SNRs). It early reaches the Genie-bound and exhibits significant performance gains compared to the other ones.

5.7 Conclusion

The association of MIMO linear precoder and turbo equalization is investigated in this chapter. We propose a novel convergence-based SNR-dependent threshold for the two working modes of max-$d_{\min}$ precoder (referred to as max-$d_{\min}$-mod precoder). More

importantly, a new Genie-optimized MIMO linear precoder is proposed in this chapter under the assumption of an outer FEC encoder and turbo equalization at the receiver. In contrast to the conventional precoders that maximize the mutual information (MI) between finite alphabet input and the corresponding output over the precoded MIMO channel (channel capacity), the proposed precoder aims to maximize the MI between the finite alphabet input and the corresponding output of the equalizer. The optimizations were carried out at the convergence state and the initial state of the turbo equalization, which respectively correspond to the ending and starting points in the EXIT chart. Thanks to the EXIT chart analysis, we have shown that the extrinsic MI at the convergence state, *i.e.* $I_E^1(1)$, corresponding to the new precoder, is higher than $I_E^1(1)$ achieved by the other precoders. This confirms the theoretical analysis. Simulations show high improvement in terms of error-rate of the MIMO scheme at high enough SNR thanks to the new precoder, which appears to be a promising solution for modern MIMO communications whose applications require low-complex FEC codes. The increase of b, however, is necessary for the high data-rate purpose. In fact, in large scale MIMO systems [78], the constraint $\min(n_T, n_R) \geq b$ for a high value of b can be easily fulfilled.

As mentioned in Section 1.1, NB-LDPC codes show advantages compared to its binary counterpart. Therefore, the association of NB-LDPC with MIMO systems is interesting. The main challenge in the NB-LDPC encoded systems is the high computational complexity at the receiver (mostly spent at the decoding procedure due to the high-order GF). In the next chapter, we investigate the NB-LDPC encoded MIMO systems. We show that the computational complexity at the receiver can be significantly reduced by mapping multiple lower-order GF symbols onto one MIMO symbol instead of mapping one high-order GF symbol. We then propose to apply MIMO precoder at the transmitter and turbo detection at the receiver to compensate for the performance loss due to the multiple mapping.

Appendices of chapter 5

A Proof of Lemma 5

From (5.8), the problem of finding (ψ, θ) that maximize p_1 becomes

$$(\psi^\diamond, \theta^\diamond) = \arg\max_{\mathcal{F}} \left(\tfrac{\sigma_s^2}{\sigma_\eta^2} \left[\left(\sigma_1^2 \cos^2\psi - \sigma_2^2 \sin^2\psi \right) \cos^2\theta + \sigma_2^2 \sin^2\psi \right] \right). \tag{5.19}$$

Recall that the set $\mathcal{F}$ is defined by $0^o \leq \psi \leq 90^o$ and $0^o \leq \theta \leq 90^o$.

Partial derivatives of p_1 with respect to θ and ψ yield

$$\frac{\partial p_1}{\partial \psi} = \frac{\sigma_s^2}{\sigma_\eta^2} \sin 2\psi \left[\sigma_2^2 - \left(\sigma_1^2 + \sigma_2^2 \right) \cos^2\theta \right], \tag{5.20}$$

$$\frac{\partial p_1}{\partial \theta} = \frac{\sigma_s^2}{\sigma_\eta^2} \sin 2\theta \left(\sigma_2^2 \sin^2\psi - \sigma_1^2 \cos^2\psi \right). \tag{5.21}$$

The set of points (ψ, θ) of $\mathcal{F}$ that satisfy the first order conditions for a local maximum $(\nabla p_1 = 0)$, denoted by $\mathcal{F}^\diamond$, equals

$$\mathcal{F}^\diamond = \left\{ (0,0), (0, 90^o), (90^o, 0), (90^o, 90^o) \right\}. \tag{5.22}$$

Second order partial derivatives of p_1 with respect to θ and ψ yield

$$\frac{\partial^2 p_1}{\partial \psi^2} = 2\frac{\sigma_s^2}{\sigma_\eta^2} \cos 2\psi \left[\sigma_2^2 - \left(\sigma_1^2 + \sigma_2^2 \right) \cos^2\theta \right], \tag{5.23}$$

$$\frac{\partial^2 p_1}{\partial \theta^2} = 2\frac{\sigma_s^2}{\sigma_\eta^2} \cos 2\theta \left(\sigma_2^2 \sin^2\psi - \sigma_1^2 \cos^2\psi \right), \tag{5.24}$$

$$\frac{\partial^2 p_1}{\partial \psi \partial \theta} = \frac{\sigma_s^2}{\sigma_\eta^2} \sin 2\theta \sin 2\psi \left(\sigma_1^2 + \sigma_2^2 \right). \tag{5.25}$$

The only point of $\mathcal{F}^{\diamond}$ that satisfies the second order conditions for a local maximum is $(\psi, \theta) = (0, 0)$, which yields $p_1^{\max} = \frac{\sigma_s^2}{\sigma_\eta^2}\sigma_1^2$. The same optimization is applied for p_2, which yields $p_2^{\max} = \frac{\sigma_s^2}{\sigma_\eta^2}\sigma_1^2$ at $(\psi = 0, \theta = 90^o)$.

Chapter 6

Turbo Detection of NB-LDPC Codes in Precoded MIMO Systems

The content of this chapter is mainly based on the following paper:

- Nhat-Quang Nhan, Philippe Rostaing, Karine Amis, Emanuel Radoi, and Y. Rosa Zheng. "Turbo detection of NB-LDPC codes in precoded MIMO systems" Submitted to *IEEE Transaction on Vehicular Technology*, pending for review.

6.1 Introduction

This chapter investigates the association of NB-LDPC codes with MIMO linear precoders assuming non-binary turbo detection at the receiver. Let us firstly recall that, in the conventional binary encoded MIMO communications, the codeword bits are grouped and mapped onto Q-ary modulated symbols. Then, a serial-to-parallel conversion is applied to convert these symbols into MIMO symbol vectors, which define a MIMO constellation. The size of this constellation is therefore equal to Q^b, where b is the number of data streams (layers) in the MIMO system. In contrast to this conventional binary mapping, the researches reported in [79, 80] show that the GF order of NB-LDPC codes can be chosen to be equal to the MIMO constellation size. Subsequently, each GF symbol is directly mapped onto a data symbol vector. This mapping is convenient for the association of NB-LDPC in MIMO communications and we refer to it as *one-by-one mapping*.

Modern wireless communications, especially the fifth generation (5G) cellular networks, require high data-rate with low transmission latency [81]. Thus, the high-order modulation would be used extensively [82]. Using high-order modulation leads to an exponential increase of the MIMO constellation size. Thereby, a very high-order GF is needed for one-by-one mapping. Unfortunately, high-order GF NB-LDPC codes require very high computational complexity at the NB-LDPC decoder, which increases the system latency. To overcome that problem, we consider in this chapter a mapping such that several low-order GF symbols are mapped onto one MIMO symbol. We refer to this mapping as *multiple mapping*. Then, the jointly detection-decoding iterative receiver (or turbo detection) is investigated. It is shown that, with the same modulation order, the multiple mapping used with turbo detection scheme impressively reduces the computational complexity at receiver compared to the one-by-one mapping. However, this advantage is counterbalanced by two main challenges. The first challenge is that multiple mapping leads to error-rate performance degradation at the error-floor (an erroneous demodulated symbol yields multiple erroneous FEC codeword components). The second challenge includes the optimization of iteration number between detector and decoder in turbo detection scheme as well as iteration number inside of the decoder to further reduce the decoding complexity.

The NB-LDPC iterative receiver scheme has been investigated in several papers, but no paper focuses on the two challenges as listed above. Indeed, the paper in [83] presented the advantage of using a rate-1 encoder in concatenation with the NB-LDPC codes assuming sphere packing demapper. The main focus in [84] was the design and performance comparisons of different cycle code constructions. The most likely related paper can be found in [85]. However, the paper [85] focused on space-time mapping strategy and accumulator trade-off comparisons rather than on the decoding complexity or the error-rate performance improvement.

In this chapter, we apply one technique to tackle both of the mentioned challenges. The contributions of this chapter are summarized as follows. Firstly, we propose to use MIMO precoders to overcome the drawback of turbo detection scheme in terms of error-rate performance, especially at the error-floor. These precoders are selected from precoders in literature under the criterion of maximizing the channel mutual information, which are well adapted for MIMO encoded systems as presented in Chapter 2. By using the precoders, significant performance gains are achieved. Secondly, we show that the

computational complexity at receiver is remarkably reduced thanks to precoders. The optimization for the turbo detection scheme configuration is also discussed. Last but not least, we show that the iteration number inside the decoder can be minimized by using the precoders. Hence, with few iterations, the well-known belief propagation (BP) NB-LDPC decoding algorithm achieves optimality [86, Chapter 3], especially when it is used with large girth parity check matrices. Indeed, by using a small number of iterations to decode a large-girth code, only the extrinsic information are exchanged between the check and the variable nodes as in the case of cycle-free codes [87]. Therefore, the BP in log-domain (Log-BP [40]) decoding algorithm is considered throughout this chapter. However, as we consider the detector and the demapper in turbo detection scheme as two independent blocks (only soft information are exchanged), the extension to other decoding algorithms is straightforward.

Since the receiver complexity is significantly reduced and the error-rate performance is enhanced thanks to the proposed solution, the results in this chapter can be applied to the communication schemes that take into account NB-LDPC codes but require low-complex fast-decoding at the receiver such as machine-to-machine communications. The organization of this chapter is as follows: In Section 6.2, some preliminaries and the MIMO precoders are briefly reviewed. The complexity of receiver and the EXIT chart analysis of the proposed solution are introduced in Section 6.3. Simulation results are presented in Section 6.4. Finally, Section 6.5 wraps up with a summary of conclusions.

6.2 Preliminaries

6.2.1 System model

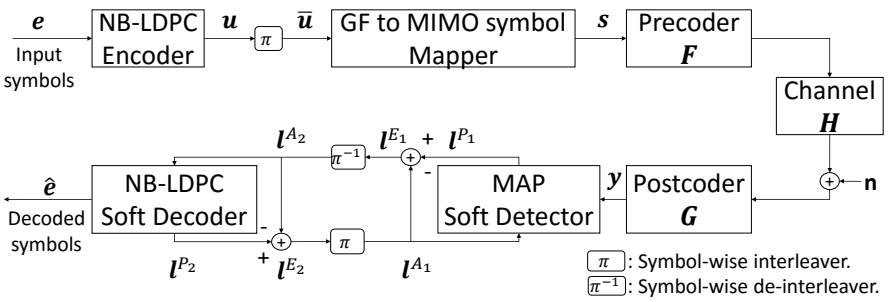


FIGURE 6.1: System model.

Let us consider a MIMO system with n_R receive, n_T transmit antennas and b independent data streams to be transmitted as shown in FIGURE 6.1. We assume full-CSI at both the transmitter and the receiver. A NB-LDPC encoder of Galois field order q (GF(q)) is used for channel coding. The obtained codeword $\mathbf{u}$ is then interleaved into $\bar{\mathbf{u}}$ by a symbol-wise interleaver. At the mapper, every b Q-ary quadrature amplitude modulation (QAM) symbols are grouped into a vector $\mathbf{s}$ of size $b \times 1$. By taking all possibilities of $\mathbf{s}$, a constellation of MIMO symbols $\mathcal{S} = \{\mathbf{s}_0, \ldots, \mathbf{s}_k, \ldots, \mathbf{s}_{Q^b-1}\}$ is defined. Following this, every n_s GF symbols in $\bar{\mathbf{u}}$ are grouped and mapped onto a MIMO symbol $\mathbf{s} \in \mathcal{S}$ with a mapping rule $\mathcal{X}$. The MIMO symbol $\mathbf{s}$ is then precoded by a precoder $\mathbf{F}$ before being transmitted through the MIMO channel. Although n_s GF symbols in $\bar{\mathbf{u}}$ can be mapped onto multiple MIMO symbols instead of one MIMO symbol, that mapping is however not interesting since it requires a higher complexity at the demapper.

Though non-binary mapping is considered, the MIMO precoding scheme as introduced in Chapter 2 does not change. Therefore, the channel transformation with the channel output presented in (2.16) is exploited again in this chapter. The receiver consists of a maximum *a posteriori* (MAP) soft-detector and an NB-LDPC soft-decoder, which iteratively exchange the extrinsic log-likelihood ratio vectors (LLRVs). This iterative loop is called *outer-loop*, where the number of iterations in this loop is denoted by n_out. Similarly, *inner-loop* refers to the iterative loop inside the decoder with the number of iterations equal to n_in. Let us denote by $\{0, \alpha_1, \ldots, \alpha_i, \ldots, \alpha_{q-1}\}$ the non-binary symbols in GF(q). Then, a LLRV of a GF symbol a is defined by $\boldsymbol{l} = [\boldsymbol{l}(\alpha_1), \ldots, \boldsymbol{l}(\alpha_i), \ldots, \boldsymbol{l}(\alpha_{q-1})]^T$, where each element of $\boldsymbol{l}$ is calculated by $\boldsymbol{l}(\alpha_i) = \ln \frac{P(a=\alpha_i)}{P(a=0)}$.

The *a priori*, *a posteriori* and extrinsic LLRVs of the detector are denoted by $\boldsymbol{l}^{A_1}$, $\boldsymbol{l}^{P_1}$ and $\boldsymbol{l}^{E_1}$. The equivalent notations for decoder are $\boldsymbol{l}^{A_2}$, $\boldsymbol{l}^{P_2}$ and $\boldsymbol{l}^{E_2}$ respectively. Recall that $\mathbf{s} = \mathcal{X}(\mathbf{a} = [a_1, \ldots, a_j, \ldots, a_{n_\mathrm{s}}])$. Therefore, for each input vector $\mathbf{y}$, the extrinsic output of detector is a matrix of n_s LLR column vectors, which is defined by $\mathbf{L}^{E_1} = [\boldsymbol{l}_1^{E_1}, \ldots, \boldsymbol{l}_j^{E_1}, \ldots, \boldsymbol{l}_{n_s}^{E_1}]$. With the corresponding *a priori* LLR matrix $\mathbf{L}_1^{A_1} = [\boldsymbol{l}_1^{A_1}, \ldots, \boldsymbol{l}_\ell^{A_1}, \ldots, \boldsymbol{l}_{n_s}^{A_1}]$, each element of $\boldsymbol{l}_j^{E_1}$ is then calculated by

$$\begin{aligned} \boldsymbol{l}_j^{E_1}(\alpha_i) = & \max_{\mathbf{s}\in\mathcal{S}|\mathbf{a}:a_j=\alpha_i}{}^{*} \left[\frac{-\|\mathbf{y} - \mathbf{H}_v\mathbf{F}_d\mathbf{s}\|^2}{\sigma^2} + \sum_{\ell=1,\ell\neq j}^{n_\mathrm{s}} \boldsymbol{l}_\ell^{A_1}(a_\ell) \right] \\ & - \max_{\mathbf{s}\in\mathcal{S}|\mathbf{a}:a_j=0}{}^{*} \left[\frac{-\|\mathbf{y} - \mathbf{H}_v\mathbf{F}_d\mathbf{s}\|^2}{\sigma^2} + \sum_{\ell=1,\ell\neq j}^{n_\mathrm{s}} \boldsymbol{l}_\ell^{A_1}(a_\ell) \right], \end{aligned} \tag{6.1}$$

$\mathbf{F}_d$	$\mathbf{F}_1$	$\mathbf{F}_2$	$\mathbf{F}_3$	$\mathbf{F}_4$	$\mathbf{F}_5$
ψ	0	$\arctan\frac{5\sqrt{2}-7}{\tan\gamma}$	$\arccos\sqrt{\frac{\alpha-\alpha\cos^2\gamma}{\alpha-2\cos^2\gamma}}$	$\arctan\frac{\sqrt{15-4\sqrt{14}}}{\tan\gamma}$	$\arctan\frac{\sqrt{2}-1}{\tan\gamma}$
θ	$\arctan(2\sin\phi)$	$\pi/4$	$\pi/4$	$\frac{1}{2}\arctan\frac{\sqrt{10}}{2}$	$\pi/4$
ϕ	$\arctan\frac{1}{6+\sqrt{3}}$	$\pi/4$	$\arctan\frac{3}{5}$	$\arctan\frac{1}{3}$	$\pi/4$
γ range (in degree)	$0^o \le \gamma < 5.12^o$	$5.12^o \le \gamma < 5.26^o$	$5.26^o \le \gamma < 8.4^o$	$8.4^o \le \gamma < 15.38^o$	$15.38^o \le \gamma \le 45^o$

TABLE 6.1: Parameters for the five matrices of max-$d_{\min}$ precoder in case of 16-QAM and $b = 2$, $\alpha = 1 + \frac{6}{\sqrt{34}}$ [63].

	Additions	Multiplications	$\overset{*}{\max}$
Detector	$\frac{N_v}{n_s}Q^b\left(7b^2+7b-2\right) + n_{\text{out}}N_v(qn_s-1)$	$\frac{N_v}{n_s}Q^b\left(3b^2+3b+1\right)$	$n_{\text{out}}[N_v q(q-1)]$
Decoder [40]	$n_{\text{out}}n_{\text{in}}[2(3d_c-4)N_c(q-1)^2 + d_c N_c(d_v-1)(q-1)]$		$n_{\text{out}}n_{\text{in}}\left[2(3d_c-4)N_c(q-1)^2\right]$

TABLE 6.2: Number of operations used at the iterative receiver for each codeword.

where $\overset{*}{\max}$ stands for the Jacobian logarithm as introduced in Chapter 1.

6.2.2 MIMO precoders

Recent research [88] has shown that significant error-rate performance improvement can be achieved by maximizing the channel capacity of the NB-LDPC encoded MIMO system assuming the one-by-one mapping. Following this, we consider again in this chapter the GOPT and max-$d_{\min}$ precoders, which have been introduced in Section 2.3.2.1 and Section 2.3.2.2 respectively. These two precoders result in high channel capacity.

Considering 16-QAM modulation and $b = 2$ data streams, Ngo *et al.* proposed in [17, 63] an extension of max-$d_{\min}$ precoder that has five different closed-form precoding matrices. The precoder switches among these matrices depending on $\gamma = \arctan\frac{\sigma_2}{\sigma_1}$. In other words, this precoder also adapts to the channel thanks to the parameter γ. The five precoding matrices, which are denoted by $\mathbf{F}_1$ to $\mathbf{F}_5$ respectively, are obtained by substituting the corresponding triplet (ψ, θ, ϕ) in TABLE 6.1 into the parameterized form of $\mathbf{F}_d$ in (6.2). The corresponding operating ranges of γ are also given in TABLE 6.1. Note that the five precoding matrices are fixed for all SNRs (SNR independent).

$$\mathbf{F}_d = \begin{pmatrix} \cos\psi & 0 \\ 0 & \sin\psi \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & e^{i\phi} \end{pmatrix}. \tag{6.2}$$

6.3 Analysis

6.3.1 Computational complexity

As mentioned, we assume that the Log-BP algorithm is used at the decoder. Let us denote by N_c and N_v the number of check nodes (CNs) and variable nodes (VNs) in an NB-LDPC code where the corresponding CNs and VNs degrees are d_c and d_v respectively. The complexity of the iterative receiver, that is required to decode every codeword, is summarized in TABLE 6.2.

It is explicitly shown that the considered system significantly reduces the decoding complexity. Indeed, let us pick up an example with $b = 2$, 16-QAM modulation, the parity check matrix derived from matrices designed in the framework of DAVINCI project [89, 90] with $N_v = 192, N_c = 96, d_c = 4$, $d_v = 2$ and the girth equal to 14. The MIMO constellation size is therefore equal to $Q^b = 256$. On one hand, if the one-by-one mapping is applied, *i.e.* $n_\mathrm{s} = 1$ and $n_\mathrm{out} = 1$, we need to use an NB-LDPC code with GF order $q = 256$. On the other hand, if the multiple mapping and iterative receiver are applied (as considered in this chapter), we can map every $n_\mathrm{s} = 2$ GF order $q = 16$ into a MIMO symbol. Note that $n_\mathrm{out} > 1$ in this case. Without loss of generality, let us select $n_\mathrm{out} = 10$ and $n_\mathrm{in} = 10$. We denote by $n_+^{\text{1-by-1}}$ and $n_\boxplus^{\text{1-by-1}}$ the total additions and $\overset{*}{\max}$ operations required in the case of one-by-one mapping. The equivalent notations for the case of multiple mapping are $n_+^{n_s\text{-by-1}}$ and $n_\boxplus^{n_s\text{-by-1}}$ respectively. Calculation yields $n_+^{n_s\text{-by-1}}/n_+^{\text{1-by-1}} \approx n_\boxplus^{n_s\text{-by-1}}/n_\boxplus^{\text{1-by-1}} \approx 0.036$. In other words, the number of additions and $\overset{*}{\max}$ operations is reduced by roughly 96.4% using the multiple mapping. Note that, in this example, we select $n_\mathrm{in} = 10$ for both one-by-one and multiple mappings. In practice, the one-by-one mapping needs more n_in for decoding than the multiple mapping, which leads to the fact that the complexity reduction by using the multiple mapping becomes even more significant. This shows the advantage of the multiple mapping assuming iterative receiver as considered in this chapter.

In addition, it comes out directly from TABLE 6.2 that the complexity of turbo detection scheme can be further reduced by minimizing the factor $\lambda = n_\mathrm{out} n_\mathrm{in}$, which is the total iterations used at the decoder. However, by doing so, the error-rate will also raise up as a compensation. In the next subsection, we propose to use MIMO precoders to take up the error-rate challenge. The selection of λ will also be discussed.

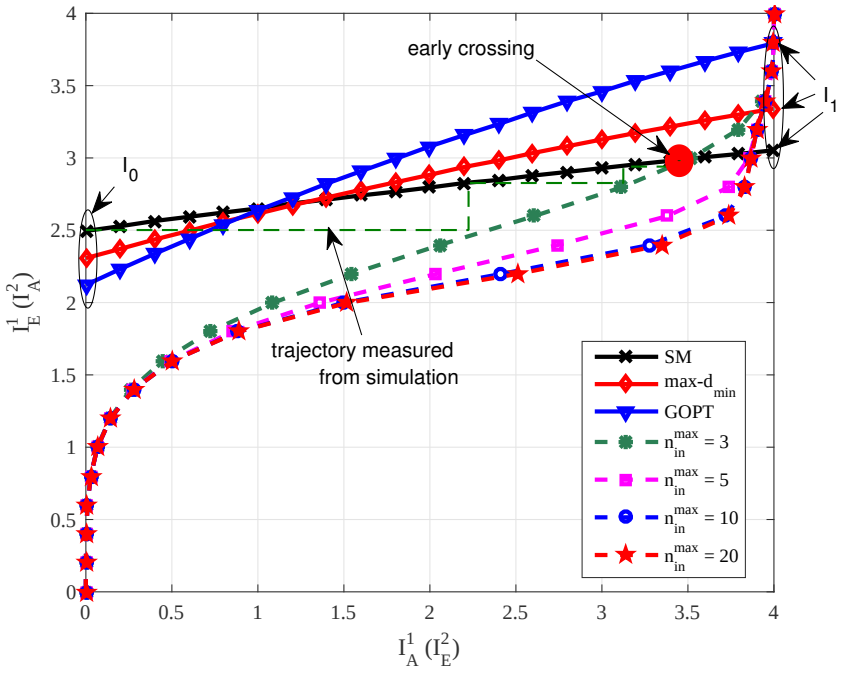


FIGURE 6.2: EXIT chart of a given channel at SNR = 18 dB.

6.3.2 Non-binary EXIT chart

The extension of extrinsic information transfer (EXIT) chart for symbol-based iterative decoding in [91] is used to analyze the iterative receiver. We consider the decoder with a given maximum number of n_{in} ($n_{\text{in}}^{\max}$) as an individual component in the chart, while the other one is the detector. The same simulation parameters as in the example given in Section 6.3.1 are considered. The SNR definition is given in (3.3). For instance, let us arbitrarily select a channel $\mathbf{H} = [0.4067-0.1801i\ 0.0419+0.2015i;\ -0.8206-0.0268i\ 0.2896-0.0428i]$ for demonstration. We denote this channel by $\mathbf{H}_{\text{ex}}$. The GOPT precoding matrices at some selected SNR along with the corresponding channel capacities, which are found by applying the algorithm in [22] over the channel $\mathbf{H}_{\text{ex}}$, are provided in TABLE 6.3. One should note that, in order to apply these matrices to the model in (2.2), we need to normalize them to satisfy the power constraint $\|\mathbf{F}\|_F^2 = 1$. The channel capacities obtained by max-$d_{\min}$ precoder, which are not far from the one achieved by GOPT precoder, are also given in TABLE 6.3.

FIGURE 6.2 shows the EXIT chart for this channel at SNR = 18 dB. The dashed lines represent the EXIT functions of the decoder for different $n_{\text{in}}^{\max}$. The solid lines are the EXIT functions of the detector when spatial multiplexing (SM), max-$d_{\min}$ (precoding matrix is given by $\mathbf{F}_4$ in TABLE 6.1 since $\gamma_{\mathbf{H}_{\text{ex}}} \approx 10^o$) and GOPT (precoding matrix

SNR	**F** (of GOPT)	Chan. cap. GOPT	Chan. cap. max-$d_{\min}$
18	$\begin{bmatrix} -0.8645+0.0360i & -0.8346+0.2190i \\ 0.5359+0.3961i & -0.1780-0.1765i \end{bmatrix}$	5.9125	5.6448
20	$\begin{bmatrix} -0.7986+0.0426i & -0.8320+0.2087i \\ 0.5654+0.4329i & -0.2648-0.2174i \end{bmatrix}$	6.6352	6.4969
22	$\begin{bmatrix} -0.7314+0.0484i & -0.8237+0.1976i \\ 0.5889+0.4642i & -0.3439-0.2541i \end{bmatrix}$	7.2743	7.2519

TABLE 6.3: GOPT precoding matrices at some selected SNR (in dB), which are found by applying the Algorithm 1 (reported in Chapter 4) with the initial parameters ($\omega = \frac{\pi}{5}, \nu = \frac{\pi}{10}$) over $\mathbf{H}_{\text{ex}}$ channel, and the corresponding channel capacities (in bits/s/Hz).

is given in TABLE 6.3 at SNR = 18 dB) precoders are used at the transmitter (or, in short, EXIT functions of the precoders). Note that only EXIT functions of detector are changed with SNR. The starting and ending points of a non-binary detector EXIT function are denoted by I_0 and I_1 respectively.

As shown in FIGURE 6.2, the receiver starts to converge from $n_{\text{in}}^{\max} = 10$. We firstly observe that the EXIT function of decoder can be used to predict the required minimum $n_{\text{in}}^{\max}$ in the conventional non-iterative scheme. In turbo detection scheme used with SM, although limiting $n_{\text{in}}^{\max}$ can reduce λ, the EXIT functions of SM detector and the decoder are early crossed. This leads to a bad error-rate performance. However, the problem of early crossing can be solved by using precoders. We observe from FIGURE 6.2 that I_1 of GOPT and max-$d_{\min}$ precoders are much higher than the one of SM. This predicts better error-floors for the system used with precoders. In contrast with I_1, the I_0 of precoders are lower than SM. This may causes intersection between EXIT functions of detector and decoder at low SNR. Moreover, the lower I_0, the more closed the tunnel between two EXIT functions, which requires a high n_{out} for convergence. Therefore, although using precoders can significantly enhance the error-rate performance and allow us to use low $n_{\text{in}}^{\max}$, the best selection of $n_{\text{in}}^{\max}$ (to reduce λ) however depends on each precoder. Indeed, simulations in Section 6.4 show that different optimized $n_{\text{in}}^{\max}$ are obtained for SM, max-$d_{\min}$ and GOPT precoders.

6.4 Simulation results

We now propose examples to demonstrate the advantages of using precoders in terms of error-rate performance as well as complexity reduction. Monte-Carlo simulations have been carried out for $(n_T = 2, n_R = 2)$ MIMO systems with $b = 2$ data streams, $Q = 16$

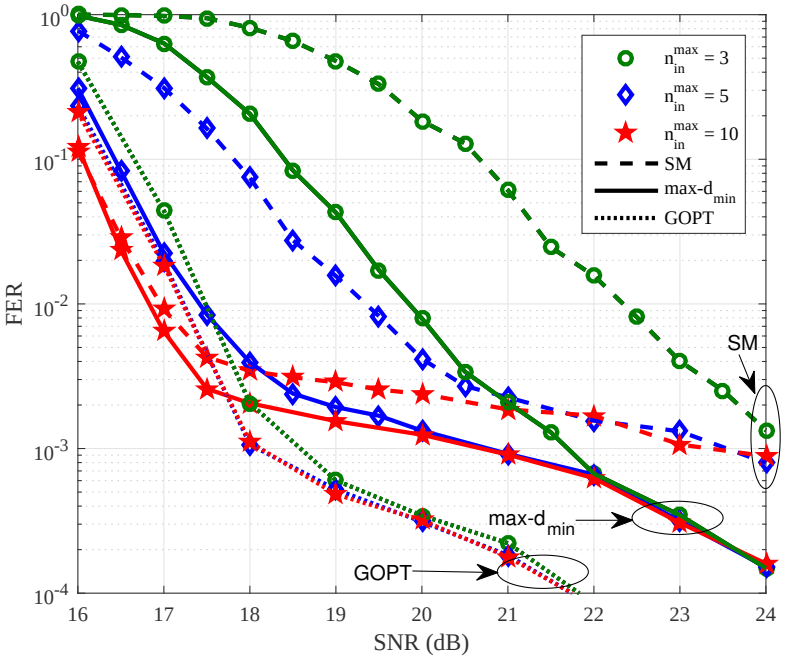


FIGURE 6.3: FER of the fixed channel $\mathbf{H}_{ex}$.

(16-QAM modulation), GF(q = 16) and n_s = 2 GF symbols per MIMO symbol. The MIMO constellation size is therefore equal to Q^b = 256. Let us recall that the NB-LDPC matrix is derived from matrices designed in the framework of DAVINCI project [89, 90] with (N_v = 192, N_c = 96, d_c = 4, d_v = 2, girth = 14) and the Log-BP decoding algorithm [40] is considered in this chapter. The maximum iteration number that has been used to limit the outer-loop is 10.

We firstly evaluate the error-rate performance over the fixed channel $\mathbf{H}_{ex}$, which has been used for the EXIT chart analysis. FIGURE 6.3 shows the system frame error-rate (FER) performance versus SNR over this channel in case SM (dashed lines), max-d_{min} (solid lines) and GOPT (dotted lines) precoders are used at the transmitter. The lines with equivalent n_{in}^{max} share the same marker. We observe that error-floor of the system is significantly reduced thanks to the precoders. With the same n_{in}^{max}, the error-rate performance is enhanced by using precoder. For example, for n_{in}^{max} = 5 at FER = 10^{-3}, the GOPT precoder achieves a gain of roughly 6 dB and 3 dB compared to SM and max-d_{min} precoder respectively. In addition, for SM and max-d_{min} precoders, the respective plots associated to different n_{in}^{max} converge at high SNR. This can be explained by the increase of I_1 to the top-right corner of the EXIT chart at high SNR. For GOPT precoder, this convergence happens at a lower SNR (see FIGURE 6.2). All of the simulation results for the fixed channel $\mathbf{H}_{ex}$ are in accordance with the EXIT chart analysis

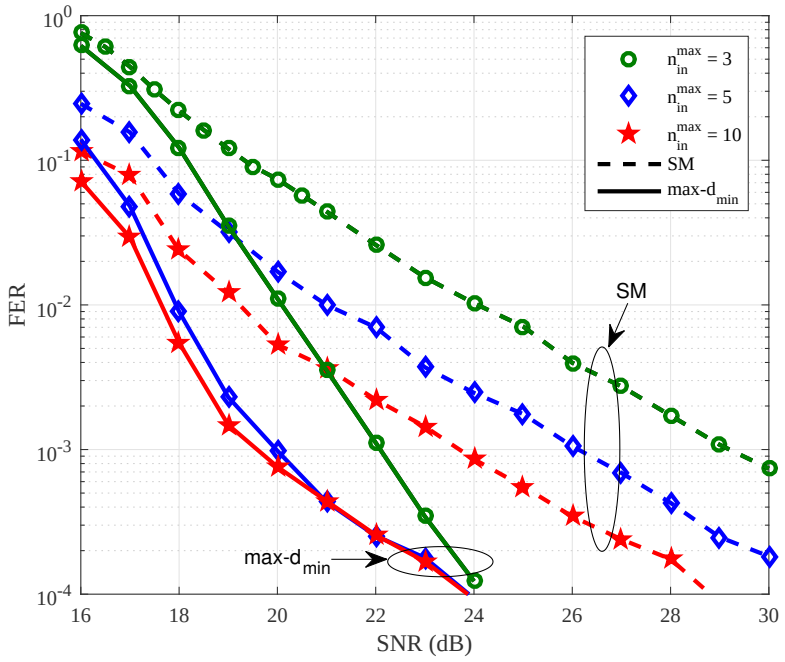
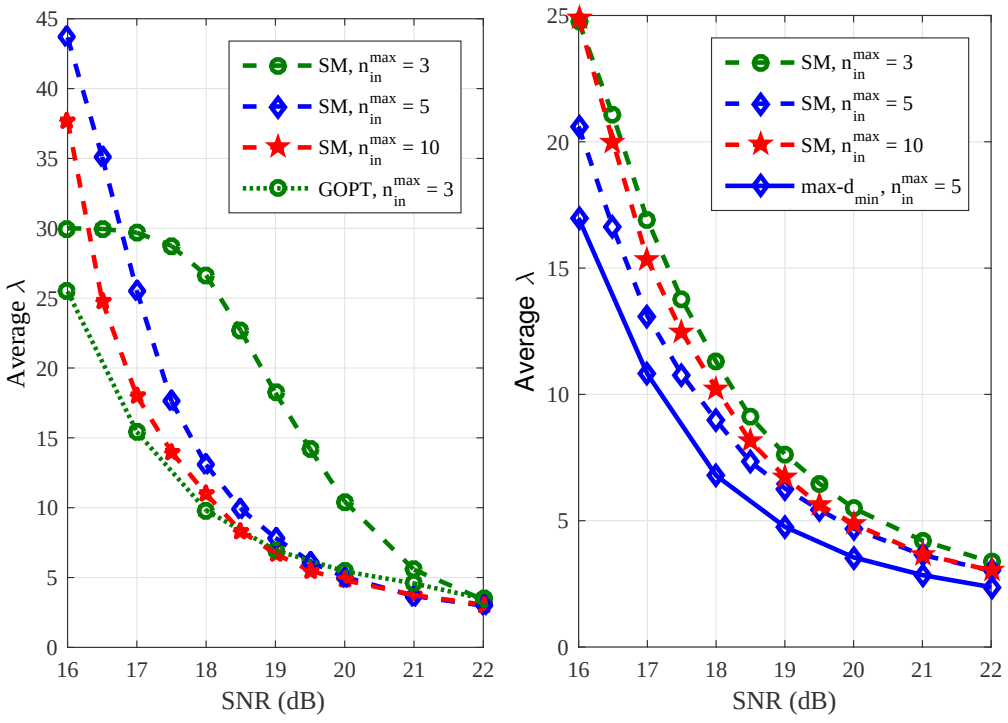


FIGURE 6.4: Average FER of random channel.

done in Section 6.3.2.

Let us step forward to evaluate the error-rate performance of the considered system over the random channel ($h_{ij} \sim \mathcal{CN}(0,1)$, where h_{ij} is element of the channel matrix $\mathbf{H}$). FIGURE 6.4 shows the average system FER performance versus SNR over the randomly generated channels in case SM (dashed lines) and max-$d_{\min}$ precoder (solid lines) are used at the transmitter. The lines with equivalent $n_{\text{in}}^{\max}$ share the same marker. As mentioned, with a large MIMO constellation size as considered in this chapter, finding the best solutions of GOPT precoder for each channel at different SNRs takes a long simulation time. Therefore, we do not consider GOPT for the random channel. However, we would like to point out that it could be practical to use GOPT for constructing a codebook of precoding matrices as done in Chapter 4. As shown in FIGURE 6.4, the performance gain achieved by using max-$d_{\min}$ precoder is even more significant compared to the one achieved over the fixed channel. The max-$d_{\min}$ precoder used with $n_{\text{in}}^{\max} = 5$ achieves a gain of roughly 5 dB compared to SM used with $n_{\text{in}}^{\max} = 10$ at FER $= 10^{-4}$.

The average λ (denoted by $\bar{\lambda}$) is also taken into account. In case SM is used over the fixed channel $\mathbf{H}_{\text{ex}}$, the lowest $\bar{\lambda}$ is obtained when $n_{\text{in}}^{\max} = 10$ as demonstrated in FIGURE 6.5(a). Note that high value of n_{in} is counterbalanced by low value of n_{out}, which totally results in a low $\bar{\lambda}$ in this case. The best $n_{\text{in}}^{\max}$ selection (in terms of $\bar{\lambda}$) of GOPT is also presented



(a) Fixed channel (b) Average over random channels

FIGURE 6.5: Receiver complexity.

in FIGURE 6.5(a) for comparison. It is observed that the GOPT precoder used with $n_{\text{in}}^{\max} = 3$ requires a lower $\bar{\lambda}$ than the lowest one of SM, namely $n_{\text{in}}^{\max} = 10$. Note that, with the corresponding $n_{\text{in}}^{\max}$, the GOPT precoder achieves a gain of roughly 4.5 dB at FER $= 10^{-3}$ compared to SM (see FIGURE 6.3). FIGURE 6.5(b) shows the plots of $\bar{\lambda}$ versus SNR for SM precoder in case the random fading channel is considered. The best $n_{\text{in}}^{\max}$ selection (in terms of $\bar{\lambda}$) of max-$d_{\min}$ precoder is also presented in FIGURE 6.5(b). We can see that, in fact, the best $n_{\text{in}}^{\max}$ for SM (in terms of $\bar{\lambda}$) over random channel is equal to 5. More importantly, the complexity reduction by using max-$d_{\min}$ precoder (with $n_{\text{in}}^{\max} = 5$) is significant. In addition, with the girth of NB-LDPC code equal to 14 as considered in this chapter, the constraint for this NB-LDPC code to become cycle-free-likely is $n_{\text{in}} \leq 6$ [85], which is easily satisfied by using the precoders ($n_{\text{in}}^{\max} = 3$ for GOPT and $n_{\text{in}}^{\max} = 5$ for max-$d_{\min}$).

6.5 Conclusion

This chapter has considered the mapping of multiple low-order GF symbols of NB-LDPC codewords onto a MIMO symbol at the transmitter to replace the conventional one-by-one high-order GF to MIMO symbol mapping, which requires high complexity for decoding when high order modulation is applied. The turbo detection scheme has been used at the receiver. Complexity analysis has been carried out to show the advantages of the considered system. We propose to use MIMO precoders to overcome the drawback of the system in terms of error-rate performance, especially at the error-floor. The simulation results, which match perfectly with the EXIT chart analysis, show that using precoders not only enhances the error-rate performance, but also further reduces the computational complexity at the receiver. Obviously, reducing (or maintaining) the complexity of NB-LDPC decoder, while improving the decoding performance in terms of error-rate, is necessary. Therefore, in the next chapter, we focus our study on low-complexity decoding algorithms for NB-LDPC codes.

Chapter 7

Multiple-Votes Parallel Symbol-Flipping Decoding Algorithm for Non-Binary LDPC Codes

The content of this chapter is mainly based on the following paper:

- Nhat-Quang Nhan, Telex M. N. Ngatched, Octavia A. Dobre, Philippe Rostaing, Karine Amis, and Emanuel Radoi. "Multiple-Votes Parallel Symbol-Flipping Decoding Algorithm for Non-Binary LDPC", In *IEEE Communication Letters*, pp. 905 - 908, Vol. 19, 2015.

7.1 Introduction

The results in Chapter 6 show that the computational complexity of decoding algorithm at the decoder plays a vital role in the total complexity of the NB-LDPC encoded systems. Therefore, the investigation into low complexity decoding algorithms is essential and will be considered in this chapter.

The algorithms for decoding NB-LDPC codes can be classified into three general categories: hard-decision decoding [92], soft-decision decoding [40, 42, 93, 94], and hybrid

decoding (also known as reliability-based decoding) [95–98]. From an implementation point of view, hard-decision decoding is the simplest in complexity. However, its simplicity results in a relatively poor performance that can be as far away as a few decibels from that of soft-decision decoding. Soft-decision decoding provides the best performance but requires the highest computational complexity. Hybrid decoding is in between the two extremes and provides a good trade-off between performance and complexity.

Among reliability-based message-passing algorithms, the parallel symbol flipping decoding (PSFD) algorithm recently introduced in [97] offers one of the best trade-off between performance and complexity. By using a flipping function which combines both the weighted check-based message of the normalized check-sums and the variable-based message of the received sequence, the algorithm identifies, in each decoding iteration, the relatively unreliable symbols in the hard-decision symbol sequence and decodes them based on the corresponding flipping symbols. The unreliable symbols are found using a voting system whereby each unsatisfied check node (CN) gives one vote to the variable node (VN) with the largest flipping function checked by it. Variable nodes thus accumulate votes and those with a total number of votes exceeding a predefined threshold are identified as unreliable. By only playing with the reliabilities of hard-decisions, PSFD is a good choice to decode high-rate finite-field NB-LDPC codes. However, simulation results show that PSFD is suitable for decoding only regular NB-LDPC codes whose parity-check matrices have large column weights, e.g., column weights of at least 8. In addition, PSFD employs a code-dependent voting threshold, which should be optimized through simulation. Furthermore, the computation of its flipping function involves a pair of scaling factors, (η, λ), which also depends on the code and is optimized through simulation.

This chapter proposes a new algorithm, referred to as multiple-votes PSFD (MV-PSFD). When compared with PSFD, it introduces a method of error estimation which results in avoiding the use of the voting threshold, and passes multiple votes from the unsatisfied CNs to the corresponding VNs. The proposed algorithm significantly outperforms PSFD for low column weight parity check matrices, as shown by simulation results. Such improvements are achieved with a similar computational complexity.

The organization of this chapter is as follows: In Section 7.2, some preliminaries and the PSFD algorithm are briefly reviewed. The proposed algorithm and its complexity

analysis are introduced in Section 7.3. Simulation results are presented in Section 7.4. Finally, Section 7.5 concludes the chapter.

7.2 Preliminaries

7.2.1 Notations and Definitions

A regular NB-LDPC code $\mathcal{C}$ of length N and dimension K over the Galois field of order q (GF(q)) is completely described by a row-column (RC)-constrained parity-check matrix of size $M \times N$ ($M \geq K$) $\mathbf{H} = [h_{j,i}]$ over GF(q), which has a constant column weight d_v and a constant row weight d_c. For practical purpose, we only consider binary extension fields, where $q = 2^p$. Let $\mathcal{M}_j = \{i : 0 \leq i < N, h_{ji} \neq 0\}$ be the set of all indices of VNs which connect to the CN c_j and $\mathcal{N}_i = \{j : 0 \leq j < M, h_{ji} \neq 0\}$ be the set of all indices of CNs which connect to the VN v_i.

Suppose that a regular (N, K) NB-LDPC code is used for error control over a binary-input additive white Gaussian noise (BIAWGN) channel with zero mean and two-sided power spectral density $N_0/2$. Assume binary phase-shift-keying (BPSK) signaling with unit energy. A codeword $\boldsymbol{c} = (c_0, c_1, \ldots, c_{N-1})$ in $\mathcal{C}$, where $c_n = (c_{n,0}, c_{n,1}, \cdots, c_{n,p-1}) \in \mathrm{GF}(2^p)$ with $c_{n,i} \in \mathrm{GF}(2)$, is mapped into the sequence $\boldsymbol{x} = (x_0, x_1, \ldots, x_{N-1})$ before its transmission, where $x_n = (x_{n,0}, x_{n,1}, \cdots, x_{n,p-1})$ with the mapping rule $x_{n,i} = (2c_{n,i} - 1) \in \{+1, -1\}$. Let $\boldsymbol{y} = (y_0, y_1, \ldots, y_{N-1})$ be the soft-decision received sequence at the output of the receiver matched filter. For $0 \leq i \leq N-1$, $y_n = (y_{n,0}, y_{n,1}, \cdots, y_{n,p-1}) = (x_{n,0} + \nu_{n,0}, x_{n,1} + \nu_{n,1}, \cdots, x_{n,p-1} + \nu_{n,p-1})$ in which the $\nu_{n,i}$'s are statistically independent Gaussian random variables with zero mean and variance $N_0/2$. The hard-decision symbol sequence at the decoder is denoted by $\mathbf{z} = (z_0, z_1, \ldots, z_{N-1})$.

For any hard-decision symbol sequence $\mathbf{z}$, the syndrome vector $\mathbf{s}$ is computed as $\mathbf{s} = \mathbf{z} \cdot \mathbf{H}^T$, where $s_j = h_{j0}z_0 + h_{j1}z_1 + \cdots + h_{j(N-1)}z_{N-1}$ with $0 \leq j < M$. s_j is known as the j^{th} check-sum of $\mathbf{z}$. $\mathbf{z}$ is a valid codeword in $\mathcal{C}$ if and only if $\mathbf{s} = \mathbf{0}$. Let $\mathcal{S}_i(z) = \{s_j : j \in \mathcal{N}_i\}$ be a set of check-sums that contain the symbol z_i at the VN v_i and $\tilde{\mathcal{S}}_i(z) = \{\tilde{s}_{ji}(z) = h_{ji}^{-1} s_j : s_j \in \mathcal{S}_i(z)\}$ be a normalized check-sum set. Finally, let $\tilde{\mathcal{S}}_i^*(z)$ be $\tilde{\mathcal{S}}_i(z)$ excluding zero elements, *i.e.*, $\tilde{\mathcal{S}}_i^*(z) = \tilde{\mathcal{S}}_i(z) \setminus \{0\}$.

7.2.2 PSFD Algorithm

Given the initial hard-decision symbol vector $\mathbf{z}^{(0)} = (z_0^{(0)}, z_1^{(0)}, \ldots, z_{N-1}^{(0)})$, the PSFD algorithm iteratively flips erroneous symbols in $\mathbf{z}^{(0)}$ until either a codeword is found or a maximum number of iterations is reached. In other words, the algorithm finds an error symbol vector $\mathbf{e} = [e_0, e_1, \ldots, e_{N-1}]$, such that $\mathbf{z} = \mathbf{z}^{(0)} - \mathbf{e}$ is a codeword. The error symbol vector is recursively computed by $\mathbf{e}^{(0)} = \mathbf{0}$ and $\mathbf{e}^{(l+1)} = \mathbf{e}^{(l)} + \boldsymbol{\varepsilon}^{(l)}$, where each component of $\boldsymbol{\varepsilon}^{(l)}$ is chosen so as to maximize a flipping function that is updated at each iteration and which is defined by

$$F_i^{(l)} = \max_{\alpha \in \tilde{\mathcal{S}}_i^{*(l)}(z)} F_i^{(l)}(\alpha). \tag{7.1}$$

In (7.1), $F_i^{(l)}(\alpha)$ with $\alpha \in GF(q)$ is given by

$$F_i^{(l)}(\alpha) = E_i^{(l)}(\alpha) - I_i(\alpha + e_i^{(l)}), \tag{7.2}$$

where the weighted check-based message about $z_i(l)$, $E_i^{(l)}(\alpha)$, and the variable-based message, $I_i(\alpha + e_i^{(l)})$, are respectively defined by

$$E_i^{(l)}(\alpha) = \sum_{j' \in \mathcal{N}_i : \tilde{s}_{j'i}^{(l)}(z) = \alpha} W_{j'i} - \sum_{j' \in \mathcal{N}_i : \tilde{s}_{j'i}^{(l)}(z) = 0} W_{j'i}, \tag{7.3}$$

and

$$I_i(\alpha) = \eta \phi_i(\alpha), \tag{7.4}$$

where η is a scaling factor. W_{ji} in (7.3) is the weighting coefficient contributed to z_i by other $z_{i'}$'s checked by s_j and is defined by

$$W_{ji} = \lambda \min_{i' \in \mathcal{M}_j \setminus \{i\}} R_{i'}, \tag{7.5}$$

where λ is a scaling factor. R_i is the reliability of the initial hard-decision symbol $z_i^{(0)}$ and is given by

$$R_i = \min_{\alpha \in GF(q)} \phi_i(\alpha), \tag{7.6}$$

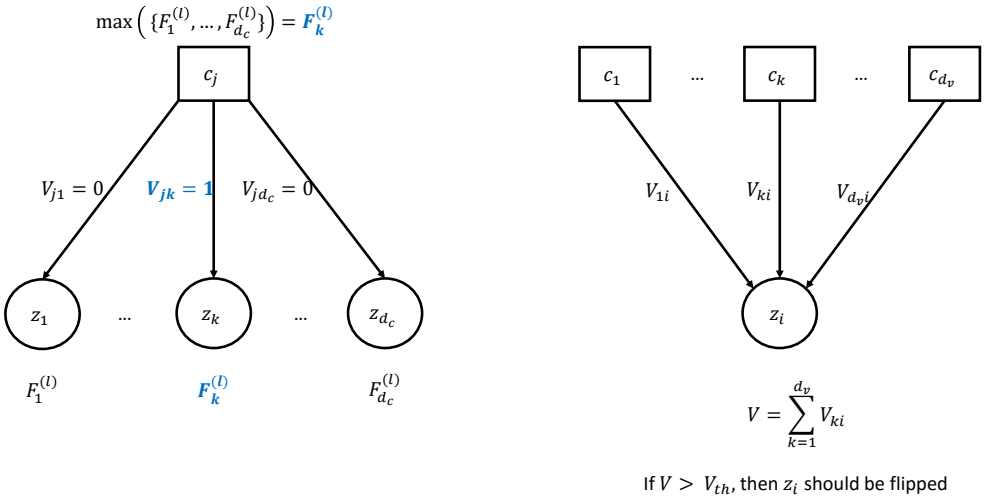


FIGURE 7.1: Conventional PSFD voting procedure.

with $\phi_i(\alpha)$ being a log-likelihood ratio representing the gap of likelihood between $z_i^{(0)}$ and $z_i^{(0)} - \alpha$ defined by

$$\phi_i(\alpha) = \frac{N_0}{4} \ln \frac{P(v_i = z_i^{(0)} | y_i)}{P(v_i = z_i^{(0)} - \alpha | y_i)}. \tag{7.7}$$

To speed up the decoding, it is essential to be able to flip multiple symbols in each decoding iteration. The PSFD algorithm achieves this by adopting a voting system whereby each unsatisfied CN gives one vote to the VN with the largest flipping function checked by it, as demonstrated in FIGURE 7.1. At the l^{th} iteration, VN v_i accumulates the number of votes, $\mathcal{V}_i^{(l)}$, from all the unsatisfied CNs. That is

$$\mathcal{V}_i^{(l)} = \sum_{j \in \mathcal{N}_i} V_{ji}^{(l)}, \tag{7.8}$$

where $V_{ji_0}^{(l)} = 1$ if $s_j^{(l)} \neq 0$; otherwise $V_{ji}^{(l)} = 0$ with $i_0 = \arg\max_{i' \in \mathcal{M}_j} F_{i'}^{(l)}$. The hard-decision $z_i^{(l)}$ will be flipped to $z_i^{(l+1)} = z_i^{(l)} - \varepsilon_i^{(l)}$ when $\mathcal{V}_i^{(l)} > V_{th}$ (as shown in FIGURE 7.2), where V_{th} is a code-dependent threshold.

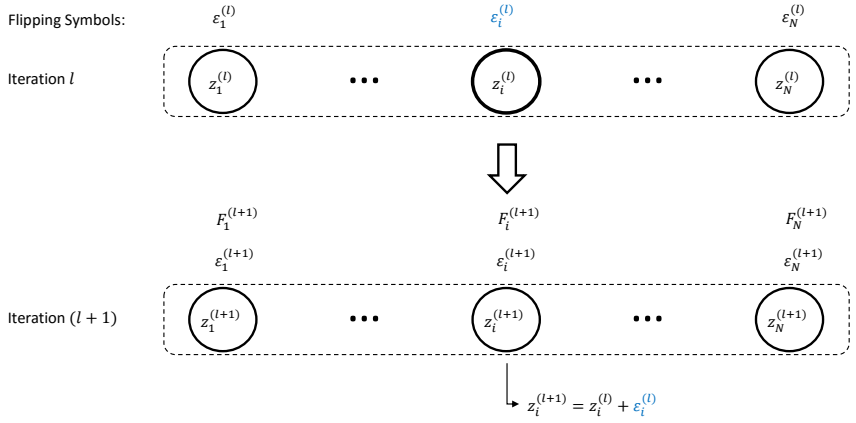


FIGURE 7.2: Conventional PSFD parallel flipping procedure.

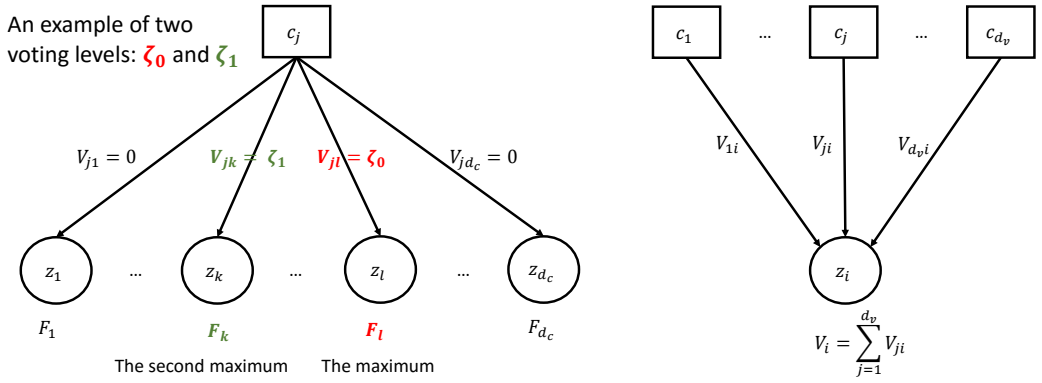


FIGURE 7.3: The proposed MV-PSFD voting procedure.

7.3 Multiple-Votes PSFD Algorithm

7.3.1 Algorithm

The voting threshold, V_{th}, in the PSFD is code-dependent and should be optimized through simulation [97]. In the sequel, we design a new decoding algorithm which mitigates this drawback while maintaining the strong aspects of the PSFD.

First, we introduce multiple voting levels, allowing each unsatisfied CN to pass more than one vote to the VNs checked by it. Without loss of generalization, let us introduce two voting levels $\zeta_0 > \zeta_1 > 0$. At each VN v_i, the voting function is still given by (7.8) but with the voting principle defined as follows

For $s_j^{(l)} \neq 0$,

- $V_{ji_0}^{(l)} = \zeta_0$ with $i_0 = \arg\max_{i' \in \mathcal{M}_j} F_{i'}^{(l)}$.
- $V_{ji_1}^{(l)} = \zeta_1$ with $i_1 = \arg\max_{i' \in \mathcal{M}_j \setminus \{i_0\}} F_{i'}^{(l)}$.
- $V_{ji}^{(l)} = 0$ elsewhere.

The unsatisfied CN c_j gives ζ_0 to the VN that has the largest flipping function and gives ζ_1 to the VN that has the second largest flipping function as shown in FIGURE 7.3. Though it might appear that the two factors ζ_0 and ζ_1 should be optimized, we will show through simulations, in the next section, that their optimal values are not code-dependent.

Secondly, we derive a relationship between the syndrome weight, the number of errors, and the parity-check matrix column weight d_v from the following Lemma [99].

Lemma 7. For regular LDPC matrices, the average syndrome weight increases linearly with the number of errors.

Proof. (Heuristic.) For any regular LDPC matrix, a plot of the average syndrome weight in terms of the number of errors yield a straight line with slope d_v. □

Based on the above Lemma, at iteration l, the number of errors in the hard-decision vector $\mathbf{z}^{(l)}$, and hence the number of symbols to be flipped, can be estimated by

$$N_e{}^{(l)} = \left\lfloor \frac{w^{(l)}(\mathbf{s})}{d_v} \right\rfloor, \tag{7.9}$$

where $\lfloor x \rfloor$ denotes the greatest integer less than or equal to x, and $w^{(l)}(\mathbf{s})$ is the syndrome weight at iteration l defined by

$$w^{(l)}(\mathbf{s}) = \sum_j (s_j^{(l)} \neq 0). \tag{7.10}$$

The use of a voting threshold can therefore be avoided by flipping the first $N_e{}^{(l)}$ symbols with the highest votes $\mathcal{V}_i^{(l)}$. This procedure is demonstrated in FIGURE 7.4.

The steps of the new algorithm, which we call Multiple-Votes Parallel Symbol-Flipping Decoding (MV-PSFD), can be summarized as shown in Algorithm 3.

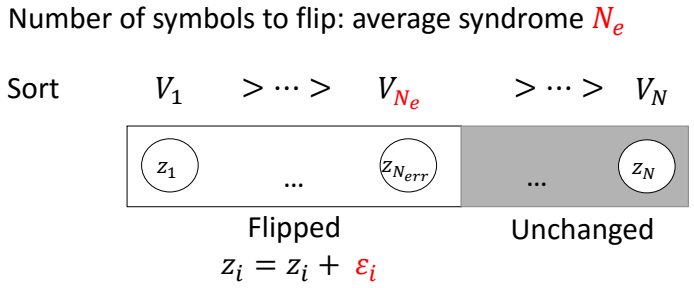


FIGURE 7.4: The proposed MV-PSFD parallel flipping procedure.

Algorithm 3 MV-PSFD algorithm

1: **Inputs:** Maximum number of iterations l_{max}, hard-decision $\mathbf{z}^{(0)}$ and the received symbol vector $\mathbf{y}$.
2: **Initialization:** Set iteration index $l = 0$, maximum number of iterations to l_{max}, and error symbol vector $\mathbf{e} = \mathbf{0}$. Find $\phi_i(\alpha)$ by (7.7) and the variable-based message $I_i(\alpha)$ by (7.4). Compute the coefficients W_{ji} by (7.5) and store them.
3: **while** $l \leq l_{max}$ **do**
4: Compute the syndrome $\mathbf{s}^{(l)}$
5: **if** $\mathbf{s}^{(l)} = \mathbf{0}$ **then**
6: Stop the while loop
7: **end if**
8: With $0 \leq i < N$, compute the flipping function $F_i^{(l)}$ and find its corresponding flipping symbol $\alpha = \varepsilon_i^{(l)}$ by (7.1)
9: Compute $\mathcal{V}_i^{(l)}$ by (7.8) using the multiple-votes principle
10: Estimate the number of errors ${N_e}^{(l)}$ by (7.9)
11: With $0 \leq i < N$, sort VNs according to descending order of $\mathcal{V}_i^{(l)}$
12: Flip the first ${N_e}^{(l)}$ VNs by setting new hard-decision symbols $z_i^{(l+1)} = z_i^{(l)} - \varepsilon_i^{(l)}$ and new error symbols $e_i^{(l+1)} = e_i^{(l)} + \varepsilon_i^{(l)}$
13: $l \leftarrow l + 1$
14: **end while**
15: **Output:** $\mathbf{z}^{(l)}$.

7.3.2 Complexity Computation

We evaluate the computational complexity of the new algorithm based on the number of integer addition (IA), multiplication (IM) and comparison (IC) operators as in [40].

At the initialization, the computations of $I_i(\alpha)$ and R_i (BPSK) need $\mathcal{O}(Nqlog_2(q))$ IAs, $\mathcal{O}(Nq)$ IMs, and $\mathcal{O}(N(log_2(q)-1))$ ICs. Additionally, $\mathcal{O}(M(2d_c-3))$ ICs and $\mathcal{O}(Md_c)$ IMs are needed for W_{ji}.

At each iteration, check-sums computations require $\mathcal{O}(M(d_c-1))$ IAs and $\mathcal{O}(Md_c)$ IMs. The normalized check-sums $\tilde{\mathcal{S}}_i^{*(l)}$ need $\mathcal{O}(Md_c)$ IMs. The flipping functions and flipping

TABLE 7.1: Complexity by operators at each iteration of NB-LDPC decoding algorithms.

	IA	**IM**	**IC**
PSFD [97]	3σ	2σ	$2\sigma - M$
MV-PSFD	$3\sigma + M - 1$	$2\sigma + 1$	$2\sigma - M - N + Nlog_2(N)$

IA: Integer Addition **IM**: Integer Multiplication **IC**: Integer Comparison
$\sigma = Md_c(= Nd_v)$: Number of non-zero elements in parity check matrix.

symbols require at most $\mathcal{O}(2N(d_v - 1))$ IAs for $F_i^{(l)}(\alpha)$ and at most $\mathcal{O}(N(d_v - 1))$ ICs for $F_i^{(l)}$. The error estimation needs at most $\mathcal{O}(M - 1)$ IAs for ${d_s}^{(l)}$ and $\mathcal{O}(1)$ IM for ${N_e}^{(l)}$. For voting function $\mathcal{V}_i^{(l)}$, it requires at most $\mathcal{O}(M(d_c - 1))$ ICs and at most $\mathcal{O}(M)$ IAs. If the Quick Sort algorithm [100] is used, the sorting needs at most $\mathcal{O}(Nlog_2(N))$ ICs. Note that the Quick Sort is not the best choice in terms of operators saving, but it is the fastest algorithm to achieve high system throughput. Finally, $\mathcal{O}(2N)$ IAs are used for $z_i^{(l+1)}$ and $e_i^{(l+1)}$.

The total computational complexity of MV-PSFD and PSFD algorithm for each iteration is shown in TABLE 7.1. From [97], it is proved that PSFD is the algorithm that costs the lowest decoding complexity among the reliability-based decoding algorithms. The MV-PSFD requires slightly more comparison operators than PSFD due to the sorting process. However, in practice, the cost of comparators is much cheaper than those of multiplications and additions. Thus, for small to moderate length codes, we can conclude that the complexity of MV-PSFD is close to the original PSFD.

7.4 Simulation Results

In this section, using Monte Carlo simulations, the error performance in terms of bit-error rate (BER), frame-error rate (FER), and average number of iterations as a function of the rate-normalized signal-to-noise ratio (SNR) per information bit (E_b/N_0), of the proposed algorithm is compared with that of the PSFD algorithm on a number of NB-LDPC codes of various constructions with short to medium block lengths. These constructions include Euclidean geometry [37], progressive edge-growth [101] and the original method of Gallager. The maximum number of iterations is set to 50 for both algorithms. At each SNR value, at least 100 erroneously received codewords are detected.

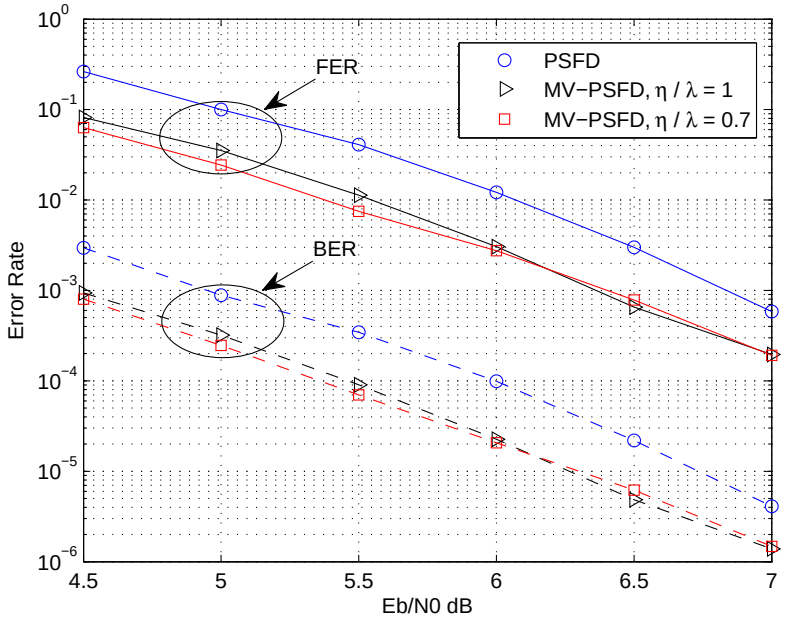


FIGURE 7.5: FER (solid) and BER (dashed) performance versus rate-normalized SNR of MV-PSFD and PSFD for Code 1 (low column weight).

The optimal values of η, λ, V_{th}, ζ_1, and ζ_0, are selected through simulations for each code. It should be noted that only the ratios $\frac{\eta}{\lambda}$ and $\frac{\zeta_1}{\zeta_0}$ are actually needed. An interesting result is that the optimum ratio $\frac{\zeta_1}{\zeta_0}$ is $\frac{2}{3}$ for all codes. Moreover, no significant difference in error performance was observed using $\frac{\zeta_1}{\zeta_0} = \frac{2}{3}$ and $\frac{\zeta_1}{\zeta_0} = 1$. Thus, the parameters ζ_1 and ζ_0 in MV-PSFD are not code-dependent.

We firstly present results for two codes, referred to as Code 1 and 2, whose parity-check matrices have small column weights, *i.e.*, the column weights are less than 8. Code 1 is a $(d_v = 3, d_c = 6)$ [102] and Code 2 is a $(d_v = 5, d_c = 10)$ [103]; both are regular (102, 204) NB-LDPC codes over GF(2^4) constructed based on Gallager's method, whose parity-check matrix satisfies the RC-constraint.

FIGURE 7.5 shows the BER and FER performances of Code 1. It can be observed that, unlike PSFD, the performance of the proposed algorithm is not sensitive to the parameters η and λ. Moreover, the proposed algorithm outperforms the PSFD by about 0.5 dB at the BER of 10^{-5}. Note that the PSFD is plotted with the optimized $\frac{\eta}{\lambda}$ and V_{th}.

FIGURE 7.6 shows the BER and FER performances of Code 2. Contrary to the previous case, the performance of the proposed algorithm is sensitive to the parameters η and λ.

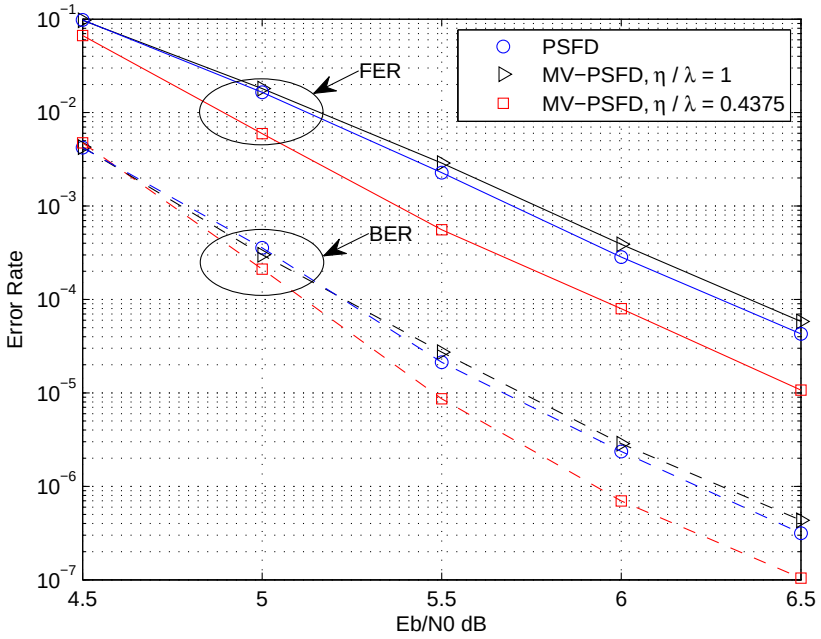


FIGURE 7.6: FER (solid) and BER (dashed) performance versus rate-normalized SNR of MV-PSFD and PSFD for Code 2 (low column weight).

However, using an optimized value of $\frac{\eta}{\lambda}$, the MV-PSFD outperforms the PSFD with a gain of about 0.3 dB at the BER of 10^{-5} and 0.4 dB at the BER of 10^{-6}. It can also be observed that, when MV-PSFD is used without the optimized values of η and λ, the performance is close to that of PSFD.

We would like to point out that, although the proposed algorithm could be generalized to more than two voting levels, no significant performance improvement was observed for more than two voting levels for all the codes simulated. This is illustrated by FIGURE 7.7 for Code 1. Therefore, taking into account the additional complexity, two voting levels appears to be the best choice for the proposed algorithm. Furthermore, for regular NB-LDPC codes whose parity-check matrices have large column weights, *i.e.*, column weights of at least 8, the proposed algorithm yields similar performance to the standard one. Indeed, let us take another (8,13)-regular (175, 255) NB-LDPC code over GF(2^4), whose parity-check matrix also satisfies the RC-constraint. This code is constructed by Progressive Edge Growth method. We refer to this code as Code 3. The FER and BER performances of Code 3 are shown in FIGURE 7.8. We observe that there is a slightly improvement of MV-PSFD in terms of FER compared to PSFD, while the performances of the two algorithms are similar in terms of BER.

With these performance studies in mind, we move on to the average number of iterations

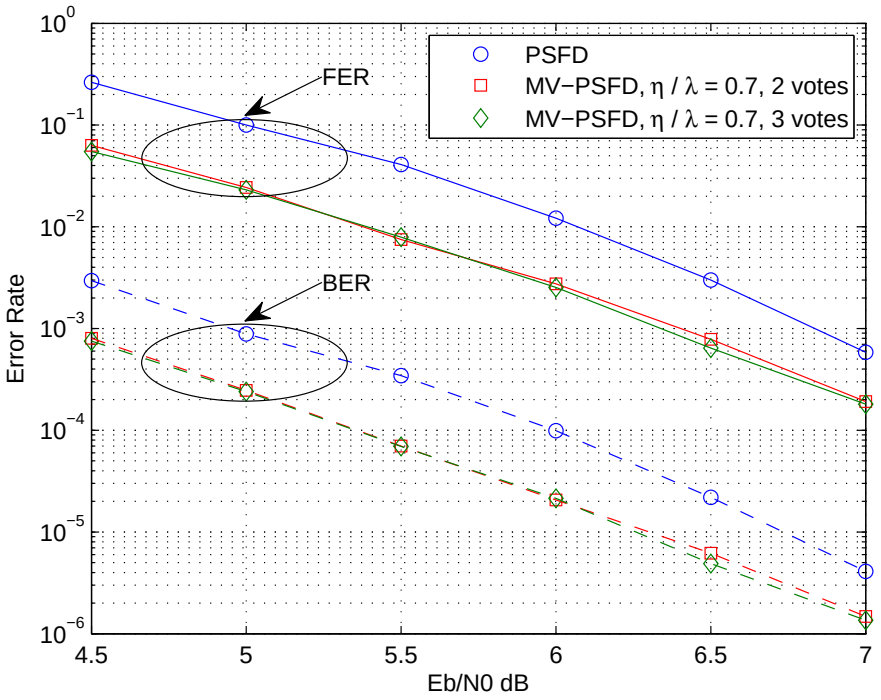


FIGURE 7.7: Voting levels comparison for Code 1 (low column weight).

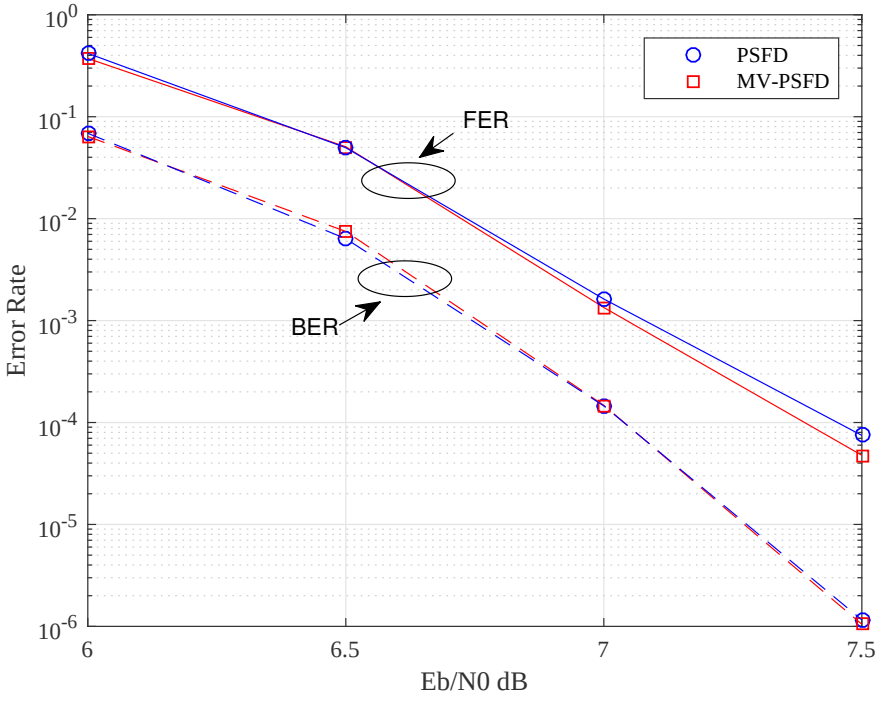


FIGURE 7.8: FER (solid) and BER (dashed) performance versus rate-normalized SNR of MV-PSFD and PSFD for Code 3 (high column weight).

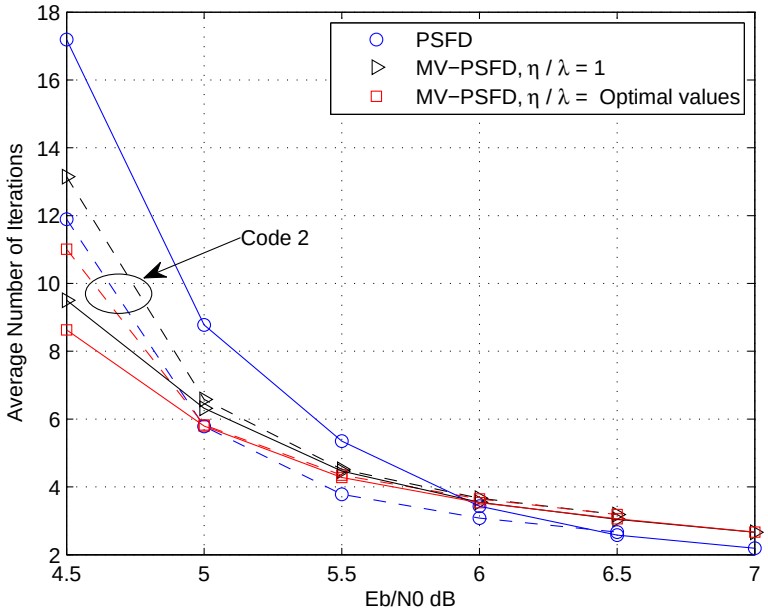


FIGURE 7.9: Average number of iterations versus rate-normalized SNR of MV-PSFD and PSFD for Code 1 (solid) and 2 (dashed).

comparison. FIGURE 7.9 depicts the average number of iterations for Codes 1 and 2. We see that, at low to medium SNR, the proposed algorithm with optimized parameters converges faster than the standard one. However, at high SNR, the convergence of the two algorithms is similar. Overall, one can conclude that there is no significant difference between the two algorithms in terms of convergence rate. As such, the excellent performance of the proposed algorithm is not obtained at the expense of the convergence rate.

7.5 Conclusion

In this chapter, we investigated the low complexity decoding algorithms for NB-LDPC codes in order to extend our study about complexity reduction for the NB-LDPC encoded MIMO communication systems. A new algorithm based on the PSFD algorithm for regular NB-LDPC codes has been proposed. Simulation results performed on a number of NB-LDPC codes of various column weights show that the proposed algorithm significantly outperforms the standard one for low column weight parity-check matrices, *i.e.*, column weights less than 8, with a similar complexity. Since we focus on the decoding

algorithm, single-input single-output (SISO) communication scheme with binary phase-shift keying (BPSK) is considered in this chapter for simplicity. However, the extension to MIMO scheme and to higher order modulation is straight forward. The MV-PSFD solution proposed in this chapter is a hybrid decoding algorithm with hard-output. For the perspective, we propose to apply the list decoding method [104] in order to have a hybrid decoding soft-output algorithm, which can be apply to the iterative receiving scheme considered in Chapter 6.

Conclusion and perspectives

Since the LTE-A standard is expected to continue playing a vital role in the 5G era [1], the studies in this thesis were targeted for LTE-A applications. As mentioned in the introduction, the most important requirements of LTE-A are high data rate and high quality of service, which assures low error-rate and low latency. Besides, low complexity is also an important criterion to design the next generation wireless communication systems. Throughout this thesis, we focused on the optimization of linear precoders for coded MIMO systems with various iterative receivers. We then optimized the error-rate performance while taking into account the complexity of these systems in order to meet the modern demand of LTE-A.

In Chapter 1, we briefly recalled the encoding and decoding algorithms for the FEC codes that were exploited throughout this thesis. In addition, the turbo detection and turbo equalization iterative receivers were presented to support for their later applications in the other chapters. Finally, EXIT chart, which is a good tool to analyze the convergence behavior of the iterative receivers, was introduced. In Chapter 2, the precoded MIMO systems along with channel transformation technique were introduced. The existing diagonal and non-diagonal precoders were presented and compared in order to find the best references for comparison with our new precoder propositions.

In Chapter 3, the concatenation of MIMO precoder with an outer FEC code assuming turbo detection at the receiver was investigated. We proposed a first new precoder for the usual scheme consisting of binary to Q-ary symbol conversion followed by Q-ary symbol to MIMO symbol conversion. This precoder aims to maximize the minimum Euclidean distance between symbols with binary mapped sequences differing by one position. In addition, we investigated a MIMO symbol mapper that directly maps the interleaved

FEC-encoded binary sequence into MIMO symbols, which allowed us to apply the maximum squared Euclidean weight (MSEW) mapping strategy on received constellation. Thanks to the EXIT chart analysis, we proposed a second novel precoder, which is adapted to be used with the MSEW mapping applied on the received constellation. Simulation results showed that, by using the new precoders, the improvement in terms of error-rate of the precoded MIMO system assuming turbo detection is significant.

Precoder optimization for finite alphabet signals over MIMO random channels was investigated in Chapter 4. We proposed a novel sub-optimal low-complexity precoding algorithm and compared it to an optimal one, which globally maximizes the channel mutual information. The new solution not only achieves a lower computational complexity but also avoids the use of initial values, which must be carefully selected for each channel and SNR for fast convergence in the case of GOPT. Another advantage of the new algorithm is that the resulting precoder has a fixed form of received constellation. This allowed us to optimize the symbol mapping on the received constellation. Simulations, in consistency with EXIT chart analysis, showed that the proposed low-complexity precoder achieves error-rate performance that is close to performance of the optimal one when the conventional Gray mapping is used. In addition, the new precoder used with optimized mapping at received constellation showed significant error-rate performance improvement. It is the best precoder for the turbo detection scheme introduced in Section 1.2.2. This precoder is more complex than $\mathbf{F}_{\ell_1}$ and $\mathbf{F}_{\ell_1}$-mod due to the power optimization at each SNR and channel realization. However, the complexity of this precoder is significantly reduced compared to GOPT.

In Chapter 5, we focused on the optimization of a linear MIMO precoder assuming an outer forward error correction code and an iterative MMSE-based interference cancellation (turbo equalization) at the receiver side. Given perfect channel state information at both sides of the communication, we proposed a novel precoder that is specifically designed to use with turbo equalization. In contrast to the conventional precoders that maximize the mutual information between channel input and channel output symbols (channel capacity), the proposed precoder aims to maximize the mutual information between the channel input and symbols at output of the equalizer. The precoder was targeted for applications that require low complexity, where simple FEC codes are used. Simulation results showed the error-rate performance gain of the resulting precoder compared to two other reference precoders presented in the literature, which are derived from

the maximization of channel capacity.

MIMO systems encoded by NB-LDPC codes were investigated in Chapter 6. Assuming high-order modulation, a group of low-order GF codeword symbols were mapped onto one Q-ary modulated symbol vector (multiple mapping), instead of mapping one high-order GF symbol (one-by-one mapping). At the receiver, the iterative detection-decoding of NB-LDPC codes was considered. It was shown that the multiple mapping technique leads to significant complexity reduction at the receiver. MIMO precoders were introduced to not only further reduce the computational complexity of the iterative receiver but also significantly enhance error-rate performance of the system compared to the spatial multiplexing scheme. EXIT chart was exploited to analyze the iterative receiver performance. Finally, simulation results were presented to assess our analysis.

To complete our study about the complexity reduction for the communication systems that use NB-LDPC codes, we proposed in Chapter 7 a novel decoding algorithm for NB-LDPC codes. The algorithm builds on the recently designed parallel symbol-flipping decoding (PSFD) algorithm and combines a technique of error estimation and a method of multiple voting levels from each unsatisfied check-sum to the corresponding variable nodes. Simulations results, performed on a number of NB-LDPC codes of various lengths and column weights constructed using several methods, showed that the new algorithm not only avoids using code-dependent voting threshold but also improves the error rate performance of the PSFD algorithm, particularly for low column weight parity-check matrices.

The wireless communication systems for the next generation mobile networks employ a high number of antennas at the transceivers (commonly referred to as massive MIMO systems). Therefore, the increase of number of data streams is important. As future work, we could adapt the proposed precoding solutions to high data streams according to two directions. On the first direction, by taking into account the systems models considered in this thesis, we could look for other precoders that are optimized for $b > 2$ (note that this has been done for LCOPT precoder in Chapter 4 but there is still room for improvement). However, optimizing precoder for a very high number of b could lead to high complexity for the practical design of precoder. Therefore, as the second direction, we propose to apply the channel transformation and split the MIMO system into multiple subsystems in order to apply the precoding solutions proposed in this thesis.

For example, with $b = 4$, we propose to precode the virtual channel by a block diagonal matrix $\mathbf{F}'_d = [\mathbf{F}_{d1} \quad \mathbf{0}; \mathbf{0} \quad \mathbf{F}_{d2}]$, where $\mathbf{F}_{d1}$, $\mathbf{F}_{d2}$ are the 2×2 precoding matrices that were optimized during this thesis.

The extension to higher order modulation (*e.g.* $Q = 16$) is also interesting. With higher order modulation the complexity of soft-demapper is dramatically increased. Hence, we must focus on the turbo equalization scheme in order to reduce the complexity. However, the extension of the proposed precoder for 4-QAM to higher order modulation is not straightforward. Indeed, with high order modulation, applying the Genie-optimized precoder proposed in Chapter 5 (which maximizes the mutual information between channel input and output of interference canceller) leads to a very low starting point in the EXIT chart, which causes non-convergence of the iterative receiver. We propose to investigate three solutions. The first one would search the best mapping to increase the initial mutual information at the detector (interference canceller revolves around SBC and BSC) output. The second one would focus on the precoder definition. We could mix the max-$d_{\min}$ and the Genie-optimized precoder. The third one would consider the receiver structure. We could use a MAP detection at first iteration and a MMSE IC for the following iterations of a iterative receiver.

Throughout this thesis, we assume perfect CSI at the transmitter. In practical applications, the imperfect CSI at the transmitter can lead to performance loss. Therefore, the extension of the studies in this thesis to imperfect CSI scheme could also be interesting for future work.

The results from this thesis immediately apply to MIMO-OFDM (orthogonal frequency division multiplexing) and its extension to frequency selective channels assuming single carrier, precoding and detection in the frequency domain could be investigated.

List of Publications

International peer-reviewed conferences

- **Nhat-Quang Nhan**, Philippe Rostaing, Karine Amis, Ludovic Collin, and Emanuel Radoi. "Association and joint optimization of max-dmin precoder with MIMO turbo equalization". In *IEEE Global Communications Conference (GLOBECOM)*, pages 1 - 6, 2015.

- **Nhat-Quang Nhan**, Philippe Rostaing, Karine Amis, Ludovic Collin, and Emanuel Radoi. "Optimized MIMO symbol mapping to improve the turbo cliff region of iterative precoded MIMO detection". In *the 2015 European Signal Processing Conference (EUSIPCO)*, pages 909 - 913, 2015.

- **Nhat-Quang Nhan**, Philippe Rostaing, Karine Amis, Ludovic Collin, and Emanuel Radoi. "Optimized maxdmin precoder assuming maximum squared Euclidean weight-mapping and turbo detection". To appear in *the International Symposium on Turbo Codes and Iterative Receivers (ISTC)*, 2016.

International peer-reviewed journals

- **Nhat-Quang Nhan**, Telex M. N. Ngatched, Octavia A. Dobre, Philippe Rostaing, Karine Amis, and Emanuel Radoi. "Multiple-Votes Parallel Symbol-Flipping Decoding Algorithm for Non-Binary LDPC", In *IEEE Communication Letters*, pp. 905 - 908, Vol. 19, 2015.

- **Nhat-Quang Nhan**, Philippe Rostaing, Karine Amis, Ludovic Collin, and Emanuel Radoi. "Optimization of linear MIMO precoding assuming MMSE-based turbo

equalization". Submitted to *IEEE Transaction on Wireless Communications*, pending for review.

- **Nhat-Quang Nhan**, Philippe Rostaing, Karine Amis, Emanuel Radoi, and Y. Rosa Zheng. "Turbo detection of NB-LDPC codes in precoded MIMO systems" Submitted to *IEEE Transaction on Vehicular Technology*, pending for review.

- **Nhat-Quang Nhan**, Philippe Rostaing, Karine Amis, Ludovic Collin, and Emanuel Radoi. "Joint Optimization of MIMO Precoding and Symbol Mapping for Turbo Detection". Submitted to *IEEE Transaction on Wireless Communications*, pending for review.

- **Nhat-Quang Nhan**, Philippe Rostaing, Karine Amis, Ludovic Collin, and Emanuel Radoi. "Complexity Reduction for the Optimization of Linear Precoders over Random MIMO Channel". Submitted to *IEEE Transaction on Wireless Communications*, pending for review.

International Collaborations

The PhD research project offered the opportunity of two international collaborations initiated through the following scientific stays

- Two-month stay at the Missouri University of Science and Technology, USA, in cooperation with Prof. Yahong Rosa Zheng and Prof. Chengshan Xiao. The research results are presented in Chapter 6.
- Two-month stay at the Memorial University of Newfoundland, Canada, in cooperation with Prof. Octavia Dobre and Prof. Telex M. N. Ngatched. The research results are presented in Chapter 7.

Bibliography

[1] Boyd Bangerter, Shilpa Talwar, Reza Arefi, and Kyle Stewart. Networks and devices for the 5G era. *IEEE Commun. Magazine*, 52(2):90–96, 2014.

[2] Christopher Cox. *An introduction to LTE: LTE, LTE-advanced, SAE and 4G mobile communications*. John Wiley & Sons, 2012.

[3] Amitabha Ghosh and Rapeepat Ratasuk. *Essentials of LTE and LTE-A*. Cambridge University Press, 2011.

[4] Khan Farooq. *LTE for 4G mobile broadband*. Cambridge University Press, 2009.

[5] Ian F Akyildiz, David M Gutierrez-Estevez, and Elias Chavarria Reyes. The evolution to 4G cellular systems: LTE-advanced. *Elsevier Physical Commun.*, 3(4): 217–244, 2010.

[6] Akhil Gupta and Rakesh Kumar Jha. A survey of 5G network: architecture and emerging technologies. *IEEE Access*, 3:1206–1232, 2015.

[7] Mamta Agiwal, Abhishek Roy, and Navrati Saxena. Next generation 5G wireless networks: a comprehensive survey. *IEEE Commun. Surveys & Tutorials*, 2016.

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

[9] Emre Telatar. Capacity of multi-antenna gaussian channels. *European Trans. Telecommun.*, 10(6):585–595, 1999.

[10] Q. Li, G. Li, W. Lee, M. Lee, D. Mazzarese, B. Clerckx, and Z. Li. MIMO techniques in WiMAX and LTE: a feature overview. *IEEE Commun. Mag.*, 48(5): 86–92, 2010.

[11] A. J. Paulraj, D. A. Gore, R. U. Nabar, and H. Bolcskei. An overview of MIMO communications - a key to gigabit wireless. *Proc. of the IEEE*, 92(2):198–218, 2004.

[12] Siavash Alamouti. A simple transmit diversity technique for wireless communications. *IEEE J. Sel. Areas Commun.*, 16(8):1451–1458, 1998.

[13] Vahid Tarokh, Hamid Jafarkhani, and A Robert Calderbank. Space-time block coding for wireless communications: performance results. *IEEE J. Sel. Areas in Commun.*, 17(3):451–460, 1999.

[14] P Bouvet, Maryline Hélard, and Vincent Le Nir. Low complexity iterative receiver for linear precoded MIMO systems. In *IEEE Symposium on Spread Spectrum Techniques and Applications*, pages 17–21, 2004.

[15] P Bouvet, Maryline Hélard, Jerome Le Masson, and Charlotte Langlais. Iterative receiver for linear precoded MIMO systems. In *IEEE ITG-Conference on Source and Channel Coding*, pages 1–6, 2006.

[16] Ludovic Collin, Olivier Berder, Philippe Rostaing, and Gilles Burel. Optimal minimum distance-based precoder for MIMO spatial multiplexing systems. *IEEE Trans. Signal Processing*, 52(3):617–627, 2004.

[17] Quoc-Tuong Ngo, Olivier Berder, Baptiste Vrigneau, and Olivier Sentieys. Minimum distance based precoder for MIMO-OFDM systems using a 16-QAM modulation. In *IEEE ICC*, pages 1–5, 2009.

[18] Quoc-Tuong Ngo, Olivier Berder, and Pascal Scalart. Neighbor-dmin precoder for three data-stream MIMO systems. In *The 19th European Signal Processing Conf.*, pages 81–85, 2011.

[19] Quoc-Tuong Ngo, Olivier Berder, and Pascal Scalart. Minimum euclidean distance based precoders for MIMO systems using rectangular QAM modulations. *IEEE Trans. Signal Processing*, 60(3):1527–1533, 2012.

[20] Quoc-Tuong Ngo, Olivier Berder, and Pascal Scalart. Minimum euclidean distance-based precoding for three-dimensional multiple input multiple output spatial multiplexing systems. *IEEE Trans. on Wireless Commun.*, 11(7):2486–2495, 2012.

[21] Quoc-Tuong Ngo, Olivier Berder, and Pascal Scalart. General minimum euclidean distance-based precoder for MIMO wireless systems. *EURASIP J. Advances in Signal Processing*, 2013(1):1–12, 2013.

[22] Chengshan Xiao, Yahong Rosa Zheng, and Zhi Ding. Globally optimal linear precoders for finite alphabet signals over complex vector gaussian channels. *IEEE Trans. Signal Processing*, 59(7):3301–3314, 2011.

[23] P. Rostaing, O. Berder, G. Burel, and L. Collin. Minimum BER diagonal precoder for MIMO digital transmissions. *Signal Processing*, 82(10):1477–1480, 2002.

[24] Hemanth Sampath, Petre Stoica, and Arogyaswami Paulraj. Generalized linear precoder and decoder design for MIMO channels using the weighted MMSE criterion. *IEEE Trans. Commun.*, 49(12):2198–2206, 2001.

[25] Petre Stoica and Girish Ganesan. Maximum-SNR spatial-temporal formatting designs for MIMO channels. *IEEE Trans. Signal Processing*, 50(12):3036–3042, 2002.

[26] Angel Lozano, Antonia Maria Tulino, and Sergio Verdú. Optimum power allocation for parallel gaussian channels with arbitrary input distributions. *IEEE Trans. Info. Theory*, 52(7):3033–3051, 2006.

[27] M Sayed Hassan and Karine Amis. On the design of full-rate full-diversity space-time block codes for MIMO systems with a turbo minimum mean square error equaliser at the receiver side. *IET Communications*, 6(18):3065–3074, 2012.

[28] Lalit Bahl, John Cocke, Frederick Jelinek, and Josef Raviv. Optimal decoding of linear codes for minimizing symbol error rate. *IEEE Trans. on Info. Theory*, 20 (2):284–287, 1974.

[29] Sergio Benedetto, Dariush Divsalar, Guido Montorsi, and Fabrizio Pollara. Soft-output decoding algorithms for continuous decoding of parallel concatenated convolutional codes. In *IEEE ICC*, volume 1, pages 112–117, 1996.

[30] Robert G Gallager. Low-density parity-check codes. *IRE Trans. on Info. Theory*, 8(1):21–28, 1962.

[31] David JC MacKay and Radford M Neal. Good codes based on very sparse matrices. In *Cryptography and Coding*, pages 100–111. 1995.

[32] David JC MacKay. Good error-correcting codes based on very sparse matrices. *IEEE Trans. on Info. Theory*, 45(2):399–431, 1999.

[33] Matthew C Davey and David JC MacKay. Low-density parity-check codes over GF(q). *IEEE Commun. Lett.*, 2(6):165–167, Jun. 1998.

[34] Matthew C Davey and David JC MacKay. Low density parity check codes over GF (q). In *Info. Theory Workshop*, pages 70–71, 1998.

[35] Dai Kimura, Frédéric Guilloud, and Ramesh Pyndiah. Application of non-binary LDPC codes for small packet transmission in vehicle communications. In *The 5th Int. conf. ITS Telecommun.*, pages 1–4, 2005.

[36] Xiao-Yu Hu, Evangelos Eleftheriou, and Dieter-Michael Arnold. Regular and irregular progressive edge-growth tanner graphs. *IEEE Trans. on Info. Theory*, 51 (1):386–398, 2005.

[37] Lingqi Zeng, Lan Lan, Ying Yu Tai, Bo Zhou, Shu Lin, and Khaled AS Abdel-Ghaffar. Construction of nonbinary cyclic, quasi-cyclic and regular LDPC codes: a finite geometry approach. *IEEE Trans. on Commun.*, 56(3):378–387, 2008.

[38] D Costello and Shu Lin. *Error control coding*. Pearson Higher Education, 2004.

[39] LDPC toolkit for MATLAB. URL http://arun-10.tripod.com/ldpc/encode.html.

[40] Henk Wymeersch, Heidi Steendam, and Marc Moeneclaey. Log-domain decoding of LDPC codes over GF (q). In *IEEE ICC*, volume 2, pages 772–776, 2004.

[41] Hongxin Song and JR Cruz. Reduced-complexity decoding of Q-ary LDPC codes for magnetic recording. *IEEE Trans. on Magnetics*, 39(2):1081–1087, 2003.

[42] David Declercq and Marc Fossorier. Decoding algorithms for nonbinary LDPC codes over GF. *IEEE Trans. on Commun.*, 55(4):633–643, 2007.

[43] A. Picart, P. Didier, and A. Glavieux. Turbo-detection: a new approach to combat channel frequency selectivity. In *IEEE Int. Conf. on Commun. Towards the Knowledge Millennium*, volume 3, pages 1498–1502, 1997.

[44] S. Haykin, M. Sellathurai, Y. de Jong, and T. Willink. Turbo-MIMO for wireless communications. *IEEE Commun. Mag.*, 42(10):48–53, 2004.

[45] C. Berrou and A. Glavieux. Near optimum error correcting coding and decoding: turbo-codes. *IEEE Trans. on Commun.*, 44(10):1261–1271, 1996.

[46] Catherine Douillard, Michel Jézéquel, Claude Berrou, Département Electronique, Annie Picart, Pierre Didier, and Alain Glavieux. Iterative correction of intersymbol interference: turbo-equalization. *European Trans. Telecommun.*, 6(5):507–511, 1995.

[47] Xiaodong Wang and H Vincent Poor. Iterative (turbo) soft interference cancellation and decoding for coded CDMA. *IEEE Trans. Commun.*, 47(7):1046–1061, 1999.

[48] Melanie Witzke, Stephan Baro, Frank Schreckenbach, and Joachim Hagenauer. Iterative detection of MIMO signals with linear detectors. In *IEEE Asilomar Conf. Signals, Systems and Computers*, volume 1, pages 289–293, 2002.

[49] Xavier Wautelet, Antoine Dejonghe, and Luc Vandendorpe. MMSE-based fractional turbo receiver for space-time BICM over frequency-selective MIMO fading channels. *IEEE Trans. Signal Processing*, 52(6):1804–1809, 2004.

[50] Raphaël Visoz, Antoine O Berthet, and Sami Chtourou. A new class of iterative equalizers for space-time BICM over MIMO block fading multipath AWGN channel. *IEEE Trans. Commun.*, 53(12):2076–2091, 2005.

[51] Karine Amis, L Le Josse, and Christophe Laot. Efficient frequency-domain MMSE turbo equalization derivation and performance comparison with the time-domain counterpart. In *IEEE Int. Conf. on Wireless and Mobile Commun.*, pages 65–65, 2007.

[52] Karine Amis, Guillaume Sicot, and Dominique Leroux. Reduced complexity near-optimal iterative receiver for WiMAX full-rate space-time code. In *IEEE Int. Symp. on Turbo Codes and Related Topics*, pages 102–106, 2008.

[53] Stephan ten Brink. Convergence of iterative decoding. *IET Electronics Letters*, 35 (10):806–808, 1999.

[54] Stephan ten Brink, Joachim Speidel, and Ran-Hong Yan. Iterative demapping and decoding for multilevel modulation. In *IEEE GLOBECOM*, volume 1, pages 579–584, 1998.

[55] S. ten Brink. Convergence behavior of iteratively decoded parallel concatenated codes. *IEEE Trans. Commun.*, 49(10):1727–1737, 2001.

[56] Robert G Maunder and Lajos Hanzo. Iterative decoding convergence and termination of serially concatenated codes. *IEEE Trans. Veh. Tech.*, 59(1):216–224, 2010.

[57] J. Hagenauer. The EXIT chart-introduction to extrinsic information transfer in iterative processing. In *Proc. 12th EUSIPCO*, pages 1541–1548, 2004.

[58] Fredrik Brannstrom, Lars K Rasmussen, and Alex J Grant. Convergence analysis and optimal scheduling for multiple concatenated codes. *IEEE Trans. on Info. Theory*, 51(9):3354–3364, 2005.

[59] Gerard J Foschini and Michael J Gans. On limits of wireless communications in a fading environment when using multiple antennas. *Wireless personal commun.*, 6 (3):311–335, 1998.

[60] Sumeet Sandhu, Robert Heath, and Arogyaswami Paulraj. Space-time block codes versus space-time trellis codes. In *IEEE ICC*, volume 4, pages 1132–1136, 2001.

[61] Mai Vu and Arogyaswami Paulraj. MIMO wireless linear precoding. *IEEE Signal Processing Magazine*, 24(5):86–105, 2007.

[62] Fernando Pérez-Cruz, Miguel RD Rodrigues, and Sergio Verdú. MIMO gaussian channels with arbitrary inputs: optimal precoding and power allocation. *IEEE Trans. Info. Theory*, 56(3):1070–1084, 2010.

[63] NGO Quoc-Tuong. Generalized minimum euclidean distance based precoders for mimo spatial multiplexing systems. In *Ph.D. dissertation, IRISA, University of Rennes 1, Rennes*, 2012.

[64] Daniel Pérez Palomar and Sergio Verdú. Gradient of mutual information in linear vector gaussian channels. *IEEE Trans. on Info. Theory*, 52(1):141–154, 2006.

[65] Sarah J Johnson. *Iterative error correction: turbo, low-density parity-check and repeat-accumulate codes.* Cambridge University Press, 2009.

[66] J. Tan and G. L. Stuber. Analysis and design of symbol mappers for iteratively decoded BICM. *IEEE Trans. on Wireless Commun.*, 4(2):662–672, 2005.

[67] Kaisa Miettinen. *Nonlinear multiobjective optimization*, volume 12. Springer, 1999.

[68] Weiliang Zeng, Chengshan Xiao, and Jianhua Lu. A low-complexity design of linear precoding for MIMO channels with finite-alphabet inputs. *IEEE Wireless Commun. Lett.*, 1(1):38–41, 2012.

[69] I.D. Coope and P.F. Renaud. Trace inequalities with applications to orthogonal regression and matrix nearness problems. *Journal of Inequaities in Pure and Applied Math*, 10(4), 2009.

[70] Marcus Grossmann. SVD-based precoding for single carrier mimo transmission with frequency domain MMSE turbo equalization. *IEEE Signal Processing Lett.*, 16(5):418–421, 2009.

[71] Juha Karjalainen, Marian Codreanu, Antti Tölli, Markku Juntti, and Tad Matsumoto. EXIT chart-based power allocation for iterative frequency domain MIMO detector. *IEEE Trans. Signal Processing*, 59(4):1624–1641, 2011.

[72] Valtteri Tervo, Antti Tolli, Juha Karjalainen, and Tad Matsumoto. Transmission power variance constrained power allocation for iterative frequency domain multiuser SIMO detector. In *IEEE ICASSP*, pages 3493–3497, 2014.

[73] Jérôme Le Masson, Charlotte Langlais, and Claude Berrou. Linear precoding with low complexity MMSE turbo-equalization and application to the wireless LAN system. In *IEEE ICC*, volume 4, pages 2352–2356, 2005.

[74] Kimmo Kansanen and Tad Matsumoto. An analytical method for MMSE MIMO turbo equalizer EXIT chart computation. *IEEE Trans. Wireless Commun.*, 6(1): 59–63, 2007.

[75] Seok-Jun Lee, Andrew C Singer, and Naresh R Shanbhag. Linear turbo equalization analysis via BER transfer and EXIT charts. *IEEE Trans. Signal Processing*, 53(8):2883–2897, 2005.

[76] Walter Gander and Jiri Hrebicek. *Solving problems in scientific computing using Maple and Matlab®*. Springer Science & Business Media, 2004.

[77] Stephan ten Brink. Designing iterative decoding schemes with the extrinsic information transfer chart. *AEU Int. J. Electron. Commun*, 54(6):389–398, 2000.

[78] Kan Zheng, Long Zhao, Jie Mei, Bin Shao, Wei Xiang, and Lajos Hanzo. Survey of large-scale MIMO systems. *IEEE Commun. Surveys and Tutorials*, 2015.

[79] Stephan Pfletschinger and David Declercq. Getting closer to MIMO capacity with non-binary codes and spatial multiplexing. In *IEEE GLOBECOM*, pages 1–5, 2010.

[80] Stephan Pfletschinger and David Declercq. Non-binary coding for vector channels. In *IEEE SPAWC*, pages 26–30, 2011.

[81] Jonathan Rodriguez. *Fundamentals of 5G Mobile Networks*. John Wiley & Sons, 2015.

[82] Haesik Kim. Coding and modulation techniques for high spectral efficiency transmission in 5G and satcom. In *EUSIPCO*, volume 1, pages 2746–2750, 2015.

[83] Osamah Alamri, Soon Xin Ng, Feng Guo, S Zummo, and Lajos Hanzo. Sphere-packing modulated space-time coding using non-binary LDPC-coded iterative-detection. In *IEEE WCNC*, pages 106–111, 2008.

[84] Ronghui Peng and Rong-Rong Chen. Application of nonbinary LDPC cycle codes to MIMO channels. *IEEE Trans. Wireless Commun.*, 7(6):2020–2026, 2008.

[85] Nicholas B Chang and David L Romero. Non-binary coded modulation and iterative detection for high spectral efficiency in MIMO. In *IEEE ASILOMAR*, pages 458–462, 2012.

[86] Michele Franceschini, Gianluigi Ferrari, and Riccardo Raheli. *LDPC coded modulations*. Springer, 2009.

[87] Henk Wymeersch, Heidi Steendam, and Marc Moeneclaey. Interleaved coded modulation for non-binary codes: a factor graph approach. In *IEEE GLOBECOM*, volume 1, pages 525–529, 2004.

[88] Tarek Chehade, Ludovic Collin, Philippe Rostaing, Oussama Bazzi, and Emanuel Radoi. Adaptive minimum-distance based precoder for NB-LDPC coded MIMO transmission. In *IEEE GLOBECOM*, Dec. 2015.

[89] Charly Poulliat, Marc Fossorier, and David Declercq. Design of regular (2, d_c)-LDPC codes over GF(q) using their binary images. *IEEE Trans. Commun.*, 56 (10):1626–1635, 2008.

[90] Stephan Pfletschinger, Alain Mourad, Eduardo Lopez, David Declercq, and Giacomo Bacci. Performance evaluation of non-binary LDPC codes on wireless channels. *ICT Mobile Summit*, 2009.

[91] Jörg Kliewer, Soon Xin Ng, and Lajos Hanzo. On the computation of EXIT characteristics for symbol-based iterative decoding. In *Int. Symp. on Turbo Codes&Related Topics*, pages 1–6, 2006.

[92] Xinmiao Zhang, Fang Cai, and Shu Lin. Low-complexity reliability-based message-passing decoder architectures for non-binary LDPC codes. *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, 20(11):1938–1950, Jul. 2012.

[93] Emmanuel Boutillon and Laura Conde-Canencia. Bubble check: a simplified algorithm for elementary check node processing in extended min-sum non-binary LDPC decoders. *IET Elect. Lett.*, 46(9):633–634, Apr. 2010.

[94] Chung-Li Wang, Xiaoheng Chen, Zongwang Li, and Shaohua Yang. A simplified min-sum decoding algorithm for non-binary LDPC codes. *IEEE Trans. Commun.*, 61(1):24–32, Feb. 2013.

[95] Chao Chen, Baoming Bai, Xiao Ma, and Xinmei Wang. A symbol-reliability based message-passing decoding algorithm for non-binary LDPC codes over finite fields. In *Proc. IEEE ISTC*, pages 251–255, 2010.

[96] Bing Liu, Jun Gao, Wei Tao, and Gaoqi Dou. Weighted symbol-flipping decoding algorithm for non-binary LDPC codes with flipping patterns. *Journal of System Eng. and Elect.*, 22(5):848–855, Oct. 2011.

[97] Chao-Cheng Huang, Chi-Jen Wu, Chao-Yu Chen, and Chi-chao Chao. Parallel symbol-flipping decoding for non-binary LDPC codes. *IEEE Commun. Letters*, 17 (6):1228–1231, 2013.

[98] F. Garia-Herrero and David Declercq. Non-binary LDPC decoder based on symbol flipping with multiple votes. *IEEE Commun. Lett.*, 18(5):749–752, May 2014.

[99] Martin Bossert. *Channel Coding for Telecommunications.* John Wiley & Sons, Inc., 1999.

[100] Charles AR Hoare. Quicksort. *The Computer Journal, Br Computer Soc*, 5(1): 10–16, Nov. 1962.

[101] Xiao-Yu Hu, Evangelos Eleftheriou, and D-M Arnold. Progressive edge-growth Tanner graphs. In *Proc. IEEE Globecom*, volume 2, pages 995–1001, 2001.

[102] David Mackay's code resource, . URL http://www.inference.phy.cam.ac.uk/mackay/codes/EN/C/204.33.486.

[103] David Mackay's code resource, . URL http://www.inference.phy.cam.ac.uk/mackay/codes/EN/C/204.55.153.

[104] Carrasco Rolando Antonio and Johnston Martin. Non-binary error control coding for wireless communication and data storage, 2008.

Résumé étendu du mémoire de thèse :

Optimisation de précodeurs linéaires pour les systèmes MIMO codés, à récepteurs itératifs

Préparé par

NHAN Nhat-Quang

Lab-STICC, CNRS UMR 6285
Université de Bretagne Occidentale & Telecom Bretagne
Brest, France, 2013-2016

Introduction

L'augmentation spectaculaire des utilisateurs de systèmes mobiles et le développement des réseaux sociaux ces dernières années rendent nécessaire l'amélioration continue de la capacité des réseaux cellulaires radio modernes. Les normes LTE (Long Term Evolution) et LTE-A (LTE-Advanced), élaborées en 2004 et 2008 respectivement, répondent à ce besoin. Le développement de la cinquième génération (5G) de réseaux radio mobiles est en cours, afin de remplacer l'actuelle quatrième génération (4G) LTE et (4.5G) LTE-A. Néanmoins, la transition de la 4G à la 5G pourrait prendre une dizaine d'années, voire plus. Bien que plusieurs technologies semblent constituer des éléments clefs de la prochaine génération, la norme LTE-A devrait continuer à jouer un rôle important dans l'avènement de la 5G [1]. Le principal résultat des études sur les normes LTE et LTE-A est représenté par les standards sans fil, sous les contraintes de haut débit de données et de qualité de service élevée, permettant d'assurer des taux d'erreur faibles et une latence réduite [2–5]. Par ailleurs, les systèmes de communications à faible complexité sont également essentiels pour les réseaux mobiles 5G [6, 7]. Cette thèse s'inscrit dans les tendances technologiques actuelles, en proposant des schémas de communications sans fil à complexité et latence réduites, qui utilisent des codes FEC (Forward Error Correction) simples. Nous optimisons ensuite les performances de ces systèmes en termes de taux d'erreur afin qu'ils puissent être utilisés avec les normes LTE et LTE-A.

L'un des éléments essentiels permettant de satisfaire les spécifications des normes LTE et LTE-A est la technologie MIMO (Multiple-Input Multiple-Output). En effet, les systèmes sans fil à antennes multiples, communément appelés systèmes MIMO, sont devenus de plus en plus populaires depuis la fin des années 90, après la prédiction théorique et la démonstration pratique de leurs capacités, notamment en termes de haute efficacité spectrale [8, 9]. Les avantages de l'utilisation de plusieurs antennes en émission et en réception par un système MIMO sans fil ont été intensivement exploités ces dernières années [10]. En utilisant des émetteurs-récepteurs multi-antennes, la technologie MIMO offre non seulement la possibilité du multiplexage et le gain de diversité, mais permet également d'obtenir une fiabilité de liaison point-à-point plus importante par rapport aux systèmes mono-antenne [11]. Le principal défi consiste à concevoir des systèmes

MIMO qui exploitent pleinement la présence de plusieurs antennes. Ainsi, la répétition des symboles sur plusieurs antennes a été proposée en [12, 13] pour obtenir la diversité et la robustesse de la transmission. L'association de cette technique avec les récepteurs itératifs conduit à des performances prometteuses [14, 15]. Plus important encore, dans les schémas duplex en boucle fermée dans le domaine temporel, la réponse du canal (Channel State Information - CSI) est disponible à l'émetteur via un lien de retour, ce qui permet de mettre en place un précodeur adapté aux caractéristiques du canal. Dès lors, plusieurs précodeurs linéaires ont été proposés dans la littérature. Ils ont été conçus en fonction de différents critères, tels que la maximisation de la distance Euclidienne minimale dans la constellation reçue (max-$d_{\min}$ [16]), l'optimisation de l'allocation de la puissance (minimisation du taux d'erreur binaire (BER) [17]), erreur quadratique moyenne minimale (MMSE) pondérée [18], max-SNR (maximisation du rapport signal à bruit (SNR) reçu ou formation de voie) [19], water filling (WF) [9] ou le mercury/water-filling (m/WF) [20].

Malheureusement, ni le code FEC externe, ni la structure du récepteur n'ont été pris en compte dans la conception de la plupart des précodeurs linéaires, à l'exception des codes espace-temps en bloc, qui tiennent compte de la CSI seulement côté récepteur [21]. Bien que les codes LDPC (Low Density Parity Check) aient été considérés en [22], le précodeur associé n'est pas adapté à un récepteur spécifique. Son critère de conception est la maximisation globale de l'information mutuelle entre l'alphabet fini d'entrée et la sortie du canal. Puisque ce précodeur est globalement optimisé, il est appelé précodeur GOPT. L'inconvénient majeur du précodeur GOPT est représenté par sa complexité, puisqu'il implique la recherche de la solution optimale pour chaque réalisation du canal et valeur du SNR. Bien que la mise en œuvre du précodeur GOPT soit de ce fait irréalisable en pratique, ses performances peuvent être utilisées comme limite inférieure pour la conception d'autres précodeurs.

Dans cette thèse, nous considérons la concaténation d'un codeur FEC et d'un précodeur MIMO linéaire. Nos investigations sont orientées dans deux directions. D'une part, nous optimisons le précodeur MIMO linéaire en considérant un codeur FEC binaire simple et des récepteurs itératifs. Tout d'abord, nous étudions une structure de réception qui utilise un turbo-détecteur. Nous proposons, au **chapitre 3**, un précodeur(noté $\mathbf{F}_{\ell_1}$), qui améliore considérablement le taux d'erreur du système en optimisant la distance Euclidienne minimale entre chaque couple de symboles, dont les configurations binaires associées diffèrent par un seul bit. En outre, en considérant un mapping direct de la constellation reçue, nous introduisons également un nouveau précodeur (appelé EXIT-based), associé au mapping optimal et optimisé en exploitant des diagrammes de transfert d'information extrinsèque (EXIT) afin d'améliorer le taux d'erreur du système. Nous comparons les précodeurs proposés au précodeur max-$d_{\min}$, dont les performances en termes de taux

d'erreur sont supérieures par rapport aux autres précodeurs existants pour des systèmes non-codés et une détection au sens du maximum de vraisemblance. Néanmoins, pour des systèmes utilisant des codeurs FEC, un bon critère d'optimisation pour le précodeur MIMO est la maximisation de l'information mutuelle du canal. Malheureusement, le meilleur précodeur maximisant l'information mutuelle du canal, qui est le précodeur GOPT mentionné précédemment, ne peut pas être utilisé sur des canaux aléatoires en raison de sa complexité élevée. Par conséquent, dans le **chapitre 4**, nous introduisons un algorithme combinant les critères de conception des précodeurs max-$d_{\min}$ et GOPT, et proposons un nouveau précodeur optimisé à faible complexité (appelé LCOPT), qui maximise asymptotiquement l'information mutuelle du canal, avec une complexité acceptable. Le précodeur LCOPT proposé fonctionne avec des constellations reçues fixes, et peut donc facilement utiliser les mappings proposés dans le chapitre 3. Deuxièmement, nous concentrons notre étude sur l'optimisation du précodeur en considérant une structure de réception à base de turbo-égalisation, dont la complexité est inférieure par rapport à la turbo-détection. Toutefois, dans le cas de la turbo-égalisation, les symboles reçus sont tout d'abord décomposés en flux de données parallèles pour annuler les interférences, avant d'être convertis en information souple en entrée du décodeur. Par conséquent, le mapping de la constellation reçue n'est pas essentiel pour ce récepteur. En outre, puisque le décodeur ne reçoit pas en entrée directement la sortie du canal, mais l'information souple obtenue après l'annulation des interférences, il est important de maximiser l'information mutuelle entre les symboles transmis et les symboles après l'annulation des interférences plutôt que l'information mutuelle du canal. Par conséquent, nous proposons, dans le **chapitre 5**, un nouveau précodeur qui maximise l'information mutuelle après l'annulation des interférences (appelé précodeur Génie optimisé). Les simulations réalisées en présence d'un turbo-égaliseur en réception montrent en effet un gain de performance significatif du précodeur Génie optimisé par rapport au précodeur GOPT, qui vise à maximiser l'information mutuelle du canal. D'autre part, nous considérons l'utilisation des codes FEC LDPC non-binaires (NB-LDPC). Nous étudions tout d'abord la concaténation des codes NB-LDPC avec des précodeurs MIMO linéaires. Habituellement, les codeurs NB-LDPC utilisent des corps de Galois (GF) d'ordre élevé pour augmenter le débit de données. Chaque symbole GF est généralement mappé sur un vecteur symbole MIMO. Bien que plusieurs algorithmes aient été proposés dans la littérature pour réduire la complexité du décodage, celle-ci reste toutefois significative. Nous proposons alors, dans le **chapitre 6**, de mapper plusieurs symboles GF d'ordre réduit en un vecteur symbole MIMO, en considérant un turbo-détecteur non-binaire en réception. Nous démontrons que ce mapping réduit significativement la complexité calculatoire au niveau du récepteur. En plus, nous proposons d'ajouter un précodeur MIMO linéaire afin d'améliorer le taux d'erreur du système et de diminuer encore davantage la complexité en limitant le nombre d'itérations du décodeur NB-LDPC. Enfin, pour achever notre étude

sur la réduction de la complexité des systèmes de communications qui utilisent des codes NB-LDPC, nous proposons au **chapitre 7** un nouvel algorithme de décodage à faible complexité pour ces codes, basé sur la fiabilité des décisions dures en sortie.

Le projet de recherche associé à cette thèse nous a offert l'opportunité de deux collaborations internationales initiées via des séjours scientifiques (deux mois à Memorial University of Newfoundland, en collaboration avec les professeurs Octavia Dobre et Telex M. N. Ngatched, et deux mois à Missouri University of Science and Technology, en collaboration avec les professeurs Yahong Rosa Zheng et Chengshan Xiao).

Modèle de la chaîne de transmission

Notre étude concerne les systèmes MIMO sans fil en bande de base. Considérons un système MIMO avec n_R et n_T antennes de réception et d'émission respectivement, et b flux de données indépendants à transmettre. Nous supposons une connaissance parfaite du canal à la fois à l'émetteur et au récepteur. Côté émetteur, nous étudions la concaténation d'un codeur FEC et d'un précodeur MIMO linéaire. Le modèle de la chaîne de transmission est illustré sur la Figure 1.

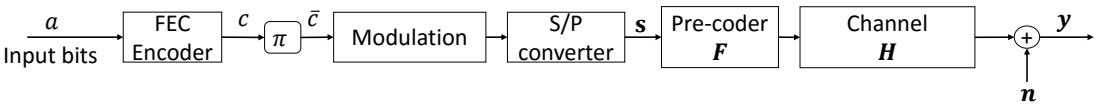


FIGURE 1: Modèle de la chaîne de transmission.

Un codeur FEC à faible complexité est utilisé pour coder les bits d'information. Les mots de code FEC sont ensuite entrelacés avant d'être mappés sur les symboles d'une modulation QAM (Quadrature Amplitude Modulation). Les symboles modulés sont convertis en un vecteur symbole b-dimensionnel $\mathbf{s}$. Le vecteur $\mathbf{s}$ est ensuite précodé par une matrice $\mathbf{F}$ et transmis dans le canal MIMO. La sortie du détecteur $\mathbf{y}$ est alors exprimée par :

$$\mathbf{y} = \mathbf{GHFs} + \mathbf{Gn}, \tag{1}$$

où $\mathbf{F}$ est la matrice de précodage, de taille $n_T \times b$, avec la contrainte de puissance $\|\mathbf{F}\|_F^2 = 1$, $\mathbf{G}$ est la matrice de détection, de taille $b \times n_R$, $\mathbf{H}$ est la matrice du canal, de taille $n_R \times n_T$, et $\mathbf{n}$ est le vecteur $n_R \times 1$ du bruit additif blanc Gaussien complexe, à symétrie circulaire. Nous supposons que $\mathrm{E}[\mathbf{nn}^\dagger] = \sigma_n^2 \mathbf{I}_{n_R}$ et $\mathrm{E}[\mathbf{ss}^\dagger] = \sigma_s^2 \mathbf{I}_b$, où $\mathrm{E}[.]$ et $(.)^\dagger$ représentent respectivement l'espérance mathématique et la transposée conjuguée, et $\mathbf{I}_{n_R}$ est la matrice identité de taille n_R.

Sous l'hypothèse de la connaissance parfaite du canal à l'émetteur, nous considérons la transformation suivante pour simplifier le modèle du système. Considérons les matrices $\mathbf{F}_d$ et $\mathbf{F}_v$ telles que $\mathbf{F} = \mathbf{F}_v \mathbf{F}_d$. La décomposition en valeurs singulières de la matrice du canal $\mathbf{H}$ de taille $n_R \times n_T$ conduit à $\mathbf{H} = \mathbf{U_H \Sigma_H V_H^\dagger}$. Avec les notations $\mathbf{F}_v =$

$\mathbf{V}_{\mathbf{H}}^{\dagger}\begin{pmatrix}\mathbf{I}_b & \mathbf{0}\end{pmatrix}^T$ et $\mathbf{G} = \begin{pmatrix}\mathbf{I}_b & \mathbf{0}\end{pmatrix}\mathbf{U}_{\mathbf{H}}^{\dagger}$, nous obtenons :

$$\mathbf{GHF} = \begin{pmatrix}\mathbf{I}_b & \mathbf{0}\end{pmatrix}\mathbf{U}^{\dagger}\mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\dagger}\mathbf{V}\begin{pmatrix}\mathbf{I}_b & \mathbf{0}\end{pmatrix}^T\mathbf{F}_d = \mathbf{H}_v\mathbf{F}_d, \tag{2}$$

où $\mathbf{H}_v = [\mathbf{\Sigma}_{\mathbf{H}}]_b = \text{diag}(\sigma_1, ..., \sigma_b)$ est la matrice propre du canal, de taille $b \times b$, $\{\sigma_1, ..., \sigma_b\}$ sont les b valeurs singulières les plus significatives de $\mathbf{H}$ triées par ordre décroissant, et $\mathbf{F}_d$ représente la nouvelle matrice de précodage, qui satisfait également la contrainte de puissance $\|\mathbf{F}_d\|_F^2 = 1$. L'équation (1) peut être alors réécrite sous la forme :

$$\mathbf{y} = \mathbf{H}_v\mathbf{F}_d\mathbf{s} + \boldsymbol{\eta}, \tag{3}$$

où $\boldsymbol{\eta}$ est le vecteur bruit de taille $b \times 1$, avec $\text{E}[\boldsymbol{\eta}\boldsymbol{\eta}^{\dagger}] = \sigma_{\eta}^2\mathbf{I}_b$ et $\sigma_{\eta}^2 = \sigma_n^2$.

Nous considérons le plus souvent dans cette thèse le cas $b = 2$, qui est largement utilisé dans la 4$^{\text{ème}}$ génération (4G) de réseaux cellulaires (les normes LTE et LTE-A). Il est à noter que la transformation ci-dessus requiert l'inégalité $b \leq \text{rank}(\mathbf{H}) \leq \min(n_T, n_R)$, de sorte que n_T et n_R soient supérieurs à b. Par conséquent, les résultats des analyses effectuées en considérant $b = 2$, ne sont pas forcément limités aux systèmes MIMO 2×2.

Soit la décomposition en valeurs singulières de $\mathbf{F}_d$: $\mathbf{F}_d = \mathbf{U}_{F_d}\mathbf{\Sigma}_{F_d}\mathbf{V}_{F_d}^{\dagger}$. Dans le cas $b = 2$, $\mathbf{\Sigma}_{F_d} = \begin{pmatrix}\cos\psi & 0 \\ 0 & \sin\psi\end{pmatrix}$. Cette forme polaire satisfait également la contrainte de puissance $\|\mathbf{F}_d\|_F^2 = 1$. Pour une valeur de ψ, il peut être démontré que le meilleur choix de $\mathbf{U}_{F_d}$, qui maximise les valeurs singulières de $\mathbf{H}_v\mathbf{F}_d$ est $\mathbf{U}_{F_d} = \mathbf{I}_b$ [16]. Il est également connu que la matrice unitaire 2×2 restante $\mathbf{V}_{F_d}^{\dagger}$ peut être décomposée en $\mathbf{V}_{F_d}^{\dagger} = \mathbf{D}\begin{pmatrix}\cos\theta & \sin\theta e^{i\phi} \\ -\sin\theta & \cos\theta e^{i\phi}\end{pmatrix}$ [22], où $\mathbf{D}$ est la matrice diagonale unitaire. Sans aucune perte de généralité, considérons $\mathbf{D} = \mathbf{I}_2$. Par conséquent, $\mathbf{F}_d$ admet la forme paramétrée ci-dessous :

$$\mathbf{F}_d = \begin{pmatrix}\cos\psi & 0 \\ 0 & \sin\psi\end{pmatrix}\begin{pmatrix}\cos\theta & \sin\theta \\ -\sin\theta & \cos\theta\end{pmatrix}\begin{pmatrix}1 & 0 \\ 0 & e^{i\phi}\end{pmatrix}, \tag{4}$$

où ψ ($0^o \leq \psi \leq 90^o$) est lié à l'allocation de puissance sur les canaux propres, tandis que θ ($0^o \leq \theta \leq 90^o$) et ϕ ($0^o \leq \phi \leq 90^o$) permettent de mélanger et faire pivoter respectivement des symboles sur les deux canaux propres.

En outre, la conversion des coordonnées cartésiennes en coordonnées polaires de $\mathbf{H}_v$ donne :

$$\mathbf{H}_v = \begin{pmatrix}\sigma_1 & 0 \\ 0 & \sigma_2\end{pmatrix} = \rho\begin{pmatrix}\cos\gamma & 0 \\ 0 & \sin\gamma\end{pmatrix}, \tag{5}$$

où ρ et γ représentent le gain et l'angle du canal respectivement. Puisque $\sigma_1 \geq \sigma_2 > 0$, on en déduit que $0^o < \gamma \leq 45^o$. Nous définissons alors le SNR instantané en réception par :

$$\text{SNR} = \frac{\sigma_s^2}{\sigma_\eta^2}\|\mathbf{H}\|_F^2 = \frac{\sigma_s^2}{\sigma_\eta^2}\rho^2. \tag{6}$$

Avec cette définition du SNR, le canal est caractérisé seulement par l'angle γ. Il en résulte une façon intéressante de représenter n'importe quel canal dans le cas $b = 2$ (notons que n_T et n_R peuvent être supérieurs à b).

Les contributions originales de cette thèse sont présentées dans les prochains chapitres. La numérotation de ces chapitres est la même que dans le manuscrit complet, qui se trouve après ce résumé étendu.

Chapitre 3

Optimisation conjointe du précodeur MIMO et du mapping symbole pour turbo-détection

3.1 Principales contributions

Dans ce chapitre, notre étude concerne l'optimisation du précodeur $\mathbf{F}_d$, en considérant un turbo-détecteur en réception, et a comme résultat quatre contributions principales. Tout d'abord, nous proposons une nouvelle étude de cas, qui tient compte de la concaténation du précodeur MIMO avec un code FEC externe, en supposant un récepteur itératif.

Deuxièmement, nous proposons un nouveau précodeur, qui est particulièrement adapté à la turbo-détection dans le cas du schéma de mapping usuel, représenté par la conversion du train binaire en symboles Q-aires, suivie par la conversion des symboles Q-aires en symboles MIMO (mapping Gray-M). Les paramètres de ce précodeur sont fixes pour tous les canaux, ce qui réduit la complexité de la conception.

Comme 3ème contribution, nous introduisons un mapping des symboles MIMO, qui peut être vu comme une correspondance directe avec la constellation reçue, et qui remplace l'habituelle concaténation du modulateur et du convertisseur des symboles Q-aires en symboles MIMO. Nous démontrons que la robustesse du précodeur linéaire est améliorée en appliquant la stratégie de mapping MSEW (Minimum Squared Euclidean Weight).

Enfin, nous proposons un nouveau précodeur basé sur l'analyse des diagrammes EXIT, et adapté au mapping MSEW. Le précodeur max-$d_{\min}$, qui présente le meilleur taux d'erreur pour des systèmes non-codés et une détection au sens du maximum de vraisemblance, est considéré ici pour comparaison. Les résultats numériques montrent que le nouveau

précodeur surpasse largement le précodeur max-$d_{\min}$ dans cette configuration (codeur FEC externe, turbo-détection, mapping MSEW).

3.2 Modèle du système

Un code RSC (Recursive Systematic Convolutional) binaire est utilisé comme codeur FEC externe pour coder les bits d'information. Les mots de code FEC sont ensuite entrelacés avant d'être appliqués au modulateur. Dans le modulateur, la séquence binaire encodée et entrelacée $\bar{\mathbf{c}}$ est mappée sur une séquence s de symboles Q-aires QAM, qui est ensuite convertie en b flux parallèles, c'est-à-dire que les b symboles de s sont regroupés et transposés en un symbole MIMO $\mathbf{s}$ de taille $b \times 1$. Le vecteur $\mathbf{s}$ est ensuite précodé et transmis dans le canal MIMO. A la réception, après la détection MIMO, un démappeur MIMO souple échange l'information extrinsèque de manière itérative avec un décodeur souple BCJR [23]. Après l'application de la transformation de canal, nous obtenons le schéma équivalent représenté sur la figure 3.1, où $L^1_{\mathrm{A}}, L^1_{\mathrm{P}}$ et L^1_{E} sont respectivement les rapports de log-vraisemblance (LLR) *a priori*, *a posteriori*, et extrinsèque du démappeur souple, tandis que les notations équivalentes pour le décodeur BCJR souple sont $L^2_{\mathrm{A}}, L^2_{\mathrm{P}}$ et L^2_{E}.

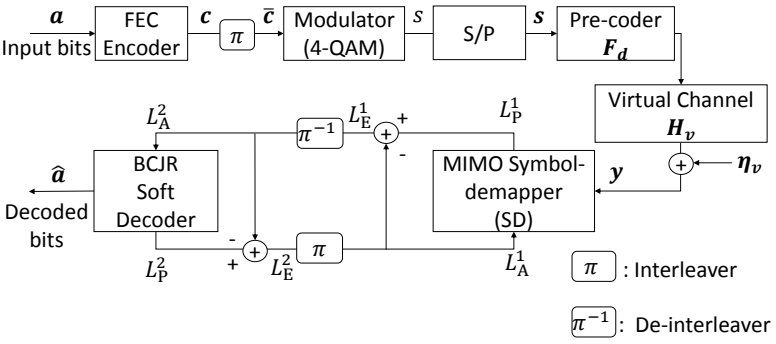


FIGURE 3.1: Système précodé à turbo-détecteur.

3.3 Précodeur optimisé pour le mapping standard Gray-M

3.3.1 Analyse

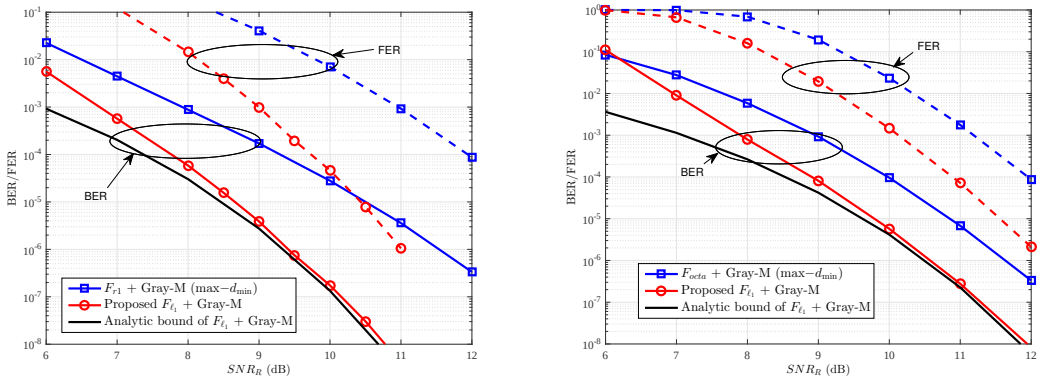
Soit ℓ le carré de la distance Euclidienne entre les symboles de la constellation reçue, dont les séquences binaires mappées diffèrent par un seul bit (distance de Hamming égale à 1). Par conséquent, ℓ s'écrit :

$$\ell = \|\mathbf{H}_v \mathbf{F}_d(\breve{\mathbf{s}} - \breve{\mathbf{s}}')\|^2, \tag{3.1}$$

où $\check{\mathbf{s}}$ et $\check{\mathbf{s}}'$ désignent tout couple de vecteurs symbole, dont les représentations binaires associées ne diffèrent que par un seul bit. La valeur minimale de ℓ sur l'ensemble de constellations reçues est notée ℓ_1. Nous démontrons tout d'abord que la maximisation de ℓ_1 permet d'améliorer le taux d'erreur du système précodé, avec turbo-détection. Deuxièmement, nous proposons un nouveau précodeur, appelé $\mathbf{F}_{\ell_1}$, qui vise à maximiser ℓ_1 dans le cas du mapping Gray-M standard.

3.3.2 Sélection de résultats obtenus

Dans le cas où deux flux de données sont transmis ($b = 2$), en utilisant la modulation 4-QAM, le précodeur max-$d_{\min}$ [16] montre de bonnes performances en termes de taux d'erreur pour un système non-codé, avec détection au sens du maximum de vraisemblance. Nous comparons les précodeurs $\mathbf{F}_{\ell_1}$ et max-$d_{\min}$ en termes de taux d'erreur. Le précodeur max-$d_{\min}$ vise à maximiser la distance Euclidienne minimale entre les points de la constellation reçue. Ce précodeur a deux formes différentes en fonction de l'angle de canal γ. Il distribue la puissance sur les deux sous-canaux (précodeur $\mathbf{F}_{octa}$) si $\gamma > \gamma_0 \simeq 17.28^o$, et utilise uniquement le premier sous-canal avec le gain du canal maximum (précodeur $\mathbf{F}_{r1}$) si $\gamma \leq \gamma_0$. Afin d'assurer une comparaison objective, les simulations ont été effectuées pour deux canaux, notés A et B, qui sont caractérisés respectivement par $\gamma = 8.3^o < \gamma_0$ and $\gamma \simeq 30^o > \gamma_0$. Les taux d'erreur des précodeurs $\mathbf{F}_{\ell_1}$ et max-$d_{\min}$ sur ces deux canaux sont montrés respectivement sur les figures 3.2(a) et 3.2(b). Nous constatons que



(a) Canal A. (b) Canal B.

FIGURE 3.2: BER (courbes en trait plein) et FER (courbes en trait discontinu), pour le mapping Gray-M et un code (13,15)-RSC.

le précodeur proposé $\mathbf{F}_{\ell_1}$ a de meilleures performances que le précodeur max-$d_{\min}$ pour les deux canaux. Il est aussi à noter que la variation des pentes des courbes pour le canal B est plus lente que pour le canal A. Il en résulte que l'utilisation du précodeur $\mathbf{F}_{\ell_1}$ est encore plus avantageuse pour les canaux caractérisés par un faible angle γ.

3.4 Amélioration des performances en utilisant le mapping des symboles MIMO

3.4.1 Analyse

Bien que les performances aient été améliorées par l'utilisation du précodeur $\mathbf{F}_{\ell_1}$, ℓ_1 peut être maximisée encore davantage si nous pouvons contrôler le mapping de la constellation reçue. Puisque celle-ci est fixe, il est possible de réaliser un mapping direct là-dessus. Par conséquent, nous proposons de réaliser un mapping direct des symboles MIMO au niveau du mappeur intérieur, qui transforme un bloc de m bits en un vecteur symbole $\mathbf{s}$. Cette opération équivaut à un mapping direct sur le vecteur symbole reçu $\mathbf{x} = \mathbf{H}_v\mathbf{F}_d\mathbf{s}$ puisque les matrices $\mathbf{H}_v$ et $\mathbf{F}_d$ sont connues. Au niveau de la constellation reçue nous utilisons la stratégie du mapping MSEW. Introduit en [24], l'objectif du mapping MSEW est d'obtenir un faible taux d'erreur asymptotique (région du plancher d'erreur) en optimisant deux critères. Tout d'abord, celui-ci maximise la distance Euclidienne quadratique minimale entre les symboles dont les séquences binaires associées ne différent que par un seul bit (c'est-à-dire ℓ_1). Deuxièmement, il minimise le nombre de couples de symboles dont les séquences binaires associées différent par un seul bit, et qui sont séparés par la distance Euclidienne quadratique minimale ℓ_1. A partir de ces deux critères, le meilleur mapping MSEW est ensuite obtenu de manière numérique. Puisque les formes des constellations reçues sont différentes en fonction du précodeur utilisé, l'optimisation MSEW doit être effectuée pour chaque précodeur $\mathbf{F}_d$.

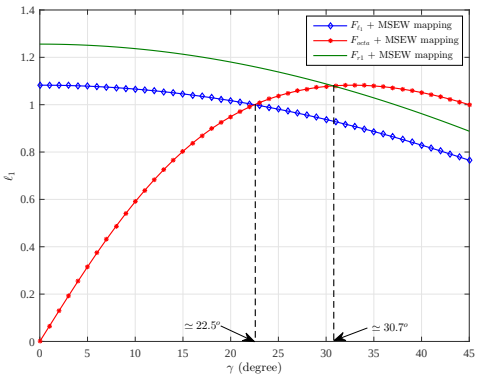
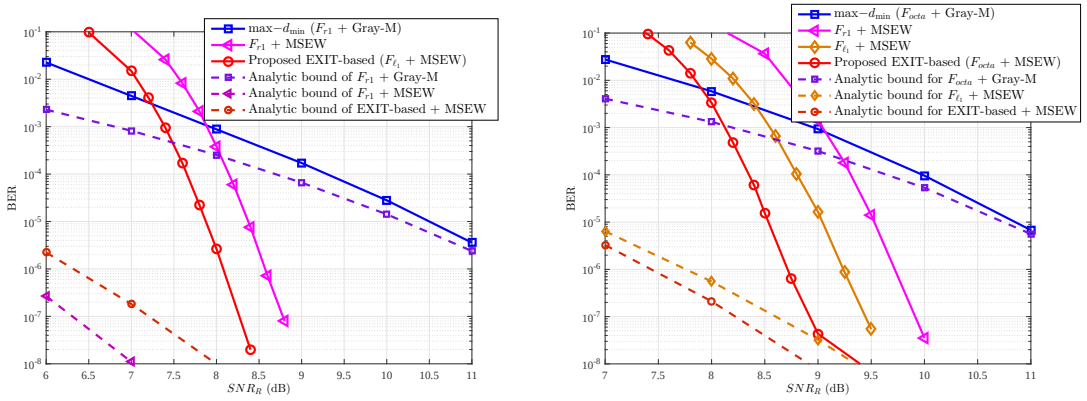


FIGURE 3.3: Variation de ℓ_1 en fonction de γ pour différents précodeurs en utilisant le mapping MSEW.

Nous comparons les précodeurs associés au mapping MSEW par le biais du paramètre ℓ_1. La figure 3.3 montre la variation de ℓ_1 en fonction de γ (c'est-à-dire tous les canaux possibles). Tel qu'il peut être constaté, en considérant comme critère la maximisation de ℓ_1, la meilleure solution consiste à associer le mapping MSEW à un précodeur qui commute de $\mathbf{F}_{r1}$ à $\mathbf{F}_{octa}$ au-delà de $\gamma_0' = 30.7^o$, qui devient la nouvelle valeur de seuil.

Toutefois, il est à noter que le critère de la maximisation de ℓ_1 montre ses avantages seulement dans la région du plancher d'erreur, qui est très faible dans le cas du mapping MSEW et hors du domaine des SNR couramment utilisés. En revanche, nous sommes surtout intéressés par la région de *turbo-cliff*, pour laquelle nous proposons une solution de précodage, basée sur l'analyse des diagrammes EXIT, à utiliser conjointement avec le mapping MSEW de la constellation reçue. Ce précodeur a deux modes de travail déterminés par le seuil $\gamma_{\ell_1} = 22.5^o$ (voir la figure 3.3). Pour les canaux ayant $\gamma \leq \gamma_{\ell_1}$, le précodeur à base de diagrammes EXIT utilise $\mathbf{F}_{\ell_1}$, alors que pour les canaux ayant $\gamma > \gamma_{\ell_1}$, celui-ci utilise $\mathbf{F}_{octa}$.

3.4.2 Sélection de résultats obtenus



(a) Canal A. (b) Canal B.

FIGURE 3.4: Performances en termes de BER des précodeurs associés aux mappings correspondants : canal A et canal B.

La figure 3.4(a) montre le taux d'erreur des précodeurs utilisés avec les mappings correspondant au canal A. Tel qu'il peut être constaté sur cette figure, le précodeur $\mathbf{F}_{\ell_1}$ permet d'obtenir un gain de 2.6 dB, pour un BER $= 10^{-5}$, comparé au précodeur max-$d_{\min}$ (avec le mode $\mathbf{F}_{r1}$ pour le canal A et mapping Gray-M). Le $\mathbf{F}_{r1}$utilisé conjointement avec le mapping MSEW conduit également à un gain important par rapport au précodeur max-$d_{\min}$, qui est d'environ 2.1 dB pour un BER $= 10^{-5}$. D'autre part, nous constatons que la limite théorique de $\mathbf{F}_{r1}$ utilisé conjointement avec le mapping MSEW, est inférieure à celle du précodeur $\mathbf{F}_{\ell_1}$. Néanmoins, la région du BER permettant d'atteindre ces limites est très faible et hors du domaine des SNR couramment utilisés. En revanche, les résultats au niveau de la région de turbo-cliff sont plus intéressants. Ainsi, nous constatons que dans cette région le précodeur $\mathbf{F}_{\ell_1}$ permet d'obtenir un gain de 0.5 dB pour un BER $= 10^{-6}$, comparé au précodeur $\mathbf{F}_{r1}$ utilisé conjointement avec le mapping MSEW. Des conclusions similaires peuvent être aussi tirées pour le canal B, tel qu'il est illustré sur la figure 3.4(b).

3.4.3 Publications en lien avec ce chapitre

[J1] N.-Q. Nhan, Philippe Rostaing, Karine Amis, Ludovic Collin, and Emanuel Radoi, *Joint optimization of MIMO precoding and symbol mapping for turbo detection*, article revue en cours de soumission.

[C1] N.-Q. Nhan, P. Rostaing, K. Amis, L. Collin, and E. Radoi, *Optimized MIMO symbol mapping to improve the turbo cliff region of iterative precoded MIMO detection*, actes de la 23$^{\text{ème}}$ conférence européenne en traitement du signal (EUSIPCO), Septembre 2015, pp. 909-913.

[C2] N.-Q. Nhan, P. Rostaing, K. Amis, L. Collin, and E. Radoi, *Optimized maxdmin precoder assuming maximum squared Euclidean weight-mapping and turbo detection*, actes du 9$^{\text{ème}}$ symposium international « Turbo Codes & Iterative Information Processing » (ISTC), Septembre 2016.

Chapitre 4

Précodeur optimisé à faible complexité pour les canaux MIMO aléatoires

4.1 Principales contributions

Dans ce chapitre, notre objectif est de diminuer la complexité de l'algorithme de précodage GOPT. Nos principales contributions concernent deux aspects.

D'une part, en utilisant une limite inférieure de l'information mutuelle du canal, il peut être démontré que maximiser la distance Euclidienne minimale $d_{\min}$ équivaut asymptotiquement à maximiser l'information mutuelle du canal [25].Par ailleurs, nous pouvons mettre la matrice de précodage sous la forme $\mathbf{F}_d = \mathbf{\Psi\Theta}$, où la matrice diagonale $\mathbf{\Psi}$ est liée à l'allocation de puissance sur les différents sous-canaux et la matrice unitaire $\mathbf{\Theta}$ dépend des formes des constellations reçues. Par conséquent, comme première contribution, nous proposons un nouvel algorithme de précodage à faible complexité optimisé, appelé LCOPT, qui réduit de manière significative la charge de calcul, comparé à GOPT. L'algorithme proposé utilise la matrice unitaire fixe $\mathbf{\Theta}$ déterminée à partir de la solution de la maximisation de $d_{\min}$, et optimise uniquement la matrice d'allocation de puissance $\mathbf{\Psi}$ selon le critère de l'information mutuelle.

Le choix de la matrice $\mathbf{\Theta}$ est discuté pour $b = 2$ et $b > 2$, sachant que ce choix détermine aussi la forme de la constellation reçue du précodeur LCOPT. Notre deuxième contribution est alors de réaliser le mapping MSEW de la constellation reçue du précodeur LCOPT proposé. Nous rappelons que le mapping MSEW a été proposé dans [24] et a déjà été appliqué dans nos travaux de recherche précédents (voir le chapitre 3 et [26]). Il

est à noter qu'il n'est pas pratique de réaliser le mapping MSEW de la constellation reçue du précodeur GOPT puisque dans ce cas la constellation change pour chaque réalisation du canal et valeur du SNR. Dans ce chapitre, nous utilisons le même modèle du système représenté sur la figure 3.1.

4.2 Analyse

L'information mutuelle du canal entre l'entrée discrète $\mathbf{s}$ et la sortie du canal $\mathbf{y}$ est donnée par [22] :

$$\mathcal{I}(\mathbf{y},\mathbf{s}) = b\log_2 Q - \frac{1}{Q^b}\sum_{m=1}^{Q^b} \mathrm{E}\left\{\log_2 \sum_{k=1}^{Q^b} e^{-\zeta_{m,k}}\right\}, \tag{4.1}$$

où $\zeta_{m,k} = \left(\|\mathbf{H}_v\mathbf{F}_d(\mathbf{s}_m - \mathbf{s}_k) + \boldsymbol{\eta}\|^2 - \|\boldsymbol{\eta}\|^2\right)/\sigma_\eta^2$ et Q est l'ordre de la modulation Q-aire. Soit $\mathbf{W} = \mathbf{F}_d^\dagger\mathbf{H}_v^\dagger\mathbf{H}_v\mathbf{F}_d$. On montre que l'information mutuelle $\mathcal{I}(\mathbf{y},\mathbf{s})$ est une fonction concave par rapport à la matrice $\mathbf{W}$. La décomposition en valeurs singulières de $\mathbf{F}_d$ conduit à $\mathbf{F}_d = \mathbf{U}_\mathbf{F}\boldsymbol{\Sigma}_\mathbf{F}\mathbf{V}_\mathbf{F}^\dagger$. Il peut être démontré [27, Proposition 1] que la matrice $\mathbf{U}_\mathbf{F}$ peut toujours être choisie égale à la matrice identité,$\mathbf{I}_b$. Par conséquent, en réécrivant $\boldsymbol{\Psi} = \boldsymbol{\Sigma}_\mathbf{F}$, $\boldsymbol{\Theta} = \mathbf{V}_\mathbf{F}^\dagger$, il en résulte :

$$\mathbf{F}_d = \boldsymbol{\Psi}\boldsymbol{\Theta}. \tag{4.2}$$

La matrice $\boldsymbol{\Psi}$ contrôle l'allocation de puissance sur chaque sous-canal, tandis que la matrice $\boldsymbol{\Theta}$ est liée à la rotation et à la mise à l'échelle des symboles de la constellation reçue. On montre que l'information mutuelle $\mathcal{I}(\mathbf{y},\mathbf{s})$ est aussi une fonction concave par rapport $\boldsymbol{\Psi}^2$.

Avec $\mathbf{F}_d = \boldsymbol{\Psi}\boldsymbol{\Theta}$, dans [22] les auteurs ont proposé un algorithme itératif (appelé GOPT), qui met à jour les matrices $\boldsymbol{\Theta}$ et $\boldsymbol{\Psi}$ en utilisant la méthode du gradient ascendant (au lieu de mettre à jour directement $\mathbf{F}_d$, qui ne semble pas faisable). Alors que la matrice $\boldsymbol{\Psi}$ peut être mise à jour directement par le biais de son gradient, dans le cas de la matrice $\boldsymbol{\Theta}$ nous devons utiliser la variation de la matrice $\mathbf{W}$, notée $\delta\mathbf{W}$. Bien que le précodeur GOPT maximise globalement la capacité du canal, la complexité de cet algorithme est trop élevée. En effet, la mise à jour de la matrice $\boldsymbol{\Theta}$ exige un temps de calcul important, notamment lorsque le nombre de flux de données b devient élevé. En outre, la convergence de l'algorithme GOPT est sensible au choix de la valeur initiale de la matrice $\boldsymbol{\Theta}$. Dans ce chapitre, nous proposons une solution sous-optimale pour contourner cet inconvénient.

Nous utilisons l'expression suivante pour la limite inférieure de l'information mutuelle du canal :

$$\begin{aligned}\mathcal{I}_{\mathrm{LB}}(\mathbf{y},\mathbf{s}) &= b\log_2 Q - (1/\ln 2 - 1)b \\ &\quad - \frac{1}{Q^b}\sum_{m=1}^{Q^b}\log_2\sum_{k=1}^{Q^b}\exp\left(-\frac{\|\mathbf{H}_v\mathbf{F}_d(\mathbf{s}_m-\mathbf{s}_k)\|^2}{2\sigma_\eta^2}\right).\end{aligned} \tag{4.3}$$

D'une part, nous pouvons en déduire que maximiser $\mathcal{I}_{\mathrm{LB}}(\mathbf{y},\mathbf{s})$ permet de maximiser l'information mutuelle du canal $\mathcal{I}(\mathbf{y},\mathbf{s})$. L'optimalité asymptotique dans la région des SNR élevés montre que maximiser $\mathcal{I}_{\mathrm{LB}}(\mathbf{y},\mathbf{s})$ à haut SNR équivaut à maximiser la distance Euclidienne minimale $d_{\min} = \min_{m\neq k}\left\{\|\mathbf{x}_m-\mathbf{x}_k\|^2\right\}$ entre les symboles de la constellation reçue, qui sont définis par $\mathbf{x}_i = \mathbf{H}_v\mathbf{F}_d\mathbf{s}_i$. D'autre part, rappelons que dans l'équation (4.2) la matrice $\mathbf{\Psi}$ contrôle l'allocation de puissance sur chaque sous-canal, tandis que la matrice $\mathbf{\Theta}$ est liée à la rotation et à la mise à l'échelle des symboles de la constellation reçue. Par conséquent, pour une matrice $\mathbf{\Theta}$ donnée, la constellation reçue est fixée. En considérant une matrice $\mathbf{\Theta}$ obtenue par la maximisation de $d_{\min}$, nous proposons alors un nouvel algorithme sous-optimal à faible complexité, qui met à jour seulement la matrice d'allocation de puissance $\mathbf{\Psi}$. Nous appelons les matrices de précodage obtenues grâce au nouvel algorithme des précodeurs à faible complexité optimisés (ou précodeurs LCOPT).

Rappelons que la matrice du canal $\mathbf{H}_v$ et la matrice d'allocation de puissance $\mathbf{\Psi}$ effectuent seulement une mise à l'échelle des constellations entre les sous-canaux, alors que la matrice $\mathbf{\Theta}$ modifie complètement la forme de la constellation reçue. Afin d'appliquer le mapping MSEW, une recherche doit donc être effectuée pour chaque matrice $\mathbf{\Theta}$. Malheureusement, dans le cas du précodeur GOPT [22], celle-ci change pour chaque réalisation du canal et valeur du SNR. Par conséquent, afin d'utiliser le mapping MSEW avec le précodeur GOPT, nous devrions rechercher le meilleur mapping pour chaque réalisation du canal et chaque SNR. Il n'est donc pas pratique d'utiliser le mapping MSEW pour la constellation reçue du précodeur GOPT. Néanmoins, grâce au précodeur LCOPT proposé, nous pouvons fixer la matrice $\mathbf{\Theta}$, et trouver ensuite le meilleur mapping MSEW pour la forme de la constellation reçue correspondante.

4.3 Sélection de résultats obtenus

Nous considérons un canal MIMO 2×2, généré de façon aléatoire, pour les simulations de Monte-Carlo, c'est-à-dire que chaque élément de $\mathbf{H}$ est distribué $\mathrm{H}_{i,j} \sim \mathcal{CN}(0,1)$. Le canal équivalent $\mathbf{H}_v$ est ensuite obtenu à partir de $\mathbf{H}$ par la transformation (3). L'encodeur FEC utilise le code $(13,15)_{octal}$-RSC, de rendement $1/2$.

Prenons tout d'abord le mapping Gray-M standard. La figure 4.1 montre le taux d'erreur trame (FER) et le taux d'erreur binaire (BER) du système lorsque les précodeurs

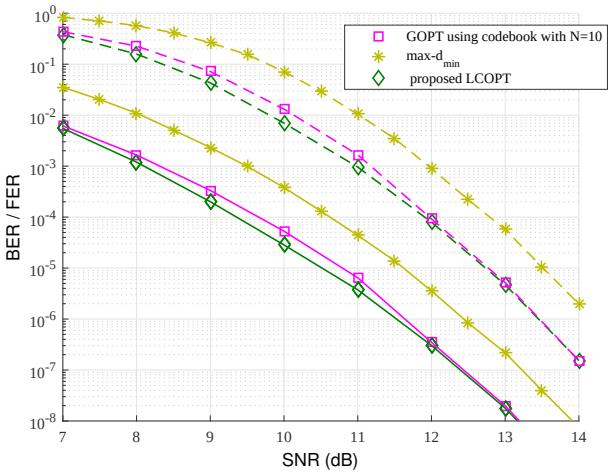


FIGURE 4.1: BER (courbes en trait plein) et FER (courbes en trait discontinu), tous les précodeurs sont utilisés avec le mapping Gray-M.

sont utilisés avec le mapping Gray-M standard. Le précodeur GOPT est simulé à l'aide d'une méthode de construction d'un dictionnaire, qui est aussi proposée dans ce chapitre, ayant une résolution de $N = 10$ précodeurs par valeur du SNR. Nous constatons que le précodeur LCOPT sous-optimal, à faible complexité, présente un taux d'erreur similaire à celui du précodeur GOPT. Notons cependant que dans ce cas le résultat est légèrement meilleur pour le LCOPT que pour le GOPT, car ce dernier n'utilise pas le précodeur optimal, mais un précodeur le plus proche, tiré du dictionnaire. Il est également à remarquer que le LCOPT surpasse le précodeur max-$d_{\min}$, à la fois en termes de FER et de BER.

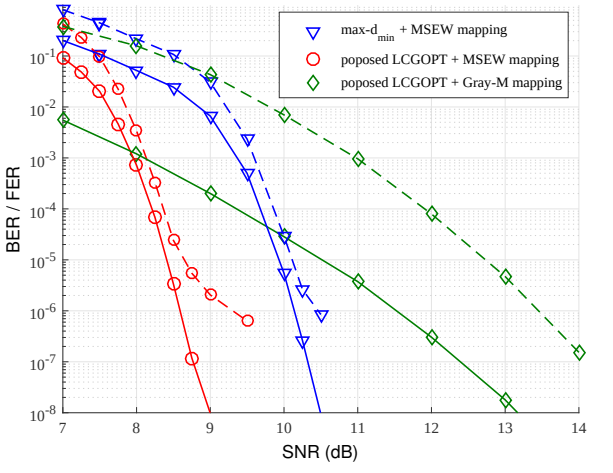


FIGURE 4.2: BER (courbes en trait plein) et FER (courbes en trait discontinu) ; les précodeurs max-$d_{\min}$ et LCOPT sont utilisés avec des mappings MSEW optimaux, tandis que le précodeur GOPT est utilisé avec le mapping Gray-M.

Deuxièmement, nous appliquons les mappings MSEW optimaux des précodeurs max-$d_{\min}$

et LCOPT, au niveau du mappeur symbole MIMO. Les résultats des simulations, présentés sur la figure 4.2, montrent que le précodeur LCOPT surpasse le précodeur max-$d_{\min}$. Sur la même figure nous présentons à nouveau le résultat du précodeur LCOPT avec le mapping Gray-M de la figure 4.1 afin de montrer l'efficacité du mapping MSEW en termes de taux d'erreur. En effet, en utilisant le mapping MSEW, nous remarquons que le précodeur LCOPT améliore considérablement le taux d'erreur par rapport au mapping Gray-M. Ainsi, le gain obtenu est d'environ 4.1 dB pour un BER = 10^{-8} et de plus de 4 dB pour un at FER = 10^{-6}.Rappelons encore une fois que la recherche des mappings MSEW optimaux dans le cas du précodeur GOPT n'est pas pratique puisque celle-ci devrait être effectuée pour chaque réalisation du canal et valeur du SNR. Toutefois, pour une forme de constellation fixée ($\mathbf{\Theta}$ fixée), le mapping MSEW associé au précodeur LCOPT est fixe pour toutes les réalisations du canal, ce qui montre l'avantage du précodeur LCOPT proposé par rapport au précodeur GOPT, dans le cas des mappings optimisés au niveau de la constellation reçue.

4.4 Publication en lien avec ce chapitre

[J2] N.-Q. Nhan, P. Rostaing, K. Amis, L. Collin, and E. Radoi, *Complexity Reduction for the Optimization of Linear Precoders over Random MIMO Channels*, article soumis à la revue IEEE Transactions on Communications.

Chapitre 5

Optimisation du précodage MIMO linéaire avec turbo-égalisation MMSE en réception

5.1 Principales contributions

Dans ce chapitre, notre objectif consiste à optimiser le précodeur $\mathbf{F}_d$ en considérant en réception l'annulation itérative des interférences (turbo-égalisation) basée sur la minimisation de l'erreur quadratique moyenne (MMSE). Nous proposons donc un nouveau précodeur, qui est spécifiquement conçu pour être utilisé avec un turbo-équaliseur au niveau du récepteur. Contrairement aux précodeurs classiques qui maximisent l'information mutuelle entre les symboles à l'entrée et la sortie du canal (capacité du canal), le précodeur proposé vise à maximiser l'information mutuelle entre les symboles à l'entrée du canal et à la sortie de l'égaliseur. Ce précodeur est adapté pour les applications nécessitant une faible complexité, où des codes FEC simples doivent être utilisés.

5.2 Modèle du système

Un système complet, incluant le modèle équivalent décrit par l'équation (3) et un turbo-égaliseur, est illustré sur la figure 5.1, où $L_{\mathrm{A}}^{1}, L_{\mathrm{P}}^{1}$ et L_{E}^{1} sont respectivement les rapports de log-vraisemblance (LLR) a priori, a posteriori, et extrinsèque du convertisseur des symboles en binaire, tandis que les notations équivalentes pour le décodeur BCJR souple sont $L_{\mathrm{A}}^{2}, L_{\mathrm{P}}^{2}$ et L_{E}^{2}. L'annulation des interférences est réalisée par des filtres avant et

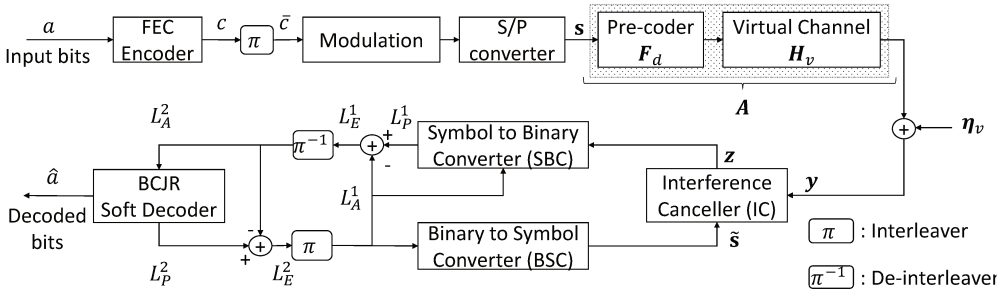


FIGURE 5.1: Système de turbo-égalisation précodé.

arrière, qui sont désignés par $\mathbf{W}$ et $\mathbf{Q}$ respectivement. Les deux filtres sont de taille $b \times b$, où b représente le nombre de flux de données transmis via le précodeur.

Nous appliquons directement ici les expressions de $\mathbf{W}$ et $\mathbf{Q}$, en prenant en compte la présence du précodeur $\mathbf{F}_d$. Des informations plus détaillées sur la turbo-égalisation peuvent être trouvées en [28–30]. Soit $\mathbf{A} = \mathbf{H}_v\mathbf{F}_d$. À la sortie du module d'annulation des interférences (IC) on obtient le vecteur détecté :

$$\mathbf{z} = \mathbf{W}\mathbf{y} - \mathbf{Q}\tilde{\mathbf{s}} \tag{5.1}$$

où $\tilde{\mathbf{s}}$, qui est la sortie du BSC (voir la figure 5.1), est l'estimation du vecteur symbole $\mathbf{s}$, avec $\mathrm{E}[\tilde{\mathbf{s}}\tilde{\mathbf{s}}^\dagger] = \sigma_{\tilde{s}}^2\mathbf{I}_b$ et $\mathrm{E}[\mathbf{s}\tilde{\mathbf{s}}^\dagger] = \sigma_{\tilde{s}}^2\mathbf{I}_b$. Les filtres $\mathbf{W}$ et $\mathbf{Q}$ sont obtenus par l'optimisation du critère MMSE. Soit $\mathbf{B} = (\sigma_s^2 - \sigma_{\tilde{s}}^2)\mathbf{A}\mathbf{A}^\dagger + \sigma_\eta^2\mathbf{I}_b$. La minimisation de l'erreur quadratique moyenne sous la contrainte $\mathbf{Q}_{k,k} = 0, \forall k \in \{1, 2, \dots b\}$ conduit à [31] :

$$\mathbf{W}_{k,:} = \sigma_s^2\mathbf{A}_{:,k}^\dagger\left(\mathbf{B} + \sigma_{\tilde{s}}^2\mathbf{A}_{:,k}\mathbf{A}_{:,k}^\dagger\right)^{-1}, \tag{5.2}$$

et

$$\mathbf{Q}_{k,:} = \mathbf{W}_{k,:}\mathbf{A} - \mathbf{W}_{k,:}\mathbf{A}_{:,k}e_k, \tag{5.3}$$

où e_k est la $k^{\text{ème}}$ ligne de $\mathbf{I}_b$, tandis que $\mathbf{A}_{:,k}$ et $\mathbf{A}_{k,:}$ représentent respectivement la $k^{\text{ème}}$ colonne et la $k^{\text{ème}}$ ligne de la matrice $\mathbf{A}$.

Considérons également l'inégalité suivante [31] :

$$0 < \mathbf{W}_{k,:}\mathbf{A}_{:,k} = \frac{\sigma_{\tilde{s}}^2\mathbf{A}_{:,k}^\dagger\mathbf{B}^{-1}\mathbf{A}_{:,k}}{1 + \sigma_s^2\mathbf{A}_{:,k}^\dagger\mathbf{B}^{-1}\mathbf{A}_{:,k}} = \mu_k < 1. \tag{5.4}$$

La sortie du module IC peut être modélisée sous la forme :

$$z_k = \mu_k s_k + \xi_k, \tag{5.5}$$

où ξ_k est indépendant de s_k, a une moyenne nulle, et la variance $\sigma^2_{\xi_k} = \sigma^2_s \mu_k (1 - \mu_k)$. Le rapport signal à bruit à la sortie du module IC est noté p_k et vaut :

$$p_k = \frac{\mu_k}{1 - \mu_k}. \tag{5.6}$$

Nous utilisons les diagrammes de transfert de l'information extrinsèque (EXIT) pour analyser la convergence du turbo-égaliseur. Une description complète de la façon de calculer les fonctions EXIT est donnée dans [32]. Nous considérons le code RSC de rendement 1/2 défini par ses polynômes générateurs sous forme octale (13, 15). La figure 5.2 illustre un diagramme EXIT complet du turbo-égaliseur, avec multiplexage spatial utilisé au niveau de l'émetteur. Le diagramme EXIT est calculé pour un canal fixe et un SNR = 10 dB.

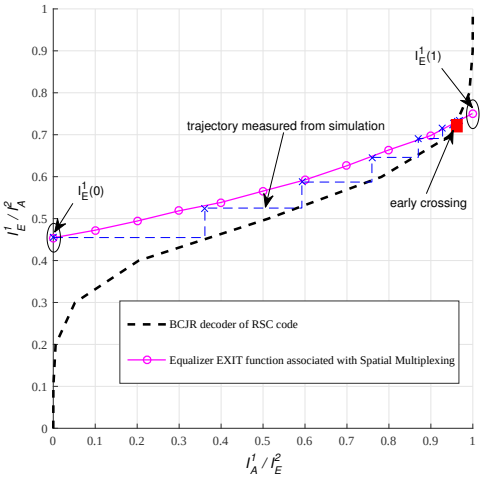


FIGURE 5.2: Diagramme EXIT du turbo-égaliseur en fonction de la trajectoire réelle (obtenu par simulation) pour un canal fixe, SNR = 10 dB, 4-QAM et un code (13,15)-RSC.

5.3 Analyse

Comme illustré sur la figure 5.1, les LLR d'entrée appliqués au décodeur FEC sont calculés à partir de la sortie $\mathbf{z}$ du module d'annulation des interférences. Par conséquent, l'information mutuelle $I(\mathbf{z}, \mathbf{s})$ joue un rôle essentiel concernant le taux d'erreur du système. En outre, à partir de (5.1) il en résulte que l'information mutuelle entre la sortie du canal $\mathbf{y}$ et l'entrée correspondante $\mathbf{s}$, est différente de $I(\mathbf{z}, \mathbf{s})$, c'est-à-dire $I(\mathbf{y}, \mathbf{s}) \neq I(\mathbf{z}, \mathbf{s})$. C'est bien cette relation qui motive la recherche d'un précodeur qui maximise $I(\mathbf{z}, \mathbf{s})$. , sachant que le précodeur GOPT [22] maximise seulement la capacité du canal $I(\mathbf{y}, \mathbf{s})$.

Puisque $\mathbf{z}$ prend en compte l'information a priori ($\tilde{\mathbf{s}}$) du décodeur FEC, il n'est pas possible de trouver un précodeur qui optimise globalement $I(\mathbf{z}, \mathbf{s})$. Cependant, la règle de la chaîne relative à l'information mutuelle, exprimée au niveau symbole, permet de

la décomposer en une somme de M informations mutuelles exprimées au niveau bit, à savoir : $I(\mathbf{z},\mathbf{s}) = \sum_{L=0}^{M-1} I(\mathbf{z},\mathbf{s}|L \text{ bits connu})$ [33, 34], où $M = (\log_2 Q)^b$ est le nombre de bits par symbole mappé.

D'une part, ten-Brink a également montré [34] que $I(\mathbf{z},\mathbf{s}|\text{aucun autre bit connu})$ ($\approx I(\mathbf{z},\mathbf{s}|\tilde{\mathbf{s}} = \mathbf{0})$) et $I(\mathbf{z},\mathbf{s}|\text{tous les autres bits connu})$($\approx I(\mathbf{z},\mathbf{s}|\tilde{\mathbf{s}} = \mathbf{s})$) correspondent respectivement aux points initial et final du diagramme EXIT. D'autre part, les résultats présentés dans [35] montrent que le fait de maximiser le point final du diagramme EXIT de l'égaliseur entraîne un plancher d'erreur bas pour la turbo-égalisation, tandis que la maximisation de son point initial conduit à une convergence rapide. En effet, tel qu'il peut être constaté sur la figure 5.2 que la maximisation de $I_E^1(1)$ évite le croisement prématuré des deux fonctions EXIT, tandis que la maximisation de $I_E^1(0)$ conduit à un bon point initial pour la trajectoire.

En résumé, nous proposons dans ce chapitre un précodeur qui maximise respectivement $I(\mathbf{z},\mathbf{s}|\tilde{\mathbf{s}} = \mathbf{s})$ ($I(\mathbf{z},\mathbf{s})$ à l'état de convergence optimale) et $I(\mathbf{z},\mathbf{s}|\tilde{\mathbf{s}} = \mathbf{0})$ ($I(\mathbf{z},\mathbf{s})$ à l'état initial). L'optimisation est effectuée en deux étapes. Tout d'abord, la priorité est d'optimiser $I(\mathbf{z},\mathbf{s}|\tilde{\mathbf{s}} = \mathbf{s})$ afin de minimiser le plancher d'erreur. Cette étape permet de trouver les deux premiers paramètres (ψ,θ) de $\mathbf{F}_d$ dans (4). Deuxièmement, avec les valeurs de (ψ,θ) déterminées précédemment, nous optimisons le dernier paramètre ϕ de $\mathbf{F}_d$ dans (4) afin d'améliorer $I(\mathbf{z},\mathbf{s}|\tilde{\mathbf{s}} = \mathbf{0})$. La solution du problème d'optimisation ci-dessus pour le précodeur proposé vaut $(\psi = 0, \theta = \pi/4, \phi = \pi/6)$. Puisque le nouveau précodeur est optimisé à l'état de convergence optimal (obtenu pour $\tilde{\mathbf{s}} = \mathbf{s}$ et également connu sous le nom de borne Génie), nous appelons celui-ci précodeur Génie optimisé (*Genie-optimized*).

5.4 Sélection de résultats obtenus

Nous illustrons maintenant, par quelques exemples, les avantages du précodeur Génie optimisé par rapport à d'autres précodeurs existants, en termes de taux d'erreur. Le premier est le précodeur [22]. A notre connaissance, ce précodeur permet d'obtenir la meilleure capacité sur des canaux MIMO complexes, car celui-ci vise à maximiser l'information mutuelle entre l'alphabet fini d'entrée et la sortie correspondante du canal. Le deuxième précodeur, qui a des expressions non diagonales distinctes, indépendantes du SNR, est le précodeur max-$d_{\min}$ [16]. Ce précodeur vise à maximiser la distance Euclidienne minimale (dmin) entre les points de la constellation reçue ($\mathbf{x}_i = \mathbf{H}_v\mathbf{F}_d\mathbf{s}_i$). Par conséquent, puisque maximiser $d_{\min}$ est équivalent à maximiser la limite inférieure de la capacité du canal MIMO à SNR élevé [25], le précodeur max-$d_{\min}$ permet également d'obtenir une capacité de transmission très proche de celle assurée par un précodeur GOPT à haut SNR.

Les canaux aléatoires sont considérés pour montrer l'avantage du précodeur Génie optimisé pour toute valeur de γ. Avec la définition du SNR donnée par (6), l'énergie du canal $\|\mathbf{H}\|_F^2 = \rho^2$ est intégrée dans le SNR. Par conséquent, le taux d'erreur ne dépend pas de ρ^2. Ainsi, en moyennant sur les canaux générés aléatoirement, la courbe de performance représentée sur la figure 5.3 montre le taux d'erreur moyen en fonction de γ. Tel qu'il peut être constaté, la solution proposée conduit à un gain de plus de 4 dB pour un BER $= 10^{-3}$ et de 2 dB pour un BER $= 10^{-6}$ respectivement, comparée au multiplexage spatial et au max-$d_{\min}$ respectively. Comparé au précodeur GOPT, le précodeur Génie optimisé permet de réaliser respectivement un gain d'environ 1 dB et de 2 dB pour un BER $= 10^{-6}$ et BER $= 10^{-7}$. En conclusion, un gain considérable peut être obtenu en utilisant un précodeur, surtout le précodeur Génie optimisé, comparativement au multiplexage spatial (aucune CSI exploitée du côté émetteur). Le précodeur Génie optimisé atteint toujours très tôt la borne Génie et montre des gains de performance comparé à d'autres précodeurs.

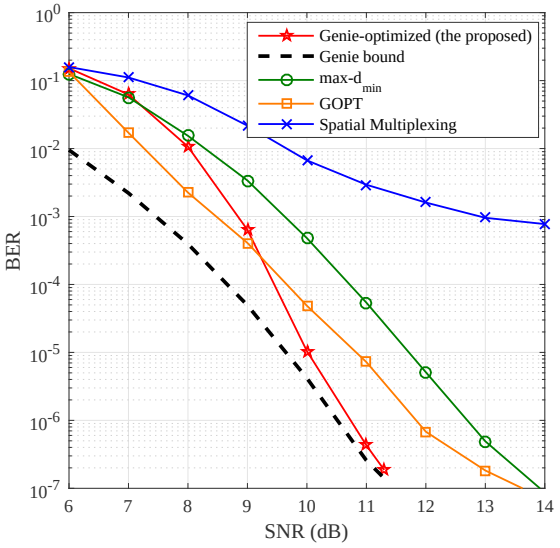


FIGURE 5.3: BER moyen sur des canaux aléatoires, 4-QAM et un code (13,15)-RSC.

5.5 Publications en lien avec ce chapitre

[J3] N.-Q. Nhan, P. Rostaing, K. Amis, L. Collin, and E. Radoi, *Optimization of linear MIMO precoding assuming MMSE-based turbo equalization*, article soumis à la revue IEEE Transactions on Wireless Communications.

[C3] N.-Q. Nhan, P. Rostaing, K. Amis, L. Collin, and E. Radoi, *Association and Joint Optimization of* max-$d_{\min}$ *Precoder with MIMO Turbo Equalization*, actes de la conférence internationale IEEE GLOBECOM, Decembre 2015, pp. 1-6.

Chapitre 6

Turbo-détection des codes NB-LDPC dans les systèmes MIMO précodés

6.1 Principales contributions

Depuis qu'ils ont été redécouverts par Mackay et al. [36], les codes LDPC ont été étudiés de manière approfondie dans le domaine du codage de canal. Parmi les codes LDPC, les codes non-binaires (NB) surpassent leurs homologues binaires, plus particulièrement pour des longueurs de code courtes à modérées [37]. Les codes NB-LDPC montrent également des avantages lorsqu'ils sont utilisés dans les systèmes MIMO. En effet, rappelons que dans les communications MIMO conventionnelles, utilisant le codage binaire, les mots de code binaires sont regroupés et mappés dans les symboles d'une modulation Q-aire. Ensuite, ces symboles sont transformés en vecteurs symboles MIMO par une conversion série-parallèle. Les vecteurs symbole ainsi obtenus constituent une constellation MIMO. La taille de cette constellation est donc égale à Q^b, où b est le nombre de flux de données (couches) du système MIMO. Contrairement à ce mapping binaire classique, les recherches présentées en [38,39] montrent que l'ordre du GF des codes NB-LDPC peut être choisi égal à la taille de la constellation MIMO. Par la suite, chaque symbole GF est directement mappé sur un vecteur symbole. Ce mapping, que nous appelons « mapping un à un », est utile pour associer les codes NB-LDPC aux communications MIMO.

Les communications sans fil modernes, en particulier les réseaux cellulaires de cinquième génération (5G), exigent des débits de données élevés et une faible latence de transmission [40]. Ces contraintes entraînent l'utilisation extensive des modulations d'ordres élevés [41], ce qui conduit à une augmentation exponentielle de la taille de la constellation

MIMO. Par conséquent, des GF d'ordres très élevés seraient nécessaires dans le cas d'un mapping un à un. Malheureusement, les codes NB-LDPC construits sur des GF d'ordre élevé induisent une complexité importante au niveau du décodeur NB-LDPC, ce qui augmente la latence du système.

Pour surmonter ce problème, nous considérons dans ce chapitre un « mapping multiple » qui transforme plusieurs symboles définis sur un GF d'ordre réduit en un symbole MIMO. Ensuite, nous étudions le récepteur itératif ou turbo-détecteur, qui effectue conjointement le décodage et la détection. Il est démontré que, pour le même ordre de modulation, le mapping multiple utilisé avec le turbo-détection réduit de manière significative la complexité du récepteur par rapport au mapping un à un. Cependant, cet avantage est associé à deux principaux défis. Le premier défi est que le mapping multiple produit des interférences inter-symboles, ce qui conduit à une dégradation des performances au niveau du plancher d'erreur. Le deuxième défi concerne l'optimisation du nombre d'itérations entre le détecteur et le décodeur dans le turbo-détection, ainsi qu'à l'intérieur du décodeur pour réduire encore davantage la complexité du décodage.

Les contributions de ce chapitre peuvent être résumées de la manière suivante. Tout d'abord, nous proposons l'utilisation des précodeurs MIMO pour surmonter l'inconvénient du turbo-détection concernant les performances, surtout au niveau du plancher d'erreur. Ces précodeurs sont sélectionnés parmi ceux présentés dans la littérature, selon le critère de la maximisation de l'information mutuelle du canal, qui est bien adapté pour les systèmes MIMO encodés. Des gains de performances importants sont réalisés en utilisant ces précodeurs. Deuxièmement, nous montrons qu'ils permettent également de réduire la complexité du récepteur et de minimiser le nombre d'itérations internes du décodeur. L'optimisation de la configuration turbo-détection est finalement discutée.

6.2 Modèle du système

Un encodeur NB-LDPC sur un GF d'ordre q (GF(q)) est utilisé pour le codage du canal (figure 6.1). Les mots de code obtenus $\mathbf{u}$ sont ensuite entrelacés au niveau symbole. Les mots de code en sortie de l'entrelaceur sont notés $\bar{\mathbf{u}}$. Au niveau du mappeur, les symboles obtenus par une modulation QAM Q-aire sont regroupés dans un vecteur $\mathbf{s}$ de taille $b \times 1$. Toutes les valeurs possibles de $\mathbf{s}$ définissent une constellation de symboles MIMO $\mathcal{S} = \{\mathbf{s}_0, \ldots, \mathbf{s}_k, \ldots, \mathbf{s}_{Q^b-1}\}$. Ensuite, les symboles GF de $\bar{\mathbf{u}}$ sont regroupés et mappés sur un symbole MIMO $\mathbf{s} \in \mathcal{S}$ selon une règle de mapping $\mathcal{X}$. Le symbole MIMO $\mathbf{s}$ est precodé par un précodeur $\mathbf{F}_d$ avant d'être transmis dans le canal MIMO. Bien que n_s symboles de $\bar{\mathbf{u}}$ puissent être mappés sur plusieurs symboles MIMO au lieu d'un seul, ce

mapping n'est toutefois pas intéressant puisqu'il induit une complexité élevée au niveau du démappeur.

Le récepteur comprend un détecteur MAP (maximum a posteriori) souple et un décodeur NB-LDPC souple, qui échangent de façon itérative les vecteurs LLR. Cette boucle itérative est ainsi appelée « boucle externe », le nombre d'itérations correspondant étant noté n_{out}. De même, la boucle itérative à l'intérieur du décodeur est appelée « boucle interne », le nombre d'itérations correspondant étant noté n_{in}.

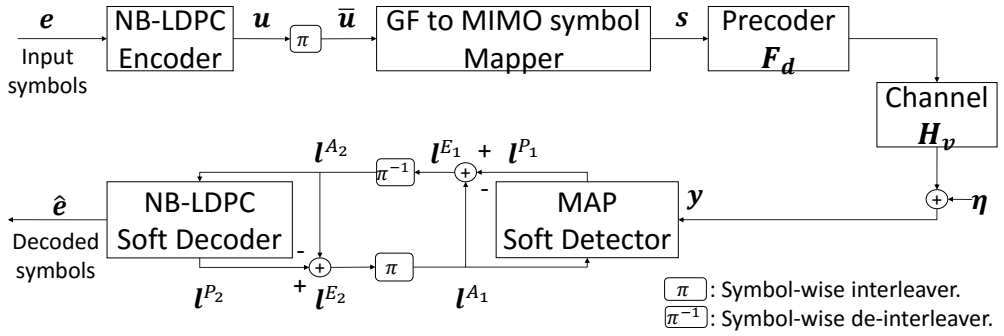


FIGURE 6.1: Modèle du système.

6.3 Analyse

L'analyse du récepteur itératif est effectuée en utilisant le diagramme EXIT pour le décodage itératif des symboles [42]. Nous considérons le décodeur avec un nombre maximal donné n_{in} ($n_{\text{in}}^{\max}$), en tant qu'élément de décodage individuel dans le diagramme, l'autre élément étant constitué par le détecteur. A titre d'exemple, prenons au hasard le canal $\mathbf{H}_{\text{ex}} = [0.4067 - 0.1801i \; 0.0419 + 0.2015i; \; -0.8206 - 0.0268i \; 0.2896 - 0.0428i]$. La figure 6.2 montre le diagramme EXIT pour ce canal et un SNR= 18 dB. Les courbes à trait discontinu représentent les fonctions EXIT du décodeur pour différentes valeurs de $n_{\text{in}}^{\max}$, tandis que les courbes à trait plein représentent les fonctions EXIT du détecteur lorsque le multiplexage spatial (SM), ou les précodeurs max-$d_{\min}$ et GOPT sont utilisés au niveau de l'émetteur. Il est à noter que seules les fonctions EXIT du détecteur varient avec le SNR. Les points initial et final d'une fonction EXIT du détecteur sont notés respectivementy I_0 et I_1.

Comme illustré sur la figure 6.2, les fonctions EXIT du décodeur commencent à converger à $n_{\text{in}}^{\max} = 10$. Nous constatons tout d'abord que la fonction EXIT du décodeur peut être utilisée pour prédire $n_{\text{in}}^{\max}$ dans le schéma non-itératif classique. Dans le schéma turbo-détection utilisé avec le SM, bien que la limitation de $n_{\text{in}}^{\max}$ diminue la complexité, les fonctions EXIT du SM et du décodeur se croisent trop tôt, ce qui conduit à de faibles performances en termes de taux d'erreur. Cependant, ce problème du croisement

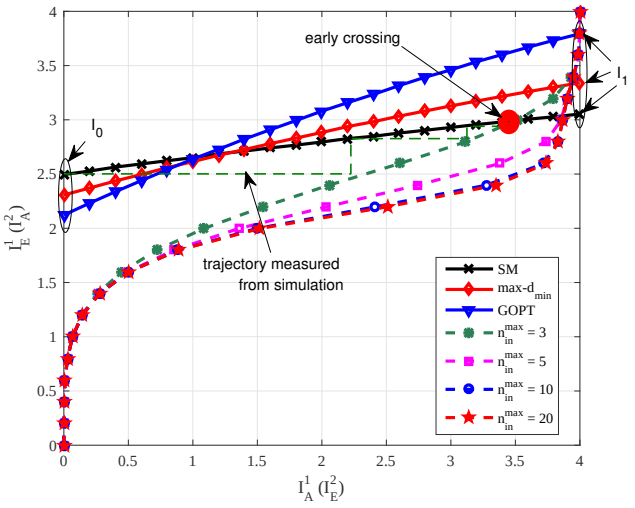


FIGURE 6.2: Diagrammes EXIT pour une réalisation du canal et un SNR = 18 dB.

prématuré des deux courbes peut être réglé en utilisant les précodeurs. Ainsi, la figure 6.2 montre que les valeurs de I_1 pour les précodeurs GOPT et max-$d_{\min}$ sont bien plus élevées que pour le SM, ce qui prédit des planchers d'erreur meilleurs pour le système précodé.

6.4 Sélection de résultats obtenus

Nous illustrons maintenant, par quelques exemples, les avantages de l'utilisation des précodeurs en termes de taux d'erreur et de réduction de complexité. Les simulations de Monte Carlo ont été effectuées pour des systèmes MIMO avec $n_T = 2, n_R = 2$, $b = 2$ flux de données, $Q = 16$ (modulation 16-QAM), GF($q = 16$) et $n_s = 2$ symboles GF par symbole MIMO, et algorithme de décodage log-BP [45]. La taille de la constellation MIMO est donc égale à $Q^b = 256$. Nous rappelons que la matrice NB-LDPC est déterminée à partir des matrices conçues dans le cadre du projet DAVINCI [43, 44] avec ($N_v = 192, N_c = 96, d_c = 4$, $d_v = 2$, longueur du plus court cycle ou *girth* = 14). Le nombre maximum d'itérations dans la boucle externe est fixé à 10.

La figure 6.3 montre le FER du système en fonction du SNR pour le canal $\mathbf{H}_{\text{ex}}$, qui a fait l'objet de l'analyse des diagrammes EXIT, dans le cas où le SM (courbes à trait discontinu), le précodeur max-$d_{\min}$ (courbes à trait plein) et le précodeur GOPT (courbes en pointillés) sont utilisés au niveau de l'émetteur. Les courbes correspondant à un même $n_{\text{in}}^{\max}$ partagent le même marqueur. Nous pouvons constater que le plancher d'erreur du système est considérablement réduit grâce aux précodeurs. Avec le même $n_{\text{in}}^{\max}$, le taux d'erreur est amélioré en présence des précodeurs.

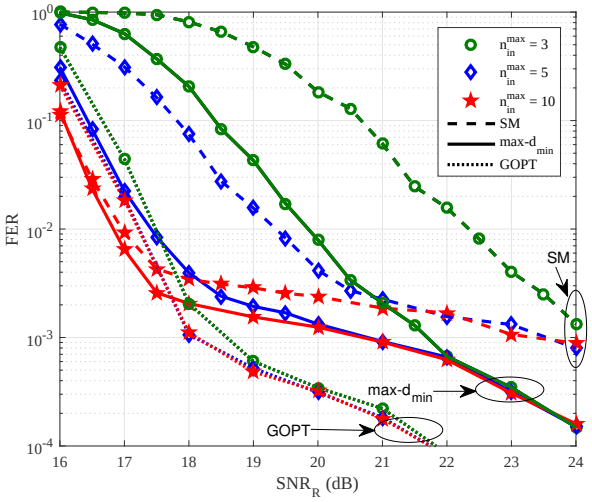


FIGURE 6.3: FER pour le canal $\mathbf{H}_{\text{ex}}$.

Par exemple, pour $n_{\text{in}}^{\max} = 5$ et FER $= 10^{-3}$, le précodeur GOPT réalise un gain d'environ 6 dB et 3 dB respectivement par rapport au SM et au précodeur max-$d_{\min}$. En outre, pour le SM et précodeur max-$d_{\min}$, les courbes correspondant aux différents $n_{\text{in}}^{\max}$ convergent à SNR élevé. Ce comportement peut être expliqué par l'augmentation de I_1 en haut à droite du diagramme EXIT à SNR élevé. Dans le cas du précodeur GOPT, cette convergence se produit à un SNR plus faible (voir la figure 6.2). Tous les résultats des simulations pour le canal $\mathbf{H}_{\text{ex}}$ sont conformes à l'analyse des diagrammes EXIT.

6.5 Publication en lien avec ce chapitre

[J4] N.-Q. Nhan, P. Rostaing, K. Amis, L. Collin, E. Radoi, and Y. Rosa Zheng, *Turbo detection of NB-LDPC codes in precoded MIMO systems*, article soumis à la revue IEEE Transactions on Vehicular Technology.

Chapitre 7

Algorithme de décodage parallèle par retournement des symboles et votes multiples pour les codes LDPC non-binaires

7.1 Principales contributions

Les algorithmes de décodage des codes NB-LDPC peuvent être regroupés en trois classes principales : décodage à décision dure [46], décodage à décision souple [45, 47–49], et décodage hybride (également connu comme décodage basé sur la fiabilité des décisions dures) [50–53]. Du point de vue de la mise œuvre, la première classe présente la complexité de décodage la plus réduite. Cependant, sa simplicité entraîne des performances relativement faibles, la différence par rapport à la deuxième classe étant typiquement de l'ordre de plusieurs dB. Le décodage à décision souple fournit les meilleures performances, mais conduit au niveau de complexité le plus élevé. Le décodage hybride se situe entre les deux, en réalisant un bon compromis entre les performances et la complexité du décodeur.

Parmi les algorithmes basés sur la fiabilité, l'algorithme de décodage parallèle par retournement des symboles (PSFD), récemment introduit dans [52] offre un des meilleurs compromis entre performance et complexité. En utilisant une fonction de retournement qui combine le message de parité pondéré issu des sommes de contrôle normalisées et le message de variable issu de la séquence reçue, l'algorithme identifie, à chaque itération de décodage, les symboles relativement non-fiables dans la séquence des symboles

obtenus par décision dure, et les décode en utilisant les symboles retournés correspondants. Les symboles non-fiables sont trouvés à l'aide d'un système de vote, où chaque nœud de contrôle (CN) non-satisfait donne une voix au nœud de variable (VN) ayant la plus grande valeur de la fonction de retournement contrôlé par celui-ci. Les nœuds de variable accumulent ainsi des voix et ceux qui dépassent un nombre total de voix au-dessus d'un seuil prédéfini sont identifiés comme non-fiables. En jouant simplement avec les fiabilités des décisions dures, l'algorithme PSFD est un bon choix pour décoder des codes NB-LDPC haut débit, définis sur des corps finis. Cependant, les résultats des simulations montrent que l'algorithme PSFD est adapté pour décoder seulement des codes NB-LDPC réguliers, dont les matrices de contrôle de parité ont des colonnes de poids élevés (au moins 8). En outre, l'algorithme PSFD utilise un seuil pour le nombre de voix qui dépend du code et qui doit être optimisé par simulation. Le calcul de sa fonction de retournement implique un couple de facteurs d'échelle, (η, λ), qui dépendent également du code et qui sont aussi à optimiser par simulation.

Ce chapitre propose un nouvel algorithme, appelé PSFD à votes multiples (MV-PSFD). Comparé au PSFD, celui-ci introduit une méthode pour l'estimation de l'erreur, qui permet d'éviter l'utilisation du seuil pour le nombre de voix, et passe les votes multiples des CN non-satisfaits aux VN correspondants. L'algorithme proposé surpasse significativement le PSFD pour des matrices de contrôle de parité avec des colonnes de faible poids, comme indiqué par les résultats des simulations. A noter que ces améliorations sont obtenues pour une complexité similaire.

7.2 Sélection de résultats obtenus

Nous présentons les résultats des simulations pour deux codes, désignés par Code 1 et Code 2. Code 1 est caractérisé par $(d_v = 3, d_c = 6)$ [54] et Code 2 par $(d_v = 5, d_c = 10)$ [55] ; les deux sont des codes NB-LDPC réguliers (102, 204) sur GF(2^4) construits avec la méthode de Gallager, et dont la matrice de contrôle de parité vérifie la contrainte RC.

La figure 7.1 montre les performances en termes de BER et de FER du Code 1. Il peut être constaté que, contrairement à l'algorithme PSFD, les performances de l'algorithme proposé ne dépendent pas des paramètres η et λ. En outre, l'algorithme proposé surpasse le PSFD optimal d'environ 0.5 dB pour un BER de 10^{-5}. A noter que les courbes de l'algorithme de référence sont tracées pour le rapport $\frac{\eta}{\lambda}$ optimisé et V_{th}.

La figure 7.2 montre les performances du Code 2 en termes de BER et de FER. Contrairement au cas précédent, les performances de l'algorithme proposé sont sensibles aux

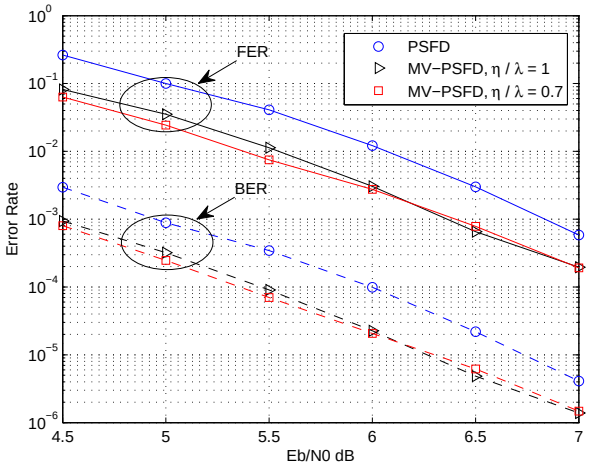


FIGURE 7.1: Performances des algorithmes MV-PSFD et PSFD en termes de FER (courbes en trait plein) et BER (courbes en trait discontinu), en fonction du SNR normalisé, pour le Code 1.

paramètres η et λ. Cependant, en utilisant la valeur optimale de $\frac{\eta}{\lambda}$, l'algorithme MV-PSFD présente des gains d'environ 0.3 dB et 0.4 dB respectivement pour des valeurs de BER=10^{-5} et BER=10^{-6}. Il peut également être constaté que, lorsque l'algorithme MV-PSFD est utilisé sans les valeurs optimales de η et λ, ses performances sont proches de celles de l'algorithme de référence.

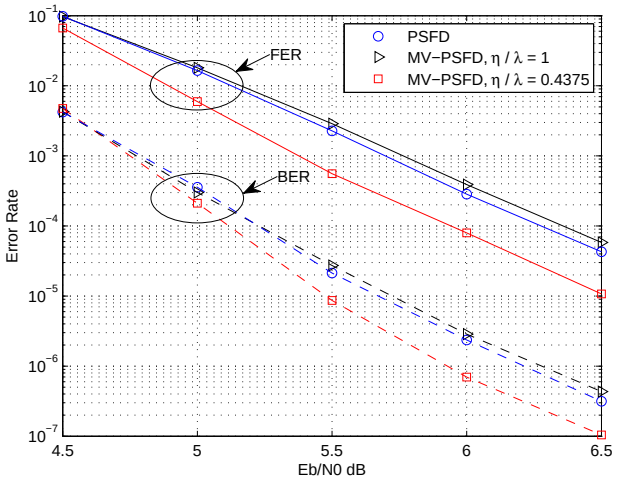


FIGURE 7.2: Performances des algorithmes MV-PSFD et PSFD en termes de FER (courbes en trait plein) et BER (courbes en trait discontinu), en fonction du SNR normalisé, pour le Code 2.

7.3 Publication en lien avec ce chapitre

[J5] N.-Q. Nhan, Telex M. N. Ngatched, Octavia A. Dobre, P. Rostaing, K. Amis, and E. Radoi. Multiple-Votes Parallel Symbol-Flipping Decoding Algorithm for Non-Binary LDPC, IEEE Communication Letters, pp. 905 - 908, Vol. 19, June 2015.

Conclusions

Nous avons investigué, dans cette thèse, des schémas de communications qui utilisent des codes FEC simples pour le codage du canal, afin de réduire la complexité et la latence des systèmes sans fil, dans le cadre des normes LTE et LTE-A. Nous avons ensuite optimisé leurs performances en termes de taux d'erreur, en vue de leur utilisation avec ces normes.

D'une part, nous avons étudié la concaténation de codes FEC binaires et précodeurs MIMO linéaires. Premièrement, en considérant un récepteur à turbo-détection, nous avons proposé différentes solutions de précodage et mapping afin d'améliorer le taux d'erreur du système. Deuxièmement, nous avons proposé un précodeur spécialement conçu pour améliorer le taux d'erreur lorsque le récepteur est un turbo-égaliseur.

D'autre part, nous avons examiné les codes FEC NB-LDPC. Nous avons introduit une technique de mapping multiple, qui permet une réduction importante de la complexité du décodage NB-LDPC. Ensuite, nous avons proposé d'utiliser des précodeurs linéaires MIMO pour améliorer le taux d'erreur du système, ainsi que réduire encore davantage la complexité du récepteur NB-LDPC. Enfin, un nouvel algorithme de décodage à faible complexité a été proposé pour achever l'étude sur la réduction de la complexité des systèmes de communications qui utilisent des codes NB-LDPC.

Les résultats obtenus pendant la thèse ont été présentés dans un article revue et trois articles publiés dans les actes de trois conférences internationales. Quatre autres articles sont également en cours de soumission dans différents journaux.

Bibliographie

[1] B. Bangerter, S. Talwar, R. Arefi, and K. Stewart, "Networks and devices for the 5G era," *IEEE Commun. Magazine*, vol. 52, no. 2, pp. 90–96, 2014.

[2] C. Cox, *An introduction to LTE : LTE, LTE-advanced, SAE and 4G mobile communications.* John Wiley & Sons, 2012.

[3] A. Ghosh and R. Ratasuk, *Essentials of LTE and LTE-A.* Cambridge University Press, 2011.

[4] K. Farooq, *LTE for 4G mobile broadband.* Cambridge University Press, 2009.

[5] I. F. Akyildiz, D. M. Gutierrez-Estevez, and E. C. Reyes, "The evolution to 4G cellular systems : LTE-advanced," *Elsevier Physical Commun.*, vol. 3, no. 4, pp. 217–244, 2010.

[6] A. Gupta and R. K. Jha, "A survey of 5G network : architecture and emerging technologies," *IEEE Access*, vol. 3, pp. 1206–1232, 2015.

[7] M. Agiwal, A. Roy, and N. Saxena, "Next generation 5G wireless networks : a comprehensive survey," *IEEE Commun. Surveys & Tutorials*, 2016.

[8] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs. Tech. J.*, vol. 1, no. 2, pp. 41–59, 1996.

[9] E. Telatar, "Capacity of multi-antenna gaussian channels," *European Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, 1999.

[10] Q. Li, G. Li, W. Lee, M. Lee, D. Mazzarese, B. Clerckx, and Z. Li, "MIMO techniques in WiMAX and LTE : a feature overview," *IEEE Commun. Mag.*, vol. 48, no. 5, pp. 86–92, 2010.

[11] A. J. Paulraj, D. A. Gore, R. U. Nabar, and H. Bolcskei, "An overview of MIMO communications - a key to gigabit wireless," *Proc. of the IEEE*, vol. 92, no. 2, pp. 198–218, 2004.

[12] S. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, 1998.

[13] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block coding for wireless communications : performance results," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 3, pp. 451–460, 1999.

[14] P. Bouvet, M. Hélard, and V. Le Nir, "Low complexity iterative receiver for linear precoded MIMO systems," in *IEEE Symposium on Spread Spectrum Techniques and Applications*, 2004, pp. 17–21.

[15] P. Bouvet, M. Hélard, J. L. Masson, and C. Langlais, "Iterative receiver for linear precoded MIMO systems," in *IEEE ITG-Conference on Source and Channel Coding*, 2006, pp. 1–6.

[16] L. Collin, O. Berder, P. Rostaing, and G. Burel, "Optimal minimum distance-based precoder for MIMO spatial multiplexing systems," *IEEE Trans. Signal Processing*, vol. 52, no. 3, pp. 617–627, 2004.

[17] P. Rostaing, O. Berder, G. Burel, and L. Collin, "Minimum BER diagonal precoder for MIMO digital transmissions," *Signal Processing*, vol. 82, no. 10, pp. 1477–1480, 2002.

[18] H. Sampath, P. Stoica, and A. Paulraj, "Generalized linear precoder and decoder design for MIMO channels using the weighted MMSE criterion," *IEEE Trans. Commun.*, vol. 49, no. 12, pp. 2198–2206, 2001.

[19] P. Stoica and G. Ganesan, "Maximum-SNR spatial-temporal formatting designs for MIMO channels," *IEEE Trans. Signal Processing*, vol. 50, no. 12, pp. 3036–3042, 2002.

[20] A. Lozano, A. M. Tulino, and S. Verdú, "Optimum power allocation for parallel gaussian channels with arbitrary input distributions," *IEEE Trans. Info. Theory*, vol. 52, no. 7, pp. 3033–3051, 2006.

[21] M. S. Hassan and K. Amis, "On the design of full-rate full-diversity space-time block codes for MIMO systems with a turbo minimum mean square error equaliser at the receiver side," *IET Communications*, vol. 6, no. 18, pp. 3065–3074, 2012.

[22] C. Xiao, Y. R. Zheng, and Z. Ding, "Globally optimal linear precoders for finite alphabet signals over complex vector gaussian channels," *IEEE Trans. Signal Processing*, vol. 59, no. 7, pp. 3301–3314, 2011.

[23] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. on Info. Theory*, pp. 284–287, Mar. 1974.

[24] J. Tan and G. L. Stuber, "Analysis and design of symbol mappers for iteratively decoded BICM," *IEEE Trans. on Wireless Commun.*, vol. 4, no. 2, pp. 662–672, 2005.

[25] W. Zeng, C. Xiao, and J. Lu, "A low-complexity design of linear precoding for MIMO channels with finite-alphabet inputs," *IEEE Wireless Commun. Lett.*, vol. 1, no. 1, pp. 38–41, 2012.

[26] N.-Q. Nhan, P. Rostaing, K. Amis, L. Collin, and E. Radoi, "Optimized MIMO symbol mapping to improve the turbo cliff region of iterative precoded MIMO detection," in *The 23rd European Signal Processing Conf. (EUSIPCO)*, 2015, pp. 909–913.

[27] Q.-T. Ngo, O. Berder, and P. Scalart, "Minimum euclidean distance based precoders for MIMO systems using rectangular QAM modulations," *IEEE Trans. Signal Processing*, vol. 60, no. 3, pp. 1527–1533, 2012.

[28] M. Tuchler, A. C. Singer, and R. Koetter, "Minimum mean squared error equalization using a priori information," *IEEE Trans. Signal Processing*, vol. 50, no. 3, pp. 673–683, 2002.

[29] M. Tuchler, R. Koetter, and A. C. Singer, "Turbo equalization : principles and new results," *IEEE Trans. Commun.*, vol. 50, no. 5, pp. 754–767, 2002.

[30] C. Laot, R. Le Bidan, and D. Leroux, "Low-complexity MMSE turbo equalization : a possible solution for edge," *IEEE Trans. Wireless Commun.*, vol. 4, no. 3, pp. 965–974, 2005.

[31] K. Amis, G. Sicot, and D. Leroux, "Reduced complexity near-optimal iterative receiver for WiMAX full-rate space-time code," in *IEEE Int. Symp. on Turbo Codes and Related Topics*, 2008, pp. 102–106.

[32] J. Hagenauer, "The EXIT chart-introduction to extrinsic information transfer in iterative processing," in *Proc. 12th EUSIPCO*, 2004, pp. 1541–1548.

[33] S. ten Brink, J. Speidel, and R.-H. Yan, "Iterative demapping and decoding for multilevel modulation," in *IEEE GLOBECOM*, vol. 1, 1998, pp. 579–584.

[34] 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, 2000.

[35] S.-J. Lee, A. C. Singer, and N. R. Shanbhag, "Linear turbo equalization analysis via BER transfer and EXIT charts," *IEEE Trans. Signal Processing*, vol. 53, no. 8, pp. 2883–2897, 2005.

[36] D. J. MacKay and R. M. Neal, "Near shannon limit performance of low density parity check codes," *Electron. Lett.*, vol. 32, no. 18, pp. 1645–1646, 1996.

[37] M. C. Davey and D. J. MacKay, "Low-density parity-check codes over GF(q)," *IEEE Commun. Lett.*, vol. 2, no. 6, pp. 165–167, Jun. 1998.

[38] S. Pfletschinger and D. Declercq, "Getting closer to MIMO capacity with non-binary codes and spatial multiplexing," in *IEEE GLOBECOM*, 2010, pp. 1–5.

[39] ——, "Non-binary coding for vector channels," in *IEEE SPAWC*, 2011, pp. 26–30.

[40] J. Rodriguez, *Fundamentals of 5G Mobile Networks*. John Wiley & Sons, 2015.

[41] H. Kim, "Coding and modulation techniques for high spectral efficiency transmission in 5G and satcom," in *EUSIPCO*, vol. 1, 2015, pp. 2746–2750.

[42] J. Kliewer, S. X. Ng, and L. Hanzo, "On the computation of EXIT characteristics for symbol-based iterative decoding," in *Int. Symp. on Turbo Codes&Related Topics*, 2006, pp. 1–6.

[43] C. Poulliat, M. Fossorier, and D. Declercq, "Design of regular (2, d_c)-LDPC codes over GF(q) using their binary images," *IEEE Trans. Commun.*, vol. 56, no. 10, pp. 1626–1635, 2008.

[44] S. Pfletschinger, A. Mourad, E. Lopez, D. Declercq, and G. Bacci, "Performance evaluation of non-binary LDPC codes on wireless channels," *ICT Mobile Summit*, 2009.

[45] H. Wymeersch, H. Steendam, and M. Moeneclaey, "Log-domain decoding of LDPC codes over GF(q)," in *IEEE ICC*, 2004, pp. 772–776.

[46] X. Zhang, F. Cai, and S. Lin, "Low-complexity reliability-based message-passing decoder architectures for non-binary LDPC codes," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 20, no. 11, pp. 1938–1950, Jul. 2012.

[47] D. Declercq and M. Fossorier, "Decoding algorithms for non-binary LDPC codes over GF(q)," *IEEE Trans. Commun.*, vol. 55, no. 4, pp. 633–643, Apr. 2007.

[48] E. Boutillon and L. Conde-Canencia, "Bubble check : a simplified algorithm for elementary check node processing in extended min-sum non-binary LDPC decoders," *IET Elect. Lett.*, vol. 46, no. 9, pp. 633–634, Apr. 2010.

[49] C.-L. Wang, X. Chen, Z. Li, and S. Yang, "A simplified min-sum decoding algorithm for non-binary LDPC codes," *IEEE Trans. Commun.*, vol. 61, no. 1, pp. 24–32, Feb. 2013.

[50] C. Chen, B. Bai, X. Ma, and X. Wang, "A symbol-reliability based message-passing decoding algorithm for non-binary LDPC codes over finite fields," in *Proc. IEEE ISTC*, 2010, pp. 251–255.

[51] B. Liu, J. Gao, W. Tao, and G. Dou, "Weighted symbol-flipping decoding algorithm for non-binary ldpc codes with flipping patterns," *Journal of System Eng. and Elect.*, vol. 22, no. 5, pp. 848–855, Oct. 2011.

[52] C.-C. Huang, C.-J. Wu, C.-Y. Chen, and C.-C. Chao, "Parallel symbol-flipping decoding for non-binary LDPC codes," *IEEE Commun. Lett.*, vol. 17, no. 6, pp. 1228–1231, Jun. 2013.

[53] F. Garia-Herrero and D. Declercq, "Non-binary LDPC decoder based on symbol flipping with multiple votes," *IEEE Commun. Lett.*, vol. 18, no. 5, pp. 749–752, May 2014.

[54] David Mackay's code resource. [Online]. Available : http://www.inference.phy.cam.ac.uk/mackay/codes/EN/C/204.33.486.

[55] David Mackay's code resource. [Online]. Available : http://www.inference.phy.cam.ac.uk/mackay/codes/EN/C/204.55.153.

RESUME

Les standards « Long-term evolution » (LTE) et LTE-Advanced (LTE-A) devraient influencer fortement l'avenir de la cinquième génération (5G) des réseaux mobiles. Ces normes exigent de hauts débits de données et une qualité de service de très bon niveau, ce qui permet d'assurer un faible taux d'erreur, avec une faible latence. Par ailleurs, la complexité doit être limitée. Dans le but de déterminer des solutions technologiques modernes qui satisfont ces contraintes fortes, nous étudions dans la thèse des systèmes de communication sans fil MIMO codés. D'abord, nous imposons un simple code convolutif récursif systématique (RSC) pour limiter la complexité et la latence. En considérant des récepteurs itératifs, nous optimisons alors la performance en termes de taux d'erreur de ces systèmes en définissant un précodage linéaire MIMO et des techniques de mapping appropriées. Dans la deuxième partie de la thèse, nous remplaçons le RSC par un LDPC non-binaire (NB-LDPC). Nous proposons d'utiliser les techniques de précodage MIMO afin de réduire la complexité des récepteurs des systèmes MIMO intégrant des codes NB-LDPC. Enfin, nous proposons également un nouvel algorithme de décodage itératif à faible complexité adapté aux codes NB-LDPC.

Mots clefs : ***Précodage linéaire MIMO, information mutuelle, mapping des symboles, turbo-détection, turbo-égalisation, codes LDPC non-binaires, diagrammes EXIT.***

ABSTRACT

The long-term evolution (LTE) and the LTE-Advanced (LTE-A) standardizations are predicted to play essential roles in the future fifth-generation (5G) mobile networks. These standardizations require high data rate and high quality of service, which assures low error-rate and low latency. Besides, as discussed in the recent surveys, low complexity communication systems are also essential in the next 5G mobile networks. To adapt to the modern trend of technology, in this PhD thesis, we investigate the multiple-input multiple-output (MIMO) wireless communication schemes. In the first part of this thesis, low-complex forward error correction (FEC) codes are used for low complexity and latency. By considering iterative receivers at the receiver side, we exploit MIMO linear precoding and mapping methods to optimize the error-rate performance of these systems. In the second part of this thesis, non-binary low density parity check (NB-LDPC) codes are investigated. We propose to use MIMO precoders to reduce the complexity for NB-LDPC encoded MIMO systems. A novel low complexity decoding algorithm for NB-LDPC codes is also proposed at the end of this thesis.

Keywords: ***Linear MIMO precoding, mutual information, symbol mapping, turbo-detection, turbo-equalization, iterative receiver, NB-LDPC codes, EXIT charts.***