Efficient Low-complexity Data Detection for Multiple-input Multiple-output Wireless Communication Systems

by

© *Fan JIANG*, BEng, MEng

A thesis submitted to the
School of Graduate Studies
in partial fulfilment of the
requirements for the degree of
Doctor of Philosophy

Faculty of Engineering & Applied Science
Memorial University of Newfoundland

May 28, 2018

St. John's Newfoundland

Contents

Abstract viii

Acknowledgements x

List of Figures xii

List of Tables xvi

List of Abbreviations and Notations xvii

1 Introduction 1

1.1 Background 1

1.2 Low-complexity Detection Schemes for MIMO systems 3

1.2.1 Low-complexity Data Detection for Small-scale MIMO 3

1.2.2 Data Detection Schemes in Massive MIMO Uplink 6

1.3 Identified Problems and Main Contributions 9

1.4 Thesis Organization 12

1.5 Published and Submitted Work 13

2 Soft-input Soft-output MMSE-SQRD based Turbo Equalization for MIMO-OFDM Systems under Imperfect Channel Estimation 18

2.1 System Model 19
2.2 SISO MMSE based Turbo Equalization under Imperfect Channel Estimation 22
2.2.1 Imperfect Channel Estimation Model 22
2.2.2 SISO MMSE Equalization 23
2.3 SISO MMSE-SQRD based Turbo Equalization 24
2.3.1 Development of the Proposed Scheme 24
2.3.2 Sorted QR decomposition 27
2.3.3 Complexity Analysis 28
2.4 Numerical Simulations 28
2.5 Summary 31

3 Low Complexity Turbo Equalizers Conditioned on Channel Estimate for MIMO Systems 33

3.1 System Model 34
3.1.1 Transmitter and Receiver Architecture 34
3.1.2 Turbo Structure 36
3.1.3 Imperfect Channel Estimation 38
3.2 MMSE-Based Turbo Equalization Conditioned on Channel Estimate 39
3.2.1 Development of the Proposed Scheme 39
3.2.2 LLR Computation 42
3.2.3 Discussion 43
3.3 SQRD-Based Turbo Equalization Conditioned on Channel Estimate 45
3.3.1 The Re-development of SQRD-based Turbo Equalization 45
3.3.2 Sorted QR Decomposition Algorithm 48
3.3.3 Schemes Generalization and Discussion 50
3.3.4 Computational Complexity Analysis 52

3.4 Performance Evaluation and Discussions 52
3.5 Summary 59

4 A New SQRD-based Soft Interference Cancelation Scheme in Multi-user MIMO SC-FDMA Systems 61

4.1 System Model 62
4.2 Sorted QR Decomposition Based Soft Interference Cancelation Scheme . . . 64
4.2.1 SISO MMSE Detection 64
4.2.2 QR decomposition Based Soft Interference Cancelation 65
4.2.3 Computational Complexity Analysis 69
4.3 Performance Evaluation and Discussions 70
4.4 Summary 71

5 Stair Matrix and its Applications to Massive MIMO Uplink Data Detection 73

5.1 System Model 75
5.1.1 Linear MMSE Data Detection 76
5.1.2 Neumann Series Expansion 77
5.1.3 Jacobi Method 79
5.2 Stair Matrix and its Applicability to Massive MIMO Systems 80
5.2.1 Stair Matrix and its Properties 81
5.2.2 Using Stair Matrix in Neumann Series Expansion 82
5.2.3 Residual Estimation Error 85
5.3 Implementation of the Stair Matrix in Iterative Method 86
5.3.1 Stair Matrix in Iterative Method 86
5.3.2 Computation of the LLR 87

5.3.3 Computational Complexity Analysis 91
5.4 Numerical Simulations and Performance Evaluation 92
5.4.1 Convergence Conditions 92
5.4.2 Matrix Inverse 94
5.4.3 Residual Estimation Error 95
5.4.4 BER Performance 98
5.5 Summary 105

6 Low Complexity and Fast Processing Algorithms for V2I Massive MIMO Uplink Detection 106

6.1 System Model 109
6.1.1 Linear MMSE Detection 112
6.1.2 Matrix Inversion Based on Cholesky Decomposition 114
6.1.3 Jacobi Method 115
6.1.4 Gauss-Seidel Method 116
6.2 Block Diagonal Matrix Inversion Based Detection Method 117
6.2.1 Development of the Proposed Scheme 117
6.2.2 A New Matrix Inversion Method 119
6.2.3 Initial Estimation 120
6.2.4 LLR Computation 122
6.2.5 Parallel Processing 123
6.2.6 Residual Estimation Error 125
6.2.7 Computational Complexity Analysis 127
6.3 Block Gauss-Seidel Method 130
6.3.1 Development of the Proposed Scheme 131
6.3.2 Parallel Processing 132

6.3.3 Residual Estimation Error . . . 132
6.3.4 Computational Complexity Analysis . . . 133
6.4 Numerical Simulations and Performance Evaluation . . . 133
6.4.1 Convergence Conditions . . . 133
6.4.2 BER Performance . . . 139
6.4.3 Processing Time . . . 143
6.4.4 Discussion . . . 147
6.5 Summary . . . 148

7 Efficient and Fast Processing Large Array Signal Detection in Underwater Acoustic Communications 149

7.1 System Model . . . 151
7.1.1 UWA Channel Model . . . 152
7.1.2 Gauss-Seidel Method . . . 152
7.2 Block Gauss-Seidel Method . . . 154
7.2.1 The Development . . . 154
7.2.2 Parallel Processing . . . 155
7.2.3 Initial Estimation . . . 157
7.2.4 LLR Computation . . . 157
7.3 Performance Evaluation and Discussion . . . 159
7.3.1 Computational Complexity . . . 159
7.3.2 Convergence Performance . . . 161
7.3.3 BER Performance . . . 161
7.4 Summary . . . 162

8 Conclusions and Future Work 163

8.1 Conclusions 163

8.2 Future Work 164

Appendix **166**

Bibliography **184**

Abstract

The tradeoff between the computational complexity and system performance in multiple-input multiple-output (MIMO) wireless communication systems is critical to practical applications. In this dissertation, we investigate efficient low-complexity data detection schemes from conventional small-scale to recent large-scale MIMO systems, with the targeted applications in terrestrial wireless communication systems, vehicular networks, and underwater acoustic communication systems.

In the small-scale MIMO scenario, we study turbo equalization schemes for multiple-input multiple-output orthogonal frequency division multiplexing (MIMO-OFDM) and multiple-input multiple-output single-carrier frequency division multiple access (MIMO SC-FDMA) systems. For the MIMO-OFDM system, we propose a soft-input soft-output sorted QR decomposition (SQRD) based turbo equalization scheme under imperfect channel estimation. We demonstrate the performance enhancement of the proposed scheme over the conventional minimum mean-square error (MMSE) based turbo equalization scheme in terms of system bit error rate (BER) and convergence performance. Furthermore, by jointly considering channel estimation error and the *a priori* information from the channel decoder, we develop low-complexity turbo equalization schemes conditioned on channel estimate for MIMO systems. Our proposed methods generalize the expressions used for MMSE and MMSE-SQRD based turbo equalizers, where the existing methods can be viewed as special cases. In addition, we extend the SQRD-based soft interference cancelation scheme to MIMO SC-FDMA systems where a multi-user MIMO scenario is considered. We show an improved system BER performance of the proposed turbo detection scheme over the conventional MMSE-based detection scheme.

In the large-scale MIMO scenario, we focus on low-complexity detection schemes because computational complexity becomes critical issue for massive MIMO applications. We

first propose an innovative approach of using the stair matrix in the development of massive MIMO detection schemes. We demonstrate the applicability of the stair matrix through the study of the convergence conditions. We then investigate the system performance and demonstrate that the convergence rate and the system BER are much improved over the diagonal matrix based approach with the same system configuration. We further investigate low-complexity and fast processing detection schemes for massive MIMO systems where a block diagonal matrix is utilized in the development. Using a parallel processing structure, the processing time can be much reduced. We investigate the convergence performance through both the probability that the convergence conditions are satisfied and the convergence rate, and evaluate the system performance in terms of computational complexity, system BER, and the overall processing time. Using our proposed approach, we extend the block Gauss-Seidel method to large-scale array signal detection in underwater acoustic (UWA) communications. By utilizing a recently proposed computational efficient statistic UWA channel model, we show that the proposed scheme can effectively approach the system performance of the original Gauss-Seidel method, but with much reduced processing delay.

Acknowledgements

I would like to take this opportunity to acknowledge the great support and help I received during my PhD program.

First of all, I would like to thank Dr. Li, Cheng, my supervisor for his continuous support, encouragement, and great help. As a supervisor, he helps me with identifying research directions, planing research work and career development. Without his support, I could not have participated in the flagship international conferences of Institute of Electrical and Electronics Engineers (IEEE), such as international conferences on communications (ICC), the global communications conference (GLOBECOM). The participation in those well-known international conferences brought me the opportunity to know the latest research work from peers, and learn from them. Besides that, Dr. Li also recommends me to act as technical program committee member in conferences and reviewer for conferences and journals. Those activities not only help develop my research capability in identifying research problems, but also improve academic writing by learning from others. He is always the first one to provide valuable discussion and suggestions whenever I have difficulty in research. I am proud to have him as the supervisor. To me, he is not only the supervisor, but also a respected father and best friend.

Many thanks to the supervisor committee members: Dr. Howard Heys and Dr. Yuanzhu Chen. During the program, especially in the early stage, they worked with me to find research problems. They have provided many valuable suggestions in my proposal and most of them have been taken in my further research work. Dr. Heys also acts as the supervisor in my teaching skills enhancement program (TSEP) and his valuable suggestions and great encouragement have helped me with gaining confidence and teaching skills.

Thanks to Dr. Cecilia Moloney, Dr. Weimin Huang for their dedicate comments and great help in both the course work and comprehensive examination. Many thanks to stu-

dents members in Wireless Networking and Mobile Computing Laboratory (WineMocol), including Ruoyu Su, Yi Zhang, Chen Zhang, Zijun Gong, Somayeh Kafaie, Sipan Ye, and many more. I will always remember the study time spent with you.

Special thanks to Faculty of Engineering and Applied Science, School of Graduate Studies, and Research and Development Corporation (RDC) of Newfoundland and Labrador, for providing financial support. Also, the administrative staff, such as Moya Crocker, Tina Dwyer, Colleen Mahoney, and many more, your dedicated assistance is of great significance to my study at Memorial University. Special thanks to Courtney Hiscock from RDC (now financial officer at department of Tourism, Culture, Industry and Innovation, Newfoundland and Labrador), for the continuous support and dedicate help in ocean industry student research award (OISRA) program.

Last but not least, I would like to acknowledge the great support and help from my family. I insist that the great support and help from them are essential for every achievement I received in the past and will receive in the future.

Thank you, all!

— Fan Jiang

List of Figures

1.1 Illustration of a typical turbo structure 3

2.1 Schematic diagram of the transmitter and receiver in MIMO-OFDM systems 20

2.2 BER performance of the proposed scheme compared to SISO MMSE turbo equalization, QPSK 29

2.3 BER performance of the proposed scheme compared to SISO MMSE turbo equalization, 16QAM 29

2.4 BER performance of the proposed scheme compared to SISO MMSE turbo equalization, 64QAM 30

3.1 Schematic diagram of the transmitter and receiver architecture in MIMO systems 35

3.2 BER performance of the MMSE-based turbo equalization schemes, 1/2 Turbo code, QPSK, 4×4 MIMO: $\lambda = 0$ and $\lambda = 0.1$. 53

3.3 BER performance of the MMSE-based turbo equalization schemes, 1/2 Turbo code, QPSK, 4×4 MIMO: $\lambda = 0.2$ and $\lambda = 0.5$. 53

3.4 BER performance of the MMSE-SQRD based turbo equalization schemes, 1/2 Turbo code, QPSK, 4×4 MIMO: $\lambda = 0$ and $\lambda = 0.1$. 55

3.5 BER performance of the MMSE-SQRD based turbo equalization schemes, 1/2 Turbo code, QPSK, 4×4 MIMO: $\lambda = 0.2$ and $\lambda = 0.5$. 56
3.6 BER performance comparison, 1/2 Turbo code, 16QAM, 4×4 MIMO: $\lambda = 0.1$ 57
3.7 BER performance comparison, 1/2 Turbo code, 16QAM, 4×4 MIMO: $\lambda = 0.5$ 58

4.1 Schematic diagram of the transmitter and receiver architecture in MIMO SC-FDMA systems 63
4.2 BER performance comparison, QPSK, 2/3 Turbo code 68
4.3 BER performance comparison, 0 iteration 69
4.4 BER performance comparison, 2 iteration 70

5.1 The normalized mean-square error for the approximation in Equation (5.33), $N_U = 25$ 89
5.2 Cumulative distribution function of the maximum eigenvalue $N_U = 25$. 92
5.3 Normalized mean-square error for the matrix inverse approximation 94
5.4 Residual Estimation Error: (a) $N_B = 150$, $N_U = 25$, average SNR= 5dB; (b) $N_B = 200$, $N_U = 25$, average SNR= 3.5dB 96
5.5 BER performance: (a) $N_B = 150$, $N_U = 25$; (b) $N_B = 250$, $N_U = 25$. LDPC code: code length 64800, code rate 0.5. 64QAM modulation. 99
5.6 BER performance: (a) $N_B = 125$, $N_U = 25$; (b) $N_B = 150$, $N_U = 25$; (c) $N_B = 175$, $N_U = 25$. LDPC code: code length 64800, code rate 0.5. 64QAM modulation. 101
5.7 BER performance: $N_B = 200$, $N_U = 25$. LDPC code: code length 64800, code rate 0.5. 64QAM modulation. 102
5.8 BER performance: (a) $N_B = 150$, $N_U = 25$; (b) $N_B = 400$, $N_U = 25$. LDPC code: code length 1152, code rate 0.5. 64QAM modulation. 103

6.1 RSU with massive antennas to serve multiple vehicles on the road 110
6.2 The baseband signal processing diagram at OBU and RSU 110
6.3 Parallel processing on each block 124
6.4 Computational complexity comparison for different block size 129
6.5 Cumulative distribution function of the maximum eigenvalues, block diagonal matrix based detection scheme, $N_U = 32$: (a) $N_B = 112$, (b) $N_B = 128$. 134
6.6 Cumulative distribution function of the maximum eigenvalues, block diagonal matrix based detection scheme, $N_U = 32$: (a) $N_B = 144$, (b) $N_B = 160$. 135
6.7 Cumulative distribution function of the maximum eigenvalues, block Gauss-Seidel method. 138
6.8 BER Performance, $N_B = 256$, $N_U = 32$. BMI: block diagonal matrix based detection scheme; BGS: block Gauss-Seidel method. (a) $B = 1$; (b) $B = 2$. . . 140
6.9 BER Performance, $N_B = 256$, $N_U = 32$. BMI: block diagonal matrix based detection scheme; BGS: block Gauss-Seidel method. (a) $B = 4$; (b) $B = 8$. . . 141
6.10 BER Performance, $N_B = 256$, $N_U = 32$. BMI: block diagonal matrix based detection scheme; BGS: block Gauss-Seidel method. (a) $B = 16$, (b) $B = 32$. 142
6.11 BER Performance versus N_B, $N_U = 32$, BMI: block diagonal matrix based detection scheme; BGS: block Gauss-Seidel method. (a) $B = 1$; (b) $B = 2$. . . 144
6.12 BER Performance versus N_B, $N_U = 32$, BMI: block diagonal matrix based detection scheme; BGS: block Gauss-Seidel method. (a) $B = 4$; (b) $B = 8$. . . 145
6.13 Processing time of the proposed detection schemes using parallel processing, BMI: block diagonal matrix based detection scheme; BGS: block Gauss-Seidel method. 146

7.1 An application scenario of UWA communications with large-scale hydrophones at platform. 151

7.2 Block diagram of the parallel processing 156

7.3 Cumulative distribution function of the maximum eigenvalue 159

7.4 BER Performance, LDPC code rate 0.5, code length 1152, 16QAM 162

List of Tables

3.1 **Algorithm 3.1:** SQRD-based Turbo Equalization 49

3.2 **Algorithm 3.2:** Sorted QR Decomposition 51

5.1 **Algorithm 5.1:** Compute the Inverse of a Stair Matrix 83

5.2 **Algorithm 5.2:** Proposed Iterative Method Using Stair Matrix 90

6.1 **Algorithm 6.1:** Compute the Inversion of a Hermitian Matrix 121

6.2 **Algorithm 6.2:** Block Diagonal Matrix Based Massive MIMO Detection . 126

7.1 **Algorithm 7.1:** Proposed Block Gauss-Seidel Method on Each Subcarrier . 160

A.1 Non-zero Terms in Equation (A.19) 171

A.2 Non-zero Terms in Equation (A.22) 173

List of Abbreviations and Notations

16QAM	16-quadrature amplitude modulation
3G	third generation
4G	fourth generation
5G	fifth generation
64QAM	64-quadrature amplitude modulation
AWGN	additive white Gaussian noise
BER	bit error rate
BICM	bit-interleaved coded modulation
BP	belief propagation
CP	cyclic prefix
DFT	discrete Fourier transform
EVA	extended vehicular A
FDE	frequency domain equalization
FFT	fast fourier transform
IBI	inter-block interference
IDFT	inverse-DFT
IFFT	inverse-FFT
ISI	inter-symbol interference
ITS	intelligent transportation systems
LAS	likelihood ascent search
LDPC	low density parity check
LLR	log likelihood ratio

LTE	long-term evolution
LTE-A	LTE-advanced
MAP	maximum *a posteriori*
MCS	modulation and coding scheme
MCMC	Markov chain Monte-Carlo
MIMO	multiple-input multiple-output
ML	maximum likelihood
MMSE	minimum mean-square error
MSE	mean-square error
MSWF	multi-stage Winner filter
NP	non-deterministic polynomial-time
NPI	noise-plus-interference
UWA	underwater acoustic
OBU	on-board unit
OFDM	orthogonal frequency division multiplexing
PAPR	peak-to-average power ratio
PDF	probability density function
PIC	parallel interference cancelation
QAM	quadrature amplitude modulation
QPSK	quadrature phase shift keying
RHS	right hand side
RSU	roadside unit
SC-FDMA	single-carrier frequency division multiple access
SIC	successive interference cancelation
SINR	signal-to-interference-plus-noise ratio

SISO	soft-input soft-output
SNR	signal-to-noise ratio
SQRD	sorted QR decomposition
V2I	vehicle-to-infrastructure
V2V	vehicle-to-vehicle
ZMCGCS	zero-mean complex Gaussian circularly symmetric
i.i.d.	independent and identically distributed
$(\cdot)^{\mathrm{T}}$	matrix transpose
$(\cdot)^{\mathrm{H}}$	matrix conjugate transpose
$(\cdot)^{-1}$	matrix inversion
$\lfloor \cdot \rfloor$	the floor function
$\mathrm{E}\{\cdot\}$	mathematical expectation
$\mathrm{cov}(\cdot,\cdot)$	mathematical covariance
$\exp(\cdot)$	exponential function
$\ln(\cdot)$	natural logarithmic function
$\mathbf{I}_L$	the size L identity matrix
$\mathbf{e}_l$	the l-th column of $\mathbf{I}_L$
$\mathbf{F}_N$	the normalized N-point DFT matrix
$\mathbf{F}_N^{\mathrm{H}}$	the normalized N-point IDFT matrix
$\mathrm{diag}\{\mathbf{A}\}$	a column vector with diagonal elements in $\mathbf{A}$
$\mathrm{diag}(\mathbf{v})$	a diagonal matrix with diagonal elements in $\mathbf{v}$

Chapter 1

Introduction

1.1 Background

Wireless communication systems play significant roles in people's lives nowadays. For past several decades, the mobile access technology has experienced a revolutionary change about every ten years [1]. It is foreseeable that by the year of 2020, the fifth generation (5G) mobile communication systems will be commercially deployed. Yet for 5G, the key requirements, the development of the standard, and the field trials, are still undergoing; increasing interests and focus have been put on this research in both academia and industry [2–4]. Meanwhile, with the evolution of 5G technologies, modern vehicular networks, aiming to providing high level user experience on the move, have attracted much attention [5–10]. In particular, the intelligent transportation systems (ITS) have been proposed to improve driving safety and comfort, as well as offering in-vehicle entertainment such as social connections, video streaming, and related services [5–7, 10–12]. It is predicted that 90 percent of vehicles will be connected via wireless links by 2020 [10]. Wireless communication techniques have been also applied to discover, investigate, and exploit oceanic resources [13]. Typically, underwater acoustic (UWA) communications have been widely studied [14–19]. While the

harsh UWA channels are agreed to be rapid time-varying and environmentally dependent, the wireless techniques developed in radio communication systems have been demonstrated to be applicable and efficient to establish reliable UWA communications [15, 16].

As one of the typical wireless communication techniques, multiple-input multiple-output (MIMO) has been successfully applied to the third and currently forth generation (3G/4G) mobile communication systems [1, 20–23], and many other application scenarios [15–17, 24, 25]. It was in the 1990s that MIMO has shown great potential to achieve high data rate low error rate wireless communications [26–28]. Generally, MIMO systems can bring spacial multiplexing and diversity gain over single antenna systems. The multiplexing gain enables multiple data stream transmitted over the same time, frequency, and code resources, resulting in high rate communication; while the diversity gain takes advantages of multi-path fading to improve error rate performance. Recently, large-scale MIMO, also known as massive MIMO, has attracted increasing research interests [2–4, 18, 19, 29–37]. Due to the huge potential multiplexing and diversity gain over the conventional small-scale MIMO and single-antenna systems, massive MIMO can increase the system spectrum and energy efficiency in orders of magnitude [2–4, 18, 19, 29–32, 35–37]. However, it is known that the optimum MIMO detection problem is non-deterministic polynomial-time hard (NP-hard) [38]. That is to say, the computational complexity for MIMO signal processing is considerable when using the optimal MIMO detection methods. For practical applications, low-complexity, yet high-performance suboptimal MIMO detection schemes are preferred, hence becoming the main research problems in this project.

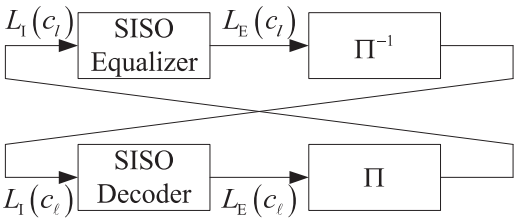


Figure 1.1: Illustration of a typical turbo structure

1.2 Low-complexity Detection Schemes for MIMO systems

1.2.1 Low-complexity Data Detection for Small-scale MIMO

In conventional small scale MIMO systems, optimal detection schemes, such as the maximum likelihood (ML) and the maximum *a posteriori* (MAP) receivers, are prohibited as their complexity increases exponentially with the number of transmit antennas (or user numbers) and the modulation symbols [38–43]. Therefore, when high order modulations, for example 16 quadrature amplitude modulation (16QAM) and 64QAM, are adopted in the system, the complexity is costly for implementations. In real applications, turbo receivers are normally adopted to achieve a good tradeoff between the system performance and the computational complexity [21, 22, 37–51]. A turbo structure consists of two soft-input soft-output (SISO) modules while the *extrinsic* information is exchanged between them, which is illustrated in Figure 1.1 [51]. Specifically, the two SISO modules are usually the SISO equalizer and the SISO channel decoder, and the *extrinsic* information is defined as the mutual information of the bits log likelihood ratio (LLR) and the coded bit stream. The SISO equalizer utilizes the *a priori* information from the channel decoder to perform the equalization process, producing *extrinsic* LLRs to the SISO channel decoder for the decoding process. Meanwhile, with the *extrinsic* LLRs of the coded bit stream, the SISO channel decoder performs the decoding process and outputs the LLRs for information and coded bit stream. The previous one is used to produce an information bit estimation, and

the latter one is feedback to the SISO equalizer. With sufficient iterations, the progress between the two SISO modules comes to convergence, achieving near-optimal system performance [40, 43, 49, 50, 52–54].

It has also been demonstrated that with iterative signal processing, a simple equalizer can eventually approach the system bit error rate (BER) performance of the optimal equalizers [49, 52–57]. Therefore, most studies have focused on the design of low complexity and fast convergence turbo equalizers. In [55], the authors have investigated the MMSE-based turbo equalization for a multi-path time-invariant channel. In their work, the *a priori* information from the channel decoder is utilized to help with symbol detection. Successful applications of the MMSE-based turbo equalization schemes have been applied to orthogonal frequency division multiplexing (OFDM) and single-carrier frequency division multiple access (SC-FDMA) systems [21, 41, 43, 58–60]. Turbo equalization schemes for single-carrier frequency domain equalization (SC-FDE) systems are studied in [40, 46, 47], and it has been reported that the low-complexity turbo equalization schemes in the frequency domain achieve high robustness compared to the time domain equalization schemes. A class of sorted QR decomposition (SQRD) based turbo equalization schemes has been studied in [22, 39, 41, 60, 61]. By using the QR decomposition algorithms, an upper triangular matrix is obtained; hence, the original estimation solutions can be obtained successively [62]. MMSE SQRD based turbo equalization schemes employ the *a priori* information from channel decoder, and construct the extended channel matrix consisting of the channel matrix and the *a priori* covariance matrix. By performing the sorted QR decomposition, the MMSE-based turbo equalization can be transformed from a parallel interference cancelation into a successive interference cancelation. Therefore, a back-forward solution can be used to complete the symbol estimation. Since the SQRD-based detection schemes avoid the matrix inversion issue in MMSE-based detection schemes, the overall computational

complexity is comparatively reduced [22, 39, 41, 60].

In practical applications, the channel matrix used for equalization is obtained through channel estimation [63, 64]. However, channel estimation error is inevitable due to the following reasons. First, a channel estimate is usually obtained based on pilot symbols or training sequence, which is not exactly the same as the channel used for data transmission. The difference is even more severe when a fast time-varying channel is studied. Second, as noise and interference are presented, there is channel estimation error in whatever channel estimation method. Channel estimation error can always cause the system performance degradation [21,39,65–71]. Therefore, accounting for channel estimation error in the design of equalization schemes is significant and necessary [21, 39, 67–71]. In [21, 39], the channel estimation error is assumed to be independent from the channel estimation. Based on that assumption, the MMSE-based and SQRD-based turbo equalization schemes are developed. Recently, equalization schemes conditioned on channel estimate have shown much enhanced system performance [67–71]. Specifically, the linear zero-forcing (ZF) equalization scheme is studied in [67], and linear MMSE equalization schemes are studied in [68,70]. In [71], the authors derived the MMSE equalization conditioned on both the channel estimate and the transmitted symbols. The results of their work show that slight improvement is achieved compared to the equalization scheme in [68, 70]. In prior work [69], the turbo equalization conditioned on channel estimate based on MMSE criteria is investigated. The results show that compared to conventional MMSE-based turbo equalization schemes where the channel estimate is directly applied to the equalization process, the turbo equalization scheme conditioned on channel estimate has much better system performance, especially when the channel estimation error is severe.

In general, in conventional small-scale MIMO systems, low-complexity detection schemes are expected to consider the *a priori* information from channel decoder and the channel esti-

mation error in the development. Prevailing MMSE-based or SQRD-based turbo detection schemes are shown to be low-complexity while achieving near-optimal system performance.

1.2.2 Data Detection Schemes in Massive MIMO Uplink

We then move the focus to large-scale MIMO (or massive MIMO) systems. Different from the conventional small-scale MIMO systems, in massive MIMO, the antenna elements at the base station can reach up to hundreds while the users under service can be tens [2, 3]. By extending the MIMO systems from a small scale to a large one, an overwhelming diversity and multiplexing gain can be achieved, leading to orders of magnitude in the improvement of the spectrum and energy efficiency [2, 3, 29–31]. The idea of using massive MIMO has been also introduced to vehicular networks [35, 36] and UWA communications [18, 19]. However, in massive MIMO systems, the optimal receivers, such as ML and MAP receivers, are prohibited due to the exponentially increased computational complexity.

To achieve good tradeoff between the system performance and the computational complexity, much effort has been made in developing low-complexity detection schemes to achieve near-optimal performance [23, 72–86]. For example, in [72, 73], the authors extend a class of likelihood ascent search (LAS) based multi-user detection schemes to large-scale MIMO. The Monte-Carlo sampling based massive MIMO detection schemes are studied in [74, 75], where the Markov chain Monte-Carlo (MCMC) strategy is introduced. A class of belief propagation (BP) based detection schemes for massive MIMO have been studied in [76, 77], where the idea of message passing through factor graph is introduced for the development. Recently, adaptive reduced rank interference suppression based detection schemes are developed for large-scale MIMO uplink detection [78, 79]. The idea behind these schemes is to project a large dimensional signal vector to a small dimensional one, and then multi-stage Winner filtering (MSWF) is adopted to perform optimization of the de-

sired signal [88,89]. It has been also reported that the complexity required for each symbol detection can be reduced to $O(N_B \times N_U)$. (N_B and N_U denotes the number of antennas at base station and the single-antenna user terminals, respectively.) Among these low complexity detection schemes for massive MIMO uplink, a class of near linear MMSE/ZF detection schemes are studied in [23, 80–87]. Generally, all those schemes attempt to avoid matrix inversion operations in linear MMSE/ZF detection schemes, and the strategies can be summarized in two categories: the first one is to approach matrix inversion, and the other is to solve linear equations with iterative methods.

The first category is to approximate the matrix inversion [23, 80, 81]. For example, in [23], the authors attempt to introduce Neumann series expansion to avoid the matrix inversion in linear MMSE detection. It has been shown that when the number of antennas at base station is much greater than the number of user equipment, the orders required for Neumann series expansion can be as few as 3 (for example, $r = N_B/N_U \geqslant 16$). In [80], the probability of the convergence condition that using the diagonal matrix in Neumann series expansion based detection scheme has been comprehensively discussed. However, Neumann series expansion suffers from matrix multiplications, and the computational complexity is comparable to the matrix inversion algorithm when the expansion order is more than two. In order to speed up the convergence rate, a diagonal banded Newton iteration based matrix inversion approach is studied in [81], where the Newton iteration structure is used. Actually, the results after P iterations in Newton iteration can be seen as the Neumann series expansion of the order $2^P - 1$ [81]. Inevitably, matrix multiplications are involved in diagonal banded Newton iteration based matrix inversion approach, and the iterations are limited to 2 for computational complexity considerations. In summary, the methods that are to approximate matrix inversion suffer from high computational complexity due to the matrix multiplications and the slow convergence rate when the ratio $r = N_B/N_U$ is not

sufficiently large.

The second category is to solve linear equations with iterative methods [82–87]. The basic idea of these methods is to transform the matrix inversion problem into solving linear equations. To solve the linear equations, an initial estimation is provided. Then following an iterative structure to converge, the final output is provided as the solutions to linear equations. For example, in [82], the Jacobi method is adopted, and by following the Jacobi iterative structure, the estimation eventually approaches the MMSE estimation. The Richardson iteration in massive MIMO uplink data detection has been studied in [83], and the authors have demonstrated that the iterative structure can converge even with zero initialization. However, as pointed out in [84,85], the convergence rate for both Jacobi method and Richardson iteration is slow, hence quite a few iterations are required for convergence. The application of the Gauss-Seidel method to massive MIMO uplink data detection is studied in [84], and the convergence performance can be greatly improved. By providing an initial estimation that is close to the MMSE estimation, the joint steepest-decent and Jacobi method based data detection is proposed in [85], and the iterations are greatly reduced. In [86], the authors formulate the MMSE estimation as a minimization problem, and use the conjugate gradient to calibrate the next estimation. However, conjugate gradient-based data detection scheme involves many division operations, which is also computational costly. Compared to the first category which is to approximate matrix inversion, solving linear equations with iterative methods is of less complexity due to the replacement of matrix multiplications with matrix-vector products. However, as summarized in [85], the overall computational complexity of the iterative methods, including the computations in both the initialization and iteration, is still high. It is worth pointing out that the convergence rate of the existing iterative methods can be speeded up by using preconditioning [90]. A potential direction to further reduce the computational complexity could be to find an iterative

method that requires less computation in initialization and fewer iterations for convergence.

1.3 Identified Problems and Main Contributions

For the MIMO-OFDM system, MMSE-SQRD based turbo equalization has been studied in [41]. However, the inevitable channel estimation error is not considered in their scheme. On the other side, the MMSE-based turbo equalization for SC-FDMA system with channel estimation error is studied in [21], where parallel interference cancelation (PIC) is utilized. The turbo equalization, taking into account of channel estimation error and using successive interference cancelation (SIC), has not been studied. Therefore, we first investigate the MMSE-SQRD based turbo equalization for MIMO-OFDM systems under imperfect channel estimation. The proposed scheme considers both the channel estimation error and the *a priori* information from channel decoder, and introduces a successive interference cancelation in data detection.

The design of data detection schemes conditioned on channel estimate has shown obvious system performance improvement in [68, 70, 71]. However, it is worth noting if the equalizer is designed to take into account of both the channel estimation error and the *a priori* information from channel decoder, the system performance can be further improved. Our second contribution is to develop low complexity turbo equalizers conditioned on channel estimation for MIMO systems. We also show the new proposed schemes can be viewed as a general expression for MMSE and SQRD based turbo equalization schemes, where existing schemes are special cases.

As OFDM systems experience high peak-to-average power ratio (PAPR), which requires large dynamic range of the amplifier, they are inapplicable for uplink transmission. To take fully advantages of OFDM systems, i.e., OFDM modulation module at transmitter side and low-complexity frequency domain equalizer at receiver side, the SC-FDMA tech-

nique is adopted as a standard uplink transmission technique at user terminals. Employing the SQRD-based turbo equalization for SC-FDMA systems in [22] has revealed potential advantages over the conventional MMSE-based turbo equalization. However, the scheme in [22] has the following limitations. First, the *a priori* information is not considered in the SQRD procedure where the user detection order is determined. Second, it only performs soft interference cancelation on detected users, while the residual interference from the undetected users remains in current user detection. Last, the expected signal component in the identified interference component is not sufficiently extracted for the current user detection. Therefore, we have proposed a new SQRD based turbo equalization scheme for multi-user SC-FDMA systems, where we have demonstrated the system BER performance improvement over existing schemes.

Extending the MIMO systems from a small scale to a large one, numerous benefits can be brought into the system. However, we note that most data detection schemes for massive MIMO uplink in existing literatures mainly utilize the diagonal matrix in the development. In [23,80], the applicability of using diagonal matrix to massive MIMO uplink data detection has been demonstrated. Nevertheless, we find some limitations for using diagonal matrix. First of all, in the massive MIMO system configuration with small ratio of $r = N_B/N_U$ (N_B and N_U denotes the number of antennas at base station and the single-antenna user terminals, respectively.), the convergence rate using the diagonal matrix is slow. Alternatively, a few iterations (or orders in Neumann series expansion) are required to provide near-optimal system performance. Besides, the convergence conditions, which are critical to data detection schemes, are met with a low probability when r is small. In other words, in some cases, the diagonal matrix may not achieve convergence. To alleviate those limitations, we propose to develop an iterative method to achieve performance near linear MMSE detection scheme using the stair matrix [91,92]. We have shown the new low-complexity iterative method can

approach the linear MMSE detection performance in massive MIMO systems, and achieve enhanced system performance in terms of better convergence performance, lower residual estimation error, and much improved BER performance, over the iterative method with the diagonal matrix in the development.

The iterative methods for massive MIMO detection have reduced the computational complexity to the order of $O\left(N_U^2\right)$. However, when considering a massive MIMO vehicle-to-infrastructure (V2I) scenario, as more and more vehicles are on the road, the computations are still considerable. Besides, the updating progress requires long processing delay for one iteration. For example, in the Gauss-Seidel method, the last user needs to wait until all previous uses have been updated. This successive detection manner makes the detection scheme not time efficient. Moreover, the successive detection manner is not efficient for hardware implementation [85]. To mitigate those defects, we propose to use the block diagonal matrix based iterative method. In this proposal, we have successfully transformed the large-scale matrix inversion issue into a set of small-scale inversions, and a new matrix inversion method is proposed to further reduce the computational complexity. The proposed iterative method utilizes the parallel processing structure to speedup the overall processing progress. To further reduce the overall computational complexity, the block Gauss-Seidel method is proposed for the updating progress. The new proposals have shown great improvement in terms of low complexity and fast processing.

It is known that the reliable wireless communications through harsh underwater environment is challenging [13–16]. By employing conventional small scale MIMO and OFDM techniques, we know that link reliability can be greatly improved. When the array of hydrophones increases, it has been recently shown that the system spectrum and energy efficiency, signal-to-interference-plus-noise (SINR) can be greatly improved [18, 19, 93]. However, the complexity of large array signal processing will be an emergent issue to be

addressed. We have applied the block Gauss-Seidel method to the signal processing in this application scenario and the results are promising. Specifically, we utilize the computationally efficient statistic UWA channel model developed in [14] for simulation. We then show that computational complexity of the new detection scheme is low compared to linear MMSE detection scheme. After that, we show that our proposal achieves the system BER performance close to original Gauss-Seidel method based detection scheme, and the overall processing delay is much reduced.

1.4 Thesis Organization

The main work of this thesis is divided into two parts. In the first part, consisting of Chapter 2 to Chapter 4, we investigate data detection schemes for conventional small-scale MIMO, where low-complexity efficient turbo equalization schemes are studied and evaluated. The second part, consisting of Chapter 5 to Chapter 7, focuses on low-complexity data detection schemes for massive MIMO uplink. The specific topic in each chapter is detailed as follows:

- **Chapter 2:** Soft-input soft-output MMSE-SQRD based turbo equalization for MIMO-OFDM systems under imperfect channel estimation;
- **Chapter 3:** Low complexity turbo equalizers conditioned on channel estimate for MIMO systems;
- **Chapter 4:** A new SQRD-based soft interference cancelation scheme in multi-user MIMO SC-FDMA system;
- **Chapter 5:** Stair matrix and its applications to massive MIMO uplink data detection;
- **Chapter 6:** Low-complexity and fast processing algorithms for V2I massive MIMO uplink

detection;

- **Chapter 7:** Efficient and fast processing large array signal detection in underwater acoustic communications.

Finally, we summarize our work and discuss the future work in **Chapter 8**.

1.5 Published and Submitted Work

- **Journal Articles**

  [**J1**] **F. Jiang**, C. Li, Z. Gong, and R. Su, "Low Complexity Turbo Equalizers Conditioned on Channel Estimate: From Small-scale to Large-scale MIMO Systems," submitted to *IEEE Transactions on Wireless Communications*, under review, 2018.

  [**J2**] **F. Jiang**, C. Li, Z. Gong, and R. Su, "A New Iterative Method Using Stair Matrix For Massive MIMO Uplink Signal Detection," submitted to *IEEE Wireless Communications Letters*, under second round review, 2018.

  [**J3**] R. Su, Z. Gong, C. Li, **F. Jiang**, and D. Zhang, "AUV-aided Underwater Acoustic Sensor Networks: Challenges and Opportunities," submitted to *IEEE Network Magazine*, under review, 2018.

  [**J4**] Z. Gong, C. Li, and **F. Jiang**, "An Efficient Spatial Filter Based Channel Estimation Method for Massive MIMO Systems in Low SNR Regime," submitted to *IEEE Transactions on Wireless Communications*, under review, 2018.

  [**J5**] **F. Jiang**, C. Li, and Z. Gong, "Accurate Analytical BER Performance for ZF Receivers under Imperfect Channel in Large Receiving Antennas and Low SNR Region," *IEEE Signal Processing Letters*, accepted, to appear, 2018.

[**J6**] **F. Jiang**, C. Li, and Z. Gong, "Low Complexity and Fast Processing Algorithms for V2I Massive MIMO Uplink Detection," *IEEE Transactions on Vehicular Technology*, to appear, DOI: 10.1109/TVT.2018.2808237, 2018.

[**J7**] Z. Gong, C. Li, and **F. Jiang**, "AUV-Aided Joint Localization and Time Synchronization for Underwater Acoustic Sensor Networks," *IEEE Signal Processing Letters*, DOI: 10.1109/LSP.2018.2799699, to appear, 2018.

[**J8**] **F. Jiang**, C. Li, Z. Gong, and R. Su, "Stair Matrix and its Applications to Massive MIMO Uplink Data Detection," *IEEE Transactions on Communications*, DOI:10.1109/TCOMM.2017.2789211, to appear, 2018.

[**J9**] Z. Gong, C. Li, **F. Jiang**, R. Su, R. Venkatesan, C. Meng, S. Han, and Y. Zhang, "Design, Analysis, and Field Testing of an Innovative Drone-assisted Zero-Configuration Localization Framework for Wireless Sensor Networks," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 11, pp. 10322-10335, 2017.

[**J10**] Z. Gong, C. Li, and **F. Jiang**, "Pilot Contamination Mitigation Strategies in Massive MIMO Systems," *IET Communications*, vol. 11, no. 16, pp. 2403-2409, 2017.

[**J11**] **F. Jiang**, C. Li, C. Meng, and Z. Gong, "A New Turbo Equalizer Conditioned on Estimated Channel for MIMO MMSE Receiver," *IEEE Communications Letters*, vol. 21, no. 4, pp. 957–960, April 2017.

[**J12**] **F. Jiang**, Y. Zhang, and C. Li, "A New SQRD-Based Soft Interference Cancellation Scheme in Multi-User MIMO SC-FDMA System," *IEEE Communications Letters*, vol. 21, no. 4, pp. 821–824, April 2017.

[**J13**] **F. Jiang**, Y. Zhang, C. Meng, and C. Li"GF(q)-based Precoding: Information Theoretical Analysis and Performance Evaluation," *Wireless Communications and Mobile Computing*, vol. 16, no. 17, pp. 3032-3044, December 2016.

[**J14**] Y. Xue, **F. Jiang**, B. Jiang, and X. Gao, "Efficient Link Adaptive Transmission Scheme with Turbo-IC Receivers," *Journal of Southeast University (Natural Science Edition)*, vol. 44, no. 1, pp. 1-6, January 2014.

[**J15**] **F. Jiang**, C. Moloney, and C. Li, "Investigation of Multi-User Interference Cancellation Techniques at Roadside Unit in Vehicular Networks," submitted to *Ad Hoc & Sensor Wireless Networks: An International Journal,* **invited paper**, under review.

- **Conference Articles**

[**C1**] **F. Jiang**, C. Li, and Z. Gong, "Efficient and Fast Processing Large Array Signal Detection in Underwater Acoustic Communications," submitted to *IEEE Global Communications Conference* (GLOBECOM'18), Abu Dhabi, UAE, 2018, under review.

[**C2**] Z. Gong, C. Li, and **F. Jiang**, "Pilot Decontamination for Cell-Edge Users in Multi-Cell Massive MIMO Based on Spatial Filter," *IEEE International Conference on Communications (ICC'18)*, Kansas City, MO, USA, 2018, accepted.

[**C3**] **F. Jiang**, C. Li, and Z. Gong, "Block Gauss-Seidel Method Based Detection in Vehicle-to-Infrastructure Massive MIMO Uplink," in *Proceedings of the IEEE Global Communications Conference (GLOBECOM'17)*, Singapore, December 2017. **Best Paper Award**.

[**C4**] **F. Jiang**, C. Li, and Z. Gong, and Y. Zhang, "Massive MIMO for Future Vehicular Networks: Compressed-sensing and Low-complexity Detection Schemes," in *Proceedings of the 10th International EAI International Wireless Internet Conference (WiCON)*, Tianjin, China, December 2017, **Invited Paper**.

[**C5**] **F. Jiang**, C. Li, and Z. Gong, "A Low Complexity Soft-output Data Detection

Scheme Based on Jacobi Method for Massive MIMO Uplink Transmission," in *Proceedings of the IEEE International Conference on Communications (ICC'17)*, Paris, France, May 2017.

[**C6**] **F. Jiang** and C. Li, "EXIT: Turbo Equalizer Performance Analysis," in *Proceedings of the 25th IEEE Newfoundland Electrical and Computer Engineering Conference (NECEC'16)*, St. Johns, Canada, November 2016.

[**C7**] X. Pang, Z. Gong, **F. Jiang**, and C. Li, "A DJI Drone Assisted Positioning System for Wireless Sensor Networks (WSNs)," in *Proceedings of the 25th IEEE Newfoundland Electrical and Computer Engineering Conference (NECEC'16)*, St. Johns, Canada, November 2016.

[**C8**] **F. Jiang** and C. Li, "Soft Input Soft Output MMSE-SQRD Based Turbo Equalization for MIMO-OFDM Systems under Imperfect Channel Estimation," in *Proceedings of the IEEE Global Communications Conference (GLOBECOM'15)*, San Diego, CA, USA, December 2015.

[**C9**] **F. Jiang** and C. Li, "A Novel Sorted QR Decomposition Based Multi-user Detection in MIMO SCFDMA Systems," in *Proceedings of the 24th IEEE Newfoundland Electrical and Computer Engineering Conference (NECEC'15)*, St. Johns, Canada, November 2015.

[**C10**] **F. Jiang**, C. Li, and R. Venkatesan, "GF(q) Precoding: Mutual Information Analysis in AWGN Channels," in *Proceedings of International Wireless Communications and Mobile Computing Conference (IWCMC'15)*, Dubrovnik, Croatia, August 2015.

[**C11**] **F. Jiang**, C. Moloney, and C. Li, "A Preliminary Investigation of Multi-User Interference Cancellation Techniques at Roadside Unit in Vehicular Networks," in *Proceedings of the 10th International Conference on Mobile Ad-hoc and Sensor*

Networks (IEEE MSN2014), Maui, Hawaii, USA, December 2014.

[C12] **F. Jiang** and C. Li, "Mutual Information Analysis for GF(2) Precoding," in *Proceedings of 23rd IEEE Newfoundland Electrical and Computer Engineering Conference (NECEC'14)*, St. Johns, NL, Canada, November 2014.

[C13] **F. Jiang**, C. Moloney, and C. Li, "An Investigation of Multi-User Interference Cancellation at Roadside Unit in Vehicular Networks," in *Proceedings of 22nd IEEE Newfoundland Electrical and Computer Engineering Conference (NECEC'13)*, St. Johns, NL, Canada, November 2013.

Chapter 2

Soft-input Soft-output MMSE-SQRD based Turbo Equalization for MIMO-OFDM Systems under Imperfect Channel Estimation

[1] In this chapter, a turbo equalization scheme for MIMO-OFDM systems under imperfect channel estimation based on SISO MMSE-SQRD is proposed. Turbo equalization schemes are preferable in practical communication systems due to their good performance and acceptable computational complexity. The MMSE-SQRD based SISO detection scheme is derived from SISO MMSE detection, and SIC is performed using *a posteriori* information

[1]The related work has been published in *the proceedings of Global Communications Conference* (GLOBECOM'15), San Diego, CA, USA, December 2015.

obtained from previous detected symbols. Compared to SISO MMSE detection, MMSE-SQRD based SISO detection is of low complexity but has significant bit error rate (BER) performance enhancement. When channel estimation error is present, it has been pointed out that the system performance will degrade. We investigate this practical issue in this Chapter, and propose the SISO MMSE-SQRD based turbo equalization under imperfect channel estimation. We first model the channel estimation error as added random Gaussian noise over the channel estimation matrix. Based on that, we re-derive the SISO MMSE detection for the data symbols. After that, we redefine the extended channel matrix and receive vector by taking into account channel estimation error. The SQRD algorithm is adjusted in accordance and MMSE-SQRD based data detection algorithm is finally performed. Numerical simulation results show that the proposed SISO MMSE-SQRD based turbo equalization for MIMO-OFDM systems under imperfect channel estimation outperforms the conventional MMSE-based SISO detection in terms of system BER performance and computational complexity.

The rest of this chapter is organized as follows. In Section 2.1, we present the system model including the transmitter and receiver structure in MIMO-OFDM systems. Following that, we review the SISO MMSE-based turbo equalization scheme with channel estimation error in Section 2.2. The detailed derivation of the proposed SISO MMSE-SQRD turbo equalization scheme is presented in Section 2.3. After that, we present the numerical simulation results in Section 2.4. Finally, we summarize this chapter in Section 2.5.

2.1 System Model

A turbo coded MIMO-OFDM system equipped with N_T transmit antennas and N_R receive antennas is considered. The schematic diagram of the transmitter and receiver is shown in Figure 2.1. At the transmitter side, the information bits are first fed into a turbo encoder,

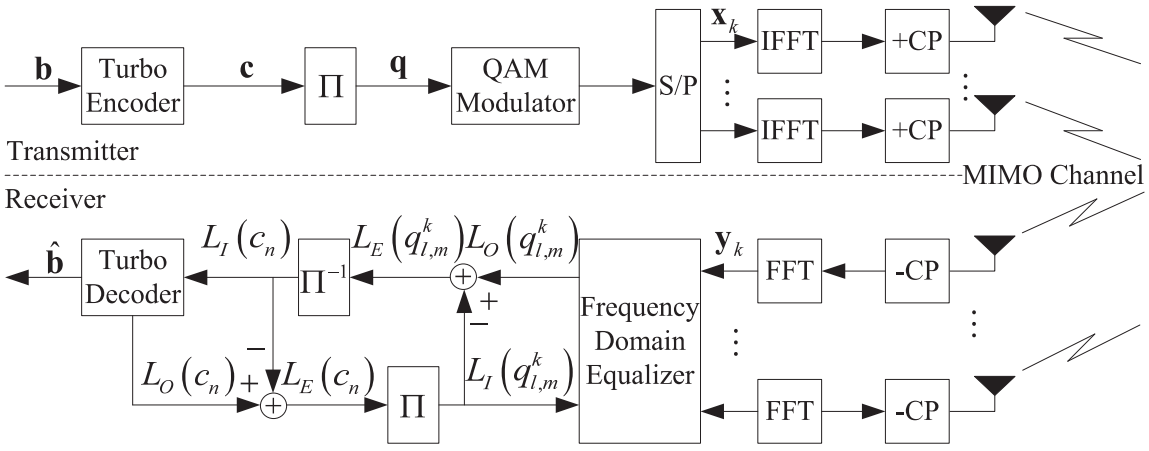


Figure 2.1: Schematic diagram of the transmitter and receiver in MIMO-OFDM systems

and then interleaved before being sent to the QAM modulator. After that, the modulated QAM symbols are mapped to transmit antennas through a series-to-parallel mapper. At each transmit antenna, OFDM modulation (an inverse FFT (IFFT) operation) is performed on QAM symbols, and a cyclic prefix (CP) is added to eliminate inter-block interference (IBI).

At the receiver side, after removing CP and FFT operations, the received signal vectors are sent to the frequency domain equalizer (FDE). The SISO FDE employs the *a priori* information, and computes *extrinsic* information of the coded bits. The *extrinsic* information is deinterleaved, and fed into the turbo decoder.

Let $\mathbf{b} = [b_0, \cdots, b_{K-1}]^{\mathrm{T}}$ denotes the message bits stream, where the message block length is K , and $\mathbf{c} = [c_0, \cdots, c_{N-1}]^{\mathrm{T}}$ denotes the output bits stream after turbo encoder, where N is the code bits length and the effective code rate is $R = K/N$. The output of the interleaver at the transmitter side is denoted by $\mathbf{q} = [q_0, \cdots, q_{N-1}]^{\mathrm{T}}$, where $q_n = \Pi(c_k)$ and $\Pi(\cdot)$ is the interleaving function. After QAM modulation and series-to-parallel mapping, the symbol vector at the k-th subcarrier are denoted by $\mathbf{x}_k = \left[x_{k,0}, \cdots, x_{k,N_T-1}\right]^{\mathrm{T}}$, where $x_{k,l} = \mathscr{Q}\left(\left[q^k_{l,0}, \cdots, q^k_{l,\log_2 M-1}\right]^{\mathrm{T}}\right)$ is the transmitted symbol from the l-th transmit antenna at

the k-th subcarrier. $\mathcal{Q}(\cdot)$ is the mapping function that generates a QAM symbol from the associated bits. $q_{l,m}^{k}$ is the corresponding m-th bit in symbol $x_{k,l}$, and M is the cardinality of the QAM symbol set χ.

The received vector after FFT is expressed as

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{z}_k, \tag{2.1}$$

where $\mathbf{y}_k = \left[y_{k,0}, \cdots, y_{k,N_R-1}\right]^{\mathrm{T}}$, and $y_{k,r}$ represents the received signal from the r-th receive antenna at the k-th subcarrier. $\mathbf{z}_k = \left[z_{k,0}, \cdots, z_{k,N_R-1}\right]^{\mathrm{T}}$ denotes the Gaussian noise vector at the k-th subcarrier, and $\mathrm{E}\{\mathbf{z}_k\} = \mathbf{0}$, $\mathrm{E}\left\{\mathbf{z}_k \mathbf{z}_k^{\mathrm{H}}\right\} = \sigma_z^2 \mathbf{I}_{N_R}$. $\mathbf{H}_k \in \mathbb{C}^{N_R \times N_T}$ is the channel matrix.

The turbo receiver exchanges *extrinsic* information between the frequency domain equalizer and turbo decoder which are separated by an interleaver $\Pi(\cdot)$ and deinterleaver $\Pi^{-1}(\cdot)$. The optimal MAP equalizer collects information of the entire block of symbols to compute the *a posteriori* information of the bits, which requires considerable computation complexity. Therefore, reduced complexity detection schemes such as MMSE equalizer are preferred in practical applications. As it is shown in [55], the MMSE equalizer first computes the estimate of the transmitted symbol, and the *a posteriori* LLR of the associated bit is computed afterwards, given by

$$\begin{aligned} L_O\left(q_{l,m}^{k}\right) &= \ln\left(\frac{P\left(q_{l,m}^{k}=1|\hat{x}_{k,l}\right)}{P\left(q_{l,m}^{k}=0|\hat{x}_{k,l}\right)}\right) \\ &= \ln\left(\frac{\sum_{s\in\chi_1^m} p\left(\hat{x}_{k,l}|x_{k,l}=s\right) P\left(x_{k,l}=s\right)}{\sum_{s\in\chi_0^m} p\left(\hat{x}_{k,l}|x_{k,l}=s\right) P\left(x_{k,l}=s\right)}\right), \end{aligned} \tag{2.2}$$

where $\chi_b^m \triangleq \{s | s \in \chi, q_m = b\}$ is the subset of QAM symbol set χ in which the m-th mapping bit of s is b. $P\left(x_{k,l}=s\right)$ is the *a priori* probability of the symbol $x_{k,l}$, given by

$$P\left(x_{k,l}=s\right) = \prod_{m=0}^{\log_2 M} P\left(q_{l,m}^{k}=b\right). \tag{2.3}$$

Substituting Equation (2.3) in Equation (2.2), the output *a posteriori* LLR of the bit $q_{l,m}^{k}$

is given by

$$L_O\left(q_{l,m}^k\right) = L_E\left(q_{l,m}^k\right) + L_I\left(q_{l,m}^k\right), \tag{2.4}$$

where $L_E\left(q_{l,m}^k\right)$ and $L_I\left(q_{l,m}^k\right)$ are *extrinsic* and *a priori* LLR of the bit $q_{l,m}^k$, respectively, given by

$$\begin{aligned} L_E\left(q_{l,m}^k\right) &= \ln\left(\frac{\sum_{s\epsilon\chi_1^m} p\left(\hat{x}_{k,l}|x_{k,l}=s\right) \prod\limits_{n\neq m} P\left(q_{l,n}^k=b\right)}{\sum_{s\epsilon\chi_0^m} p\left(\hat{x}_{k,l}|x_{k,l}=s\right) \prod\limits_{n\neq m} P\left(q_{l,n}^k=b\right)}\right), \\ L_I\left(q_{l,m}^k\right) &= \ln\left(\frac{P\left(q_{l,n}^k=1\right)}{P\left(q_{l,n}^k=0\right)}\right). \end{aligned} \tag{2.5}$$

After obtaining $L_E\left(q_{l,m}^k\right)$ from Equation (2.4), and through a deinterleaver, $L_I\left(c_n\right) = \Pi^{-1}\left(L_E\left(q_{l,m}^k\right)\right)$ is fed into the turbo decoder as *a priori* LLR of the coded bits. By using the *a priori* LLR and the trellis code structure, turbo decoder outputs the *a posteriori* LLR of the coded bits and the information bits. The previous one is used to derive the *extrinsic* LLR of the coded bits $L_E\left(c_n\right)$, while the latter one is used for final decision of $\hat{\mathbf{b}}$. At each iteration, $L_E\left(c_n\right)$ is interleaved, providing $L_I\left(q_{l,m}^k\right) = \Pi\left(L_E\left(c_n\right)\right)$ as *a priori* LLR for FDE.

2.2 SISO MMSE based Turbo Equalization under Imperfect Channel Estimation

2.2.1 Imperfect Channel Estimation Model

In practical applications, only an estimated channel matrix is available. For example, in LTE/LTE-A system, the channel estimation is performed on reference signals [21]. By using a linear MMSE channel estimator, the estimated channel matrix at the k-th subcarrier can be expressed as

$$\mathbf{H}_k = \hat{\mathbf{H}}_k + \Delta\mathbf{H}_k, \tag{2.6}$$

where $\hat{\mathbf{H}}_k \in \mathbb{C}^{N_R \times N_T}$ is the estimated channel matrix used for data detection. $\Delta\mathbf{H}_k \in \mathbb{C}^{N_R \times N_T}$ denotes the estimation error, and each entry in $\Delta\mathbf{H}_k$ is assumed to be independent and identically distributed (i.i.d), satisfying $\mathrm{E}\{\Delta\mathbf{H}_k\} = \mathbf{0}$, $\mathrm{E}\left\{\Delta\mathbf{H}_k \Delta\mathbf{H}_k^{\mathrm{H}}\right\} = N_T\sigma_e^2\mathbf{I}_{N_R}$.

With the *a priori* LLR of the associated bits for each entry in $x_{k,l}$, we can obtain the *a priori* mean and covariance:

$$\begin{aligned} \mu_{k,l} &= \sum\nolimits_{s\in\chi} s \cdot P\left(x_{k,l} = s\right) \\ v_{k,l} &= \sum\nolimits_{s\in\chi} |s|^2 P\left(x_{k,l} = s\right) - \left|\mu_{k,l}\right|^2, \end{aligned} \tag{2.7}$$

where $P\left(x_{k,l} = s\right)$ is given by Equation (2.3). Based on that, we define the *a priori* mean vector and covariance matrix as

$$\begin{aligned} \boldsymbol{\mu}_k &= \left[\mu_{k,0}, \cdots, \mu_{k,N_T-1}\right]^{\mathrm{T}}, \\ \mathbf{V}_k &= \mathrm{diag}\left\{\left[v_{k,0}, \cdots, v_{k,N_T-1}\right]^{\mathrm{T}}\right\}, \end{aligned} \tag{2.8}$$

which will be used in the following sections.

2.2.2 SISO MMSE Equalization

With the estimated channel matrix at each subcarrier, i.e., $\hat{\mathbf{H}}_k$, we can rewrite the received vector in Equation (2.1) as

$$\mathbf{y}_k = \hat{\mathbf{H}}_k\mathbf{x}_k + \mathbf{n}_k, \tag{2.9}$$

where $\mathbf{n}_k = \Delta\mathbf{H}_k\mathbf{x}_k + \mathbf{z}_k$, satisfying

$$\begin{aligned} \mathrm{E}\{\mathbf{n}_k\} &= \mathbf{0}, \\ \mathrm{E}\left\{\mathbf{n}_k\mathbf{n}_k^{\mathrm{H}}\right\} &= \boldsymbol{\Phi}_k = \sigma_n^2\mathbf{I}_{N_R}, \end{aligned} \tag{2.10}$$

where $\sigma_n^2 = \sigma_e^2\left(\sum_{l=0}^{N_T-1}\left|v_{k,l}\right|^2 + \left|\mu_{k,l}\right|^2\right) + \sigma_z^2$. The MMSE estimation $\hat{\mathbf{x}}_k = \left[\hat{x}_{k,0}, \cdots, \hat{x}_{k,N_T-1}\right]^{\mathrm{T}}$ from Equation (2.9) is given by

$$\hat{\mathbf{x}}_k = \boldsymbol{\kappa}_k\hat{\mathbf{H}}_k^{\mathrm{H}}\left(\hat{\mathbf{H}}_k\mathbf{V}_k\hat{\mathbf{H}}_k^{\mathrm{H}} + \boldsymbol{\Phi}_k\right)^{-1}\left(\mathbf{y}_k - \hat{\mathbf{H}}_k\boldsymbol{\mu}_k\right) + \boldsymbol{\kappa}_k\boldsymbol{\rho}_k\boldsymbol{\mu}_k, \tag{2.11}$$

where $\boldsymbol{\kappa}_k$ and $\boldsymbol{\rho}_k$ are both diagonal matrices, with the l-th diagonal element at each diagonal matrix are respectively given by $\kappa_{k,l} = \left(1 + \left(1 - v_{k,l}\right)\rho_{k,l}\right)^{-1}$ and $\rho_{k,l} = \mathbf{e}_l^{\mathrm{H}}\mathbf{B}_k\mathbf{e}_l$ where $\mathbf{B}_k$ is expressed as

$$\mathbf{B}_k = \hat{\mathbf{H}}_k^{\mathrm{H}}\left(\hat{\mathbf{H}}_k\mathbf{V}\hat{\mathbf{H}}_k^{\mathrm{H}} + \boldsymbol{\Phi}_k\right)^{-1}\hat{\mathbf{H}}_k. \tag{2.12}$$

From Equation (2.11), we can derive

$$\hat{x}_{k,l} = \alpha_{k,l}x_{k,l} + \xi_{k,l}, \tag{2.13}$$

where $\alpha_{k,l} = \kappa_{k,l} \cdot \rho_{k,l}$, and $\xi_{k,l}$ denotes the interference and noise component in the estimation of $x_{k,l}$. We assume $\xi_{k,l}$ is Gaussian distributed with $\mathrm{E}\left\{\xi_{k,l}\right\} = 0$, and $v_{\xi_{k,l}} = \mathrm{E}\left\{\xi_{k,l}\xi_{k,l}^{\mathrm{H}}\right\} = \kappa_{k,l}^2\rho_{k,l}\left(1 - \rho_{k,l}v_{k,l}\right)$. Therefore, the conditional probability density function (pdf) $p\left(\hat{x}_{k,l}|x_{k,l}\right)$ is given by

$$p\left(\hat{x}_{k,l}|x_{k,l} = s\right) = \frac{1}{\pi v_{\xi_{k,l}}}\exp\left(-\frac{\left|\hat{x}_{k,l} - \alpha_{k,l} \cdot s\right|^2}{v_{\xi_{k,l}}}\right), \tag{2.14}$$

Following Equations (2.2) to (2.4), the *extrinsic* LLR of the associated bits are obtained, used for turbo decoding.

2.3 SISO MMSE-SQRD based Turbo Equalization

2.3.1 Development of the Proposed Scheme

Let $\mathbf{A} = \left(\mathbf{V}_k^{1/2}\hat{\mathbf{H}}_k^{\mathrm{H}}\boldsymbol{\Phi}_k^{-1}\hat{\mathbf{H}}_k\mathbf{V}_k^{1/2} + \mathbf{I}_{N_T}\right)$, by using $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$, we obtain

$$\begin{aligned}&\left(\mathbf{V}_k^{1/2}\hat{\mathbf{H}}_k^{\mathrm{H}}\boldsymbol{\Phi}_k^{-1}\hat{\mathbf{H}}_k\mathbf{V}_k^{1/2} + \mathbf{I}_{N_T}\right)^{-1}\mathbf{V}_k^{1/2}\hat{\mathbf{H}}_k^{\mathrm{H}}\boldsymbol{\Phi}_k^{-1}\hat{\mathbf{H}}_k \\ &\quad = \mathbf{V}_k^{-1/2} - \left(\mathbf{V}_k^{1/2}\hat{\mathbf{H}}_k^{\mathrm{H}}\boldsymbol{\Phi}_k^{-1}\hat{\mathbf{H}}_k\mathbf{V}_k^{1/2} + \mathbf{I}_{N_T}\right)^{-1}\mathbf{V}_k^{-1/2}.\end{aligned} \tag{2.15}$$

In order to develop the extended channel matrix, we adjust $\hat{\mathbf{x}}_k$ in Equation (2.9) to an equivalent expression given by

$$\begin{aligned}\hat{\mathbf{x}}_k =& \kappa_k\mathbf{V}_k^{-1/2}\left(\mathbf{V}_k^{1/2}\hat{\mathbf{H}}_k^{\mathrm{H}}\boldsymbol{\Phi}_k^{-1}\hat{\mathbf{H}}_k\mathbf{V}_k^{1/2} + \mathbf{I}_{N_T}\right)^{-1}\left(\mathbf{V}_k^{1/2}\hat{\mathbf{H}}_k^{\mathrm{H}}\boldsymbol{\Phi}_k^{-1}\mathbf{y}_k + \mathbf{V}_k^{1/2}\boldsymbol{\mu}_k\right) \\ &+ \kappa_k\mathbf{V}_k^{-1} \cdot \mathrm{diag}\left\{\left(\mathbf{V}_k^{1/2}\hat{\mathbf{H}}_k^{\mathrm{H}}\boldsymbol{\Phi}_k^{-1}\hat{\mathbf{H}}_k\mathbf{V}_k^{1/2} + \mathbf{I}_{N_T}\right)^{-1}\right\}\boldsymbol{\mu}_k,\end{aligned} \tag{2.16}$$

where Equation (2.15) is used for the derivation. To obtain the conditional pdf $p\left(\hat{x}_{k,l}|x_{k,l}\right)$, Equation (2.16) is equivalent to

$$\begin{aligned}\hat{\mathbf{x}}_k =&\left(\mathbf{V}_k^{1/2}\hat{\mathbf{H}}_k^{\mathrm{H}}\boldsymbol{\Phi}_k^{-1}\hat{\mathbf{H}}_k\mathbf{V}_k^{1/2}+\mathbf{I}_{N_T}\right)^{-1}\left(\mathbf{V}_k^{1/2}\hat{\mathbf{H}}_k^{\mathrm{H}}\boldsymbol{\Phi}_k^{-1}\mathbf{y}_k+\mathbf{V}_k^{1/2}\mu_k\right)\\&+\mathbf{V}_k^{-1/2}\cdot\mathrm{diag}\left\{\left(\mathbf{V}_k^{1/2}\hat{\mathbf{H}}_k^{\mathrm{H}}\boldsymbol{\Phi}_k^{-1}\hat{\mathbf{H}}_k\mathbf{V}_k^{1/2}+\mathbf{I}_{N_T}\right)^{-1}\right\}\boldsymbol{\mu}_k.\end{aligned} \tag{2.17}$$

In addition, we rewrite $\mathbf{B}_k$ and $\boldsymbol{\rho}_k$ as

$$\begin{aligned}\mathbf{B}_k =&\mathbf{V}_k^{-1/2}\left(\mathbf{V}_k^{1/2}\hat{\mathbf{H}}_k^{\mathrm{H}}\boldsymbol{\Phi}_k^{-1}\hat{\mathbf{H}}_k\mathbf{v}_k^{1/2}+\mathbf{I}_{N_T}\right)^{-1}\mathbf{V}_k^{1/2}\hat{\mathbf{H}}_k^{\mathrm{H}}\boldsymbol{\Phi}_k^{-1}\hat{\mathbf{H}}_k\\=&\mathbf{V}_k^{-1}-\mathbf{V}_k^{-1/2}\left(\mathbf{V}_k^{1/2}\hat{\mathbf{H}}_k^{\mathrm{H}}\boldsymbol{\Phi}_k^{-1}\hat{\mathbf{H}}_k\mathbf{V}_k^{1/2}+\mathbf{I}_{N_T}\right)^{-1}\mathbf{V}_k^{-1/2},\\\boldsymbol{\rho}_k =&\mathrm{diag}\left\{\mathbf{B}_k\right\}\\=&\mathbf{V}_k^{-1}-\mathbf{V}_k^{-1}\cdot\mathrm{diag}\left\{\left(\mathbf{V}_k^{1/2}\hat{\mathbf{H}}_k^{\mathrm{H}}\boldsymbol{\Phi}_k^{-1}\hat{\mathbf{H}}_k\mathbf{V}_k^{1/2}+\mathbf{I}_{N_T}\right)^{-1}\right\}.\end{aligned} \tag{2.18}$$

From Equation (2.17), we define the extended channel matrix and received vector as

$$\underline{\mathbf{H}}_k=\begin{bmatrix}\boldsymbol{\Phi}_k^{-1/2}\hat{\mathbf{H}}_k\mathbf{V}_k^{1/2}\\\mathbf{I}_{N_T}\end{bmatrix}, \tag{2.19}$$

$$\underline{\mathbf{y}}_k=\begin{bmatrix}\boldsymbol{\Phi}_k^{-1/2}\mathbf{y}_k\\\mathbf{V}_k^{1/2}\boldsymbol{\mu}_k\end{bmatrix}. \tag{2.20}$$

Hence, Equation (2.17) is rewritten to

$$\hat{\mathbf{x}}_k=\left(\underline{\mathbf{H}}_k^{\mathrm{H}}\underline{\mathbf{H}}_k\right)^{-1}\underline{\mathbf{H}}_k^{\mathrm{H}}\underline{\mathbf{y}}_k-\mathrm{diag}\left\{\left(\underline{\mathbf{H}}_k^{\mathrm{H}}\underline{\mathbf{H}}_k\right)^{-1}\right\}\mathbf{V}_k^{-1/2}\boldsymbol{\mu}_k. \tag{2.21}$$

Similar to [62] and [41], the QR decomposition of the extended channel matrix $\underline{\mathbf{H}}_k$ is given by

$$\underline{\mathbf{H}}_k=\mathbf{QR}=\begin{bmatrix}\mathbf{Q}_1\\\mathbf{Q}_2\end{bmatrix}\mathbf{R}, \tag{2.22}$$

where $\mathbf{R}\in\mathbb{C}^{N_T\times N_T}$ is an upper triangular matrix, and $\mathbf{Q}\in\mathbb{C}^{(N_R+N_T)\times N_T}$, satisfying $\mathbf{Q}^{\mathrm{H}}\mathbf{Q}=\mathbf{I}_{N_T}$. $\mathbf{Q}_1\in\mathbb{C}^{N_R\times N_T}$ and $\mathbf{Q}_2\in\mathbb{C}^{N_T\times N_T}$ are the partitions of $\mathbf{Q}$. We have the following equations established:

$$\mathbf{Q}^{\mathrm{H}}\underline{\mathbf{H}}_k=\mathbf{Q}_1^{\mathrm{H}}\boldsymbol{\Phi}_k^{-1/2}\hat{\mathbf{H}}_k\mathbf{V}_k^{1/2}+\mathbf{Q}_2^{\mathrm{H}}=\mathbf{R}, \tag{2.23}$$

$$\mathbf{Q}_1^{\mathrm{H}}\mathbf{Q}_1 + \mathbf{Q}_2^{\mathrm{H}}\mathbf{Q}_2 = \mathbf{I}_{N_T}, \tag{2.24}$$

$$\mathbf{Q}_1\mathbf{R} = \mathbf{\Phi}_k^{-1/2}\hat{\mathbf{H}}_k\mathbf{V}_k^{-1/2}, \tag{2.25}$$

$$\mathbf{Q}_2\mathbf{R} = \mathbf{I}_{N_T}. \tag{2.26}$$

Substituting $\underline{\mathbf{H}}_k$ with $\mathbf{Q}_1$, $\mathbf{Q}_2$, and $\mathbf{R}$ in Equation (2.21), and defining $\tilde{\mathbf{x}}_k = \mathbf{R}\hat{\mathbf{x}}_k + \mathbf{R}\mathbf{V}_k^{-1/2} \cdot \mathrm{diag}\left\{\mathbf{Q}_2\mathbf{Q}_2^{\mathrm{H}}\right\}\boldsymbol{\mu}_k$, we have

$$\begin{aligned}\tilde{\mathbf{x}}_k =& \mathbf{Q}_1^{\mathrm{H}}\mathbf{\Phi}_k^{-1/2}\mathbf{y} + \mathbf{Q}_2^{\mathrm{H}}\mathbf{V}_k^{-1/2}\boldsymbol{\mu}_k \\ =& \mathbf{R}\mathbf{V}_k^{-1/2}\mathbf{x}_k - \mathbf{Q}_2^{\mathrm{H}}\mathbf{V}_k^{-1/2}\left(\mathbf{x}_k - \boldsymbol{\mu}_k\right) + \mathbf{Q}_1^{\mathrm{H}}\mathbf{\Phi}_k^{-1/2}\mathbf{n}_k.\end{aligned} \tag{2.27}$$

We redefine $\underline{\mathbf{R}} = \mathbf{R}\mathbf{V}_k^{-1/2}$, and it is easy to show that $\underline{\mathbf{R}}$ is an upper triangular matrix; therefore, we can introduce SIC algorithms to estimate the transmitted symbol $\mathbf{x}_{k,l}$. In Equation (2.27), the first component shows the interference from detected symbols; the second component reveals the interference from undetected symbols; the third component denotes the noise. The expected signal is contained in the first and second components. In SIC, the detection order is significant in system performance [41, 62]. We will discuss the ordering process in Subsection 2.3.2.

Suppose we are given the detection order $\mathbf{p} = [N_T - 1, \cdots, 0]$, and we adjust the mean vector μ_k and covariance matrix $\mathbf{v}_k$ in accordance. When the l-th transmitted symbol is in detection, with the previous detected symbols, we generate the *a posteriori* mean $\mu_{k,l'}^p$ and covariance $v_{k,l'}^p$. According to (2.24), we have $\tilde{\mu}_{k,l} = \mathrm{E}\left\{\tilde{x}_{k,l}|x_{k,l}\right\}$ given by

$$\tilde{\mu}_{k,l} = \mathbf{e}_l^H\left(\underline{\mathbf{R}} - \mathbf{Q}_2^H\mathbf{V}_k^{-1/2}\right)\mathbf{e}_l \cdot x_{k,l} + \sum_{l'=l+1}^{N_T-1}\mathbf{e}_l^H\underline{\mathbf{R}}\mathbf{e}_{l'} \cdot \mu_{k,l'}^p + \mathbf{e}_l^H\mathbf{Q}_2^H\mathbf{V}_k^{-1/2}\mathbf{e}_l\mu_{k,l}, \tag{2.28}$$

and $\tilde{v}_{k,l} = \mathrm{E}\left\{\left|\tilde{x}_{k,l} - \tilde{\mu}_{k,l}\right|^2|x_{k,l}\right\}$ given by

$$\tilde{v}_{k,l} = \sum_{l'=l+1}^{N_T-1}\left|\underline{\mathbf{R}}_{l,l'}\right|^2 v_{k,l'}^p + 1 - \left|(\mathbf{Q}_2)_{l,l}\right|^2. \tag{2.29}$$

Based on Equations (2.28) and (2.29), we can compute the conditional pdf $p\left(\tilde{x}_{k,l}|x_{k,l}\right)$ given by

$$p\left(\tilde{x}_{k,l}|x_{k,l}=s\right)=\frac{1}{\pi\tilde{v}_{k,l}}\exp\left(-\frac{\left|\tilde{x}_{k,l}-\tilde{\mu}_{k,l}\right|^2}{\tilde{v}_{k,l}}\right), \tag{2.30}$$

which can be used for the calculation of the *extrinsic* LLR of the bits associated with the symbol $x_{k,l}$, following the same procedure in Subsection 2.2.2.

The *a posteriori* mean and covariance of the estimated symbols are updated by Equation (2.7) where the only difference is using the *a posteriori* LLR of the associated bits provided by Equation (2.2).

2.3.2 Sorted QR decomposition

Ordering is significant in SIC algorithms due to the error propagation problems. In [62], a sorting algorithm is proposed for MMSE-SQRD detection; however, it does not consider an iterative processing. The authors in [41] make a few modifications in turbo equalization where the *a priori* information is taken into account. In our case, when the channel estimation errors are considered, we find the extended channel matrix has a similar expression to the extended channel matrix in [62]. Therefore, the sorting algorithm in [62] will be applied in our case. The difference is that the extended channel matrix in our case consists of the *a priori* information and channel estimation errors, which is specific for turbo equalization under imperfect channel estimation.

Following the ordering process in [62], we obtain the upper triangular matrix $\mathbf{R}$, and $\mathbf{Q}_1$, $\mathbf{Q}_2$, as well as the order $\mathbf{p}$. As indicated in Subsection 2.3.1, $\mathbf{p}$ will be used to adjust the mean vector μ_k and covariance matrix $\mathbf{V}_k$ for MMSE-SQRD detection. After computing the *extrinsic* LLR of the associated bits, $L_E\left(q_{l,m}^k\right)$ will be adjust in accordance with the detection order $\mathbf{p}$ as well.

2.3.3 Complexity Analysis

Compared to the algorithms proposed in [41], the only difference is the extended channel matrix where ours takes into account channel estimation error. The computation of the extended channel matrix is of the order of $O(N_T N_R)$. According to [41], the proposed MMSE-SQRD based turbo equalization has an overall computational complexity in the order of $O\left(N_T^3 + N_T^2 N_R + N_T M \log_2 M\right)$; therefore, the computational complexity of the proposed turbo equalization that takes into account of the channel estimation error is in the order of $O\left(N_T^3 + N_T^2 N_R + N_T M \log_2 M\right)$. As for the SISO MMSE detection, the computational complexity is in the order of $O\left(N_T N_R^3 + N_T^2 N_R + N_T M \log_2 M\right)$. Comparatively, the computation cost in our proposed turbo equalization is less than that in SISO MMSE detection when $N_R \geqslant N_T$.

2.4 Numerical Simulations

Montel-Carlo simulations are performed to evaluate the performance of the proposed turbo equalization in turbo coded MIMO-OFDM systems with imperfect channel estimation. Different modulation and coding schemes (MCS) are considered. The system bandwidth is 5MHz with 300 useful subcarriers in use. The Extended Vehicular A (EVA) model is used in the simulation with the system configuration of N_T = 4 transmit antennas and N_R = 4 receive antennas. The imperfect channel knowledge is available at the receiver and the random channel estimation error is generated by independent and identically distributed (i.i.d.) Gaussian noise.

Figure 2.2 shows the BER performance versus the average SNR at the receiver for quadrature phase shift keying (QPSK) modulated symbols with effective turbo code rate of 0.588 (MCS index 6). In the case of perfect channel estimation, as we can see, the

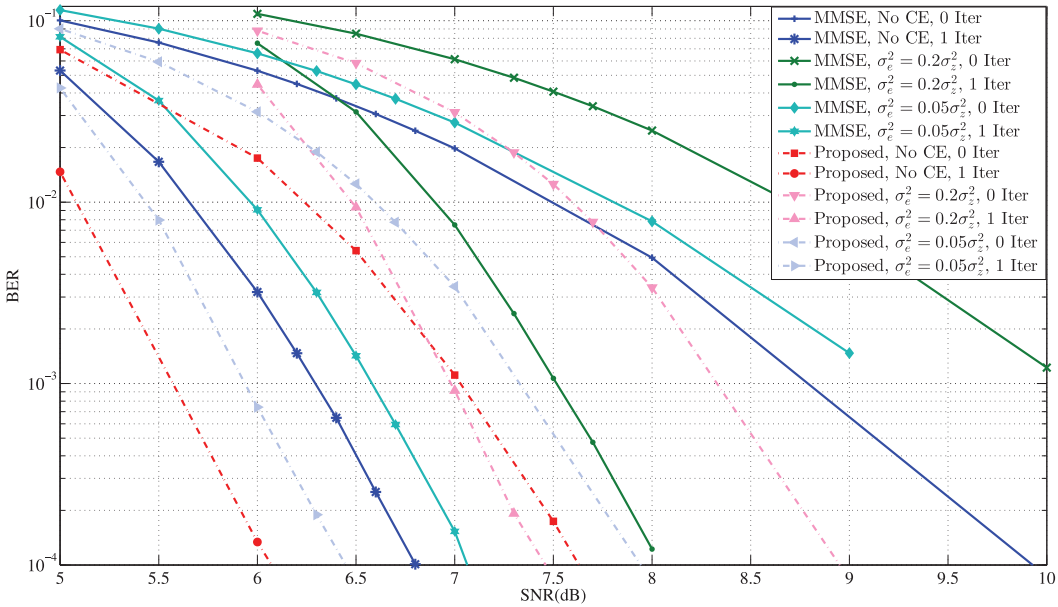


Figure 2.2: BER performance of the proposed scheme compared to SISO MMSE turbo equalization, QPSK

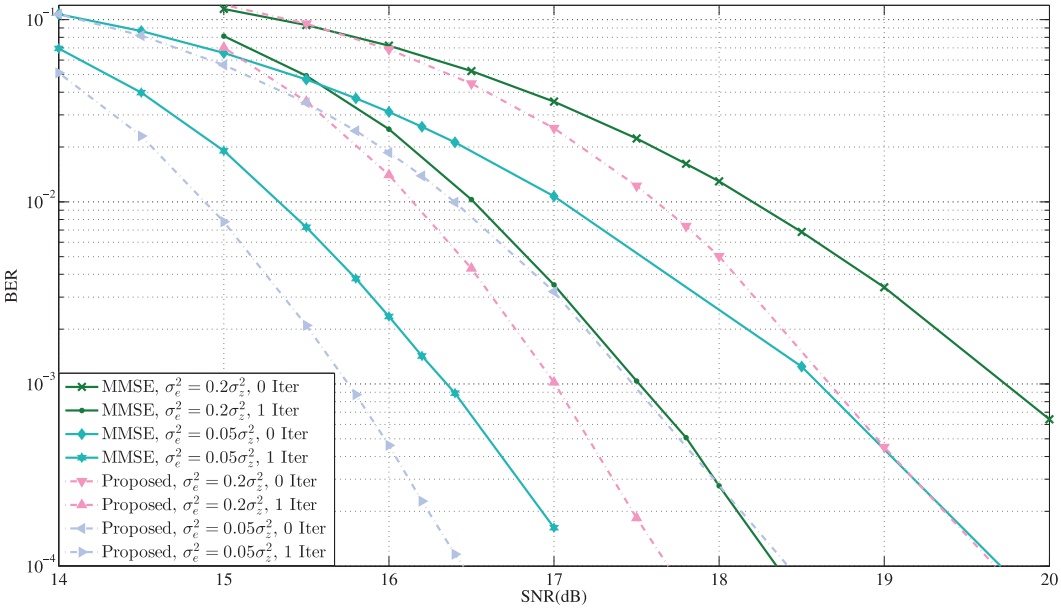


Figure 2.3: BER performance of the proposed scheme compared to SISO MMSE turbo equalization, 16QAM

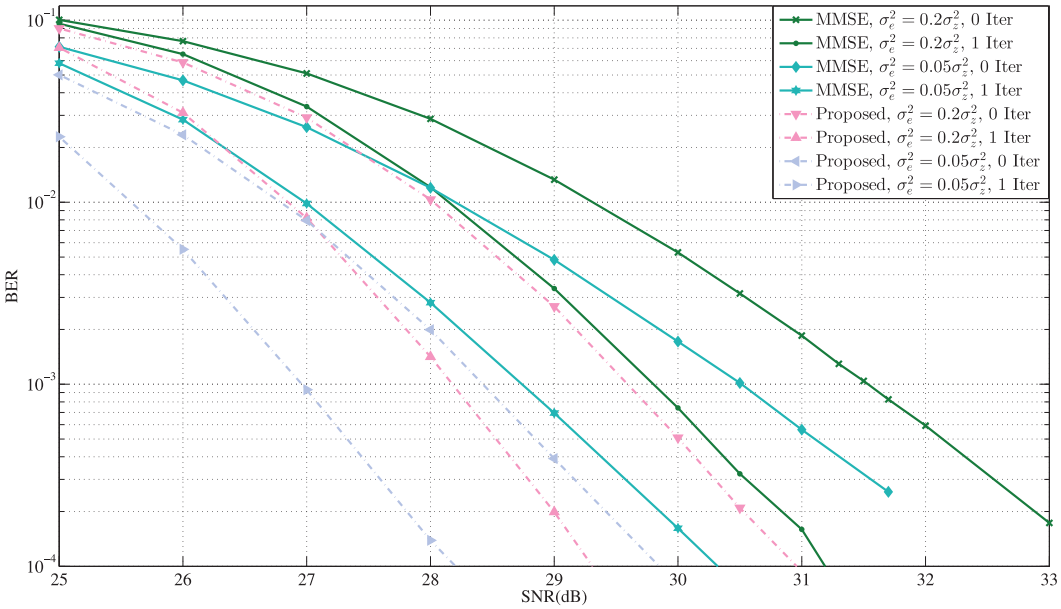


Figure 2.4: BER performance of the proposed scheme compared to SISO MMSE turbo equalization, 64QAM

BER performance of the proposed MMSE-SQRD based turbo equalization is always better than that of the traditional SISO MMSE turbo equalization. When the channel estimation error is present, both schemes experience performance degradation. However, the BER performance of the proposed MMSE-SQRD based turbo equalization always outperforms that of the MMSE scheme. It is worth noting that at high SNR region, the performance enhancement is remarkable compared to that in the low SNR region. This phenomenon can be explained as follows. Apparently, at high SNR region, inter-symbol interference (ISI) is significant in current symbol detection. By performing successive interference cancelation, the proposed MMSE-SQRD based turbo equalization can efficiently constrain the error propagation issue, thus leading to considerable performance gain.

The BER performance for 16QAM with effective code rate of 0.601 (MCS index 9) is given in Figure 2.3. Similar performance improvement of the proposed MMES-SQRD based turbo equalization scheme can be seen from Figure 2.3. Though the BER performance

degradation is caused by imperfect channel estimation, our proposed turbo equalization scheme can efficiently alleviate this practical issue. Compared to the traditional SISO MMSE turbo equalization, the proposed turbo equalization achieves better BER performance in all iterations.

Figure 2.4 presents the BER performance of 64QAM with effective code rate of 0.754 (MCS index 13). It is clear that the proposed MMSE-SQRD based turbo equalization scheme outperforms the SISO MMSE turbo equalization scheme. We even find that the BER performance of the proposed turbo equalization at the first iteration surpasses the SISO MMSE turbo equalization at the second iteration, which indicates the interference cancelation is significant in high order modulation schemes.

It is also worth noting that when the channel estimation improves, the BER performance of the proposed turbo equalization is enhanced as well. Recalling that data detection can be used to help with channel estimation [63, 64], the system performance will be further improved if the channel is re-estimated with aided data symbols. Therefore, future work of this study can be joint data detection and channel estimation by employing the proposed turbo equalization scheme in both data detection and channel estimation.

2.5 Summary

In this chapter, we conduct the study of the SISO MMSE-SQRD based turbo equalization under imperfect channel estimation. We introduce a new extended channel matrix in developing the MMSE-SQRD detection scheme. The new extended channel matrix takes into account the channel estimation error and the *a priori* information (obtained from the channel decoder). During the detection, the *a posteriori* information (updated from the previous estimated transmitted symbol) is also considered in SIC. The ordering process is adjusted by introducing the new extended channel matrix. Compared to the SISO MMSE

turbo equalization scheme, the proposed MMSE-SQRD based turbo equalization requires less computation complexity. Numerical simulations are also conducted in MIMO-OFDM system with the MIMO size of 4×4. Different modulation and coding schemes are used in the simulation. Simulation results show that the proposed MMSE-SQRD based turbo equalization scheme outperforms the SISO MMSE turbo equalization scheme when channel estimation error exists. The performance enhancement is more obvious in high SNR region where the ISI is severe. The reason can be attributed to efficient restrain of error propagation and interference cancellation in the proposed MMSE-SQRD based turbo equalization scheme. Simulation results also reveal that when the channel estimation errors decrease, the system BER performance improvement is significant. This indicates that using the detected data to further improve channel estimation can in turn improve data detection performance, which can be our future research topic.

Chapter 3

Low Complexity Turbo Equalizers Conditioned on Channel Estimate for MIMO Systems

[1] In this chapter, we investigate low-complexity turbo equalization schemes conditioned on channel estimate for MIMO systems. We start from the MMSE-based turbo receivers by taking into account the channel estimation error and the *a priori* information provided by the channel decoder. We then show that under the assumption that all transmit symbols are independent and with normalized power constraint, our results in the first iteration degrade to that in [68, 70]. In addition, we show that when the channel estimation error is

[1] The related work has been published in or submitted to:

- *IEEE Transactions on Communications*, 2018, under review.
- *IEEE Communications Letters*, vol. 21, no. 4, pp. 957–960, April 2017.
- *Proceedings of the 25th IEEE Newfoundland Electrical and Computer Engineering Conference* (NECEC'16), St. John's, NL, Canada, November 2016.

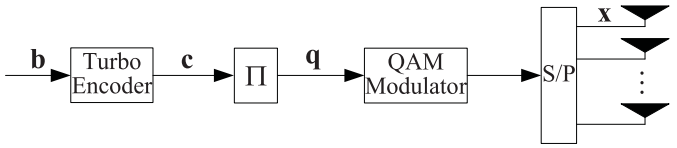
low, the MMSE-based turbo equalization scheme conditioned on channel estimate degrades to the schemes used in [21] where the channel estimation error is assumed to be independent from channel estimate. Moreover, we also show that the conventional MMSE-based turbo equalization scheme, where the channel estimate is directly applied to the equalization process, can be included as a special case in our MMSE-based turbo equalization conditioned on channel estimate. Based on the development of the MMSE-based turbo equalization, we then investigate the SQRD-based turbo equalization conditioned on channel estimate. Specially, we define the new extended channel matrix where the channel estimation error and the *a priori* covariance matrix are considered. Based on that, we develop the SQRD-based turbo equalization, and successive detection is introduced to complete the symbol estimation. We also show that the new SQRD-based turbo equalization scheme is a general expression and the SQRD-based detection schemes in [39, 41] are special cases.

The rest of this chapter is organized as follows. The system model is described in Section 3.1, followed by the development of a MMSE-based turbo equalization scheme conditioned on channel estimate in Section 3.2. In Section 3.3, we develop the SQRD-based turbo equalization scheme conditioned on channel estimate. After that, we present the numerical results in Section 3.4. We finally summarize this chapter in Section 3.5.

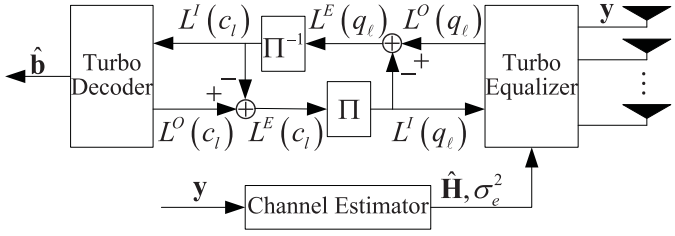
3.1 System Model

3.1.1 Transmitter and Receiver Architecture

The architecture of the transmitter and receiver is shown in Figure 3.1. As we can see from Figure 3.1(a), the original bit stream is first sent to the channel encoder. After encoding, the coded bit stream is interleaved, and then fed into QAM modulator, where the bits are mapped to QAM symbols. Through a serial-to-parallel convertor, The QAM



(a) Block diagram of the transmitter



(b) Block diagram of the receiver

Figure 3.1: Schematic diagram of the transmitter and receiver architecture in MIMO systems

symbols are spread to multiple transmit antennas. This structure is also known as bit-interleaved coded modulation [48,52,54]. We consider an iterative receiver structure, which is shown in Figure 3.1(b). From 3.1(b), we can see that two soft-input soft-output (SISO) modules, i.e., SISO turbo equalizer and SISO turbo decoder, are employed. The turbo equalizer utilizes the *a priori* information (which is the *extrinsic* information from the turbo decoder), together with the channel estimator output, to accomplish the computation of the *a posteriori* information of the bits associated with QAM symbols. In addition, the *extrinsic* information of the bits associated with the QAM symbols are obtained, interleaved, and then fed into turbo decoder as the *a priori* information of the coded bits. This turbo structure has been widely studied in [39–41, 46, 48, 49, 52–58].

To be specific, we consider a MIMO system with N_T transmit antennas and N_R receive antennas. At the transmitter side, the information bit stream, $\mathbf{b} = [b_0, b_1, \cdots, b_{K-1}]^{\mathrm{T}}$ where K is the block length, is first fed into the channel encoder. $\mathbf{c} = [c_0, c_1, \cdots, c_{L-1}]^{\mathrm{T}}$ corresponds to

the coded bit stream of length L, and $R = K/L$ is the effective code rate. After interleaving, $\mathbf{q} = [q_0, q_1, \cdots, q_{L-1}]^{\mathrm{T}}$, where $q_\ell = \Pi(c_l)$ and $\Pi(\cdot)$ is the interleaving function. QAM modulator follows, generating the QAM symbol $x_n = \mathscr{Q}(\mathbf{q}_n)$, where $\mathbf{q}_n = \left[q_{n,0}, q_{n,1}, \cdots, q_{n,\log_2 M-1}\right]^{\mathrm{T}}$, and $\mathscr{Q}(\cdot)$ denotes the operation to map $\log_2 M$ bits to a QAM symbol. x_n is the transmit QAM symbol from the n-th transmit antenna, and $q_{n,m}$ denotes the m-th mapping bit of x_n. M is the cardinality of the QAM symbol set $\chi = \{s_0, s_1, \cdots, s_{M-1}\}$. Let $\mathbf{x} = [x_0, x_1, \cdots, x_{N_T-1}]^{\mathrm{T}}$ be the symbol vector after spreading to transmit antennas. Passing through the MIMO channel, the received signal at the receiver side is given by

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

where $\mathbf{y} = [y_0, y_1, \cdots, y_{N_R-1}]^{\mathrm{T}}$ is a complex-valued $N_R \times 1$ receiving vector with y_r denoting the received symbol from the r-th receive antenna. $\mathbf{H} = [\mathbf{h}_0, \mathbf{h}_1, \cdots, \mathbf{h}_{N_T-1}]$ is the channel matrix with $\mathbf{h}_n \in \mathbb{C}^{N_R \times 1}$ denoting the channel response from the n-th transmit antenna to receive antennas. $\mathbf{z} = [z_0, z_1, \cdots, z_{N_R-1}]^{\mathrm{T}}$ is the noise vector, satisfying $E\left\{\mathbf{z}\mathbf{z}^{\mathrm{H}}\right\} = \sigma_z^2 \mathbf{I}_{N_R}$, with each entry modeled as zero-mean complex Gaussian circularly symmetric (ZMCGCS) random variable. Although the system model in (3.1) applies to single-carrier MIMO systems over flat fading channels, it also holds for OFDM/SC-FDMA systems over frequency selective channel at each subcarrier.

3.1.2 Turbo Structure

We employ the turbo receiver to recover the original bit information as shown in Figure 3.1(b). With the *extrinsic* information (i.e., log likelihood ratio (LLR) of the coded bits) from channel decoder, after interleaving, we have $L^{\mathrm{I}}(q_\ell) = L^{\mathrm{E}}(c_l)$. Therefore, the probability distribution of q_ℓ is given by

$$P(q_\ell = b) = \frac{\exp\left((2b-1)\, L^{\mathrm{I}}(q_\ell)/2\right)}{\exp\left(-L^{\mathrm{I}}(q_\ell)/2\right) + \exp\left(L^{\mathrm{I}}(q_\ell)/2\right)}, \quad b = 0, 1. \tag{3.2}$$

With the knowledge of the bit probability associated with a QAM symbol, we derive the probability of the QAM symbols through

$$P\left(x_n = s_i\right) = \prod_{m=0}^{\log_2 M-1} P\left(q_{n,m} = b_{i,m}\right), \tag{3.3}$$

where $s_i \in \chi$, $i = 0, 1, \cdots, M-1$. $b_{i,m}$ denotes the mth mapping bit of the QAM symbol s_i. In addition, we have the *a priori* mean $\mu_n = \mathrm{E}\{x_n\}$ and covariance $v_n = \mathrm{E}\left\{|x_n - \mu_n|^2\right\}$ given by

$$\mu_n = \sum_{i=0}^{M-1} s_i \cdot P\left(x_n = s_i\right), \tag{3.4}$$

$$v_n = \sum_{i=0}^{M-1} |s_i|^2 \cdot P\left(x_n = s_i\right) - |\mu_n|^2. \tag{3.5}$$

μ_n and v_n will be used for the equalization process.

The turbo equalizer outputs the *a posteriori* LLR, $L^{\mathrm{O}}\left(q_\ell\right)$. The optimal computation of $L^{\mathrm{O}}\left(q_\ell\right)$ is obtained by using the maximum *a posteriori* (MAP) equalizer, given by

$$\begin{aligned} L\left(q_{n,m}\right) &= \ln\left(\frac{P\left(q_{n,m}=1|\mathbf{y}\right)}{P\left(q_{n,m}=0|\mathbf{y}\right)}\right) = \ln\left(\frac{\sum_{s_i\in\chi_m^1} P\left(x_n=s_i|\mathbf{y}\right)}{\sum_{s_j\in\chi_m^0} P\left(x_n=s_j|\mathbf{y}\right)}\right) \\ &= \ln\left(\frac{\sum_{s_i\in\chi_m^1} p\left(\mathbf{y}|x_n=s_i\right) P\left(x_n=s_i\right)}{\sum_{s_j\in\chi_m^0} p\left(\mathbf{y}|x_n=s_j\right) P\left(x_n=s_j\right)}\right), \end{aligned} \tag{3.6}$$

where $\chi_m^b = \{s_i | s_i \in \chi, b_{i,m} = b\}$ is the subset of χ with the mth mapping bit of s_i, i.e., $b_{i,m}$, is b. It is obvious that Equation (3.6) is computationally prohibitive. In practical applications, we utilize a symbol-by-symbol detection method; as a result, we first estimate the transmitted symbol x_n, and then derive the LLR of the associated bits to approximate the optimal $L\left(q_{n,m}\right)$. Specifically, after obtaining $\hat{x}_n$, we derive

$$\begin{aligned} L^{\mathrm{O}}\left(q_{n,m}\right) &= \ln\left(\frac{P\left(q_{n,m}=1|\hat{x}_n\right)}{P\left(q_{n,m}=0|\hat{x}_n\right)}\right) = \ln\left(\frac{\sum_{\forall s_i: s_i\in\chi_m^1} P\left(x_n=s_i|\hat{x}_n\right)}{\sum_{\forall s_j: s_j\in\chi_m^0} P\left(x_n=s_j|\hat{x}_n\right)}\right) \\ &= \ln\left(\frac{\sum_{\forall s_i: s_i\in\chi_m^1} p\left(\hat{x}_n|x_n=s_i\right)\cdot P\left(x_n=s_i\right)}{\sum_{\forall s_j: s_j\in\chi_m^0} p\left(\hat{x}_n|x_n=s_j\right)\cdot P\left(x_n=s_i\right)}\right). \end{aligned} \tag{3.7}$$

Using the *a priori* information from turbo decoder, we substitute Equation (3.3) into (3.7),

yielding

$$L^{O}\left(q_{n,m}\right) = \ln\left(\frac{\sum_{\forall s_i: s_i \in \chi_m^1} p\left(\hat{x}_n | x_n = s_i\right) \cdot \prod_{l=0}^{\log_2 M-1} P\left(q_{n,l} = b_{i,l}\right)}{\sum_{\forall s_j: s_j \in \chi_m^0} p\left(\hat{x}_n | x_n = s_j\right) \cdot \prod_{l=0}^{\log_2 M-1} P\left(q_{n,l} = b_{j,l}\right)}\right) \tag{3.8}$$

$$= L^{\mathrm{E}}\left(q_{n,m}\right) + L^{I}\left(q_{n,m}\right),$$

where the *extrinsic* LLR $L^{\mathrm{E}}\left(q_{n,m}\right)$ is given by

$$L^{\mathrm{E}}\left(q_{n,m}\right) = \ln\left(\frac{\sum_{\forall s_i: s_i \in \chi_m^1} p\left(\hat{x}_n | x_n = s_i\right) \cdot \prod_{l=0, l\neq m}^{\log_2 M-1} P\left(q_{n,l} = b_{i,l}\right)}{\sum_{\forall s_j: s_j \in \chi_m^0} p\left(\hat{x}_n | x_n = s_j\right) \cdot \prod_{l=0, l\neq m}^{\log_2 M-1} P\left(q_{n,l} = b_{j,l}\right)}\right). \tag{3.9}$$

As shown in Figure 3.1(b), the $L^{\mathrm{E}}\left(q_{n,m}\right)$ is deinterleaved, providing the *a priori* LLR of the coded bits $L^{I}\left(c_l\right)$ for turbo decoder. Together with the turbo code trellis structure, after the decoding process, the turbo decoder outputs the *a posteriori* LLR of the original information bits and the coded bits $L^{O}\left(c_l\right)$. The previous one is used for the estimation of $\mathbf{b}$, while the latter one is used to derive the extrinsic LLR of the coded bits $L^{\mathrm{E}}\left(c_l\right)$.

3.1.3 Imperfect Channel Estimation

The channel estimate used for data detection is obtained through a channel estimator, and channel estimation is performed on pilot symbols or a training sequence. Similar to the works in [67, 68, 70, 71], the estimated channel matrix $\hat{\mathbf{H}}$ can be modeled as

$$\hat{\mathbf{H}} = \mathbf{H} + \Delta\mathbf{H}, \tag{3.10}$$

where $\Delta\mathbf{H} = [\Delta\mathbf{h}_0, \Delta\mathbf{h}_1, \cdots, \Delta\mathbf{h}_{N_T-1}]$ indicates the random channel estimation error, and $\Delta\mathbf{h}_n$ is independent and identically distributed, satisfying

$$\mathrm{E}\left\{\Delta\mathbf{h}_n \Delta\mathbf{h}_m^{\mathrm{H}}\right\} = \begin{cases} \mathbf{0} & n \neq m \\ \sigma_e^2 \mathbf{I}_{N_R} & n = m \end{cases}$$

Channel estimation error $\Delta\mathbf{H}$ are orthogonal to the real channel matrix, i.e., $\mathrm{E}\left\{\Delta\mathbf{H} \cdot \mathbf{H}^{\mathrm{H}}\right\} = \mathbf{0}$.

3.2 MMSE-Based Turbo Equalization Conditioned on Channel Estimate

3.2.1 Development of the Proposed Scheme

As channel estimation error is inevitable, considering channel estimation in equalizer design is significant. Recent studies clearly show that MMSE receiver design conditioned on channel estimate will greatly enhance the system BER performance. However, none of them takes into account the *a priori* information from the channel decoder. In this section, we present the detailed derivation of the MMSE-based turbo equalization process conditioned on channel estimate.

To begin with, we present the following theorem:

Theorem 1. *Let* $\mathbf{z_i}$ *and* $\mathbf{z_j}$ *be ZMCGCS random vectors with* $\mathbf{\Sigma}_{ij} = E\{\mathbf{z}_i\mathbf{z}_j^H\}$*; the conditional* $\mathbf{z_i}$ *given* $\mathbf{z_j}$ *is complex Gaussian circularly symmetric with* $E\{\mathbf{z}_i|\mathbf{z}_j\} = \mathbf{\Sigma}_{ij}\mathbf{\Sigma}_{jj}^{-1}\mathbf{z}_j$*, and* $\text{cov}\{\mathbf{z}_i, \mathbf{z}_i|\mathbf{z}_j\} = \mathbf{\Sigma}_{ii} - \mathbf{\Sigma}_{ij}\mathbf{\Sigma}_{jj}^{-1}\mathbf{\Sigma}_{ji}$.

The proof of this theorem can be found in [94]. It is worth noting that this theorem has been used in [68, 70, 71] for the derivation.

Before presenting the results of the MMSE-based turbo equalization conditioned on channel estimate, we first have the following assumptions.

Assumption 1: The noise component is independent from the channel and data symbols.

Assumption 2: Channel estimation error is independent and identically distributed.

Assumption 3: Data symbols are independent from channel data.

Assumption 4: Data symbols are independent from each other.

Assumption 1 to **4** are commonly used in the development of the turbo equalization schemes in [39, 41, 67, 68, 71]. With those assumptions, we have the following theorem.

Theorem 2. *With the received signal model given in Equation* (3.1)*, imperfect channel estimation modeled as Equation* (3.10)*, the* a priori *mean vector* $\boldsymbol{\mu} = [\mu_0, \mu_1, \cdots, \mu_{N_T-1}]^T$ *where* μ_n *is given by Equation* (3.4)*, the* a priori *covariance matrix* $\mathbf{V} = \mathrm{cov}\{\mathbf{x}, \mathbf{x}\} = diag\{\mathbf{v}\}$, $\mathbf{v} = [v_0, v_1, \cdots, v_{N_T-1}]^T$ *where* v_n *is given by Equation* (3.5)*, we have the MMSE estimation of n-th transmitted symbol, i.e.,* $\hat{x}_n$*, given by*

$$\hat{x}_n = (1-\alpha)\, \mathbf{e}_n^H \mathbf{V}\hat{\mathbf{H}}^H \left((1-\alpha)^2 \hat{\mathbf{H}}\mathbf{V}\hat{\mathbf{H}}^H + \sigma_n^2 \mathbf{I}_{N_R}\right)^{-1} \left(\mathbf{y} - (1-\alpha)\, \hat{\mathbf{H}}\boldsymbol{\mu}\right) + \mu_n, \tag{3.11}$$

where $\alpha = \frac{\sigma_e^2}{\sigma_h^2 + \sigma_e^2}$, $\sigma_n^2 = \sigma_z^2 + \alpha\sigma_h^2 \cdot tr\left(E\left\{\mathbf{x}\mathbf{x}^H\right\}\right)$.

Proof. According to ***Theorem*** 1, we have

$$\mathrm{E}\left\{\Delta\mathbf{h}_n | \hat{\mathbf{h}}_n\right\} = \alpha\hat{\mathbf{h}}_n, \tag{3.12}$$

$$\mathrm{cov}\left\{\Delta\mathbf{h}_n, \Delta\mathbf{h}_n | \hat{\mathbf{h}}_n\right\} = \alpha\sigma_h^2 \mathbf{I}_{N_R}, \tag{3.13}$$

Therefore, we have

$$\mathrm{E}\left\{\Delta\mathbf{h}_n \cdot \Delta\mathbf{h}_n^{\mathrm{H}} | \hat{\mathbf{h}}_n\right\} = \alpha^2 \hat{\mathbf{h}}_n \hat{\mathbf{h}}_n^{\mathrm{H}} + \alpha\sigma_h^2 \mathbf{I}_{N_R}. \tag{3.14}$$

In addition, we derive the conditional $\mathbf{y}$ given the channel estimate matrix $\hat{\mathbf{H}}$ as

$$\mathrm{E}\left\{\mathbf{y} | \hat{\mathbf{H}}\right\} = \sum_{n=0}^{N_T-1} \mathrm{E}\left\{\left(\hat{\mathbf{h}}_n - \Delta\mathbf{h}_n\right) x_n | \hat{\mathbf{h}}_n\right\} = \sum_{n=1}^{N_T} (1-\alpha)\, \hat{\mathbf{h}}_n \mu_n = (1-\alpha)\, \hat{\mathbf{H}}\boldsymbol{\mu}. \tag{3.15}$$

The correlation of the received vector conditioned on channel estimate matrix is obtained through

$$\begin{aligned}
\mathrm{E}\left\{\mathbf{y}\mathbf{y}^{\mathrm{H}} | \hat{\mathbf{H}}\right\} &= E\left\{ \left(\sum_{n=1}^{N_T} \left(\left(\widehat{\mathbf{h}}_n - \Delta\mathbf{h}_n\right) x_n + \mathbf{z} \right) \right) \left(\sum_{m=1}^{N_T} \left(\left(\widehat{\mathbf{h}}_m - \Delta\mathbf{h}_m\right) x_m + \mathbf{z} \right) \right)^{\mathrm{H}} \middle| \hat{\mathbf{H}} \right\} \\
&\overset{(a)}{=} E\left\{\mathbf{z}\mathbf{z}^{\mathrm{H}}\right\} + E\left\{ \left(\sum_{n=1}^{N_T} \left(\widehat{\mathbf{h}}_n - \Delta\mathbf{h}_n\right) x_n \right) \left(\sum_{m=1}^{N_T} \left(\widehat{\mathbf{h}}_m - \Delta\mathbf{h}_m\right) x_m \right)^{\mathrm{H}} \middle| \mathbf{h}_n \right\} \\
&\overset{(b)}{=} \sigma_z^2 \mathbf{I}_{N_R} + E\left\{ \sum_{n=1}^{N_T} \left(\left(\widehat{\mathbf{h}}_n - \Delta\mathbf{h}_n\right) x_n \right) \left(\left(\widehat{\mathbf{h}}_n - \Delta\mathbf{h}_n\right) x_n \right)^{\mathrm{H}} \middle| \widehat{\mathbf{h}}_n \right\} \\
&\overset{(c)}{=} \sigma_z^2 \mathbf{I}_{N_R} + \sum_{n=1}^{N_T} E\left\{x_n x_n^{\mathrm{H}}\right\} \cdot E\left\{ \left(\widehat{\mathbf{h}}_n - \Delta\mathbf{h}_n\right) \left(\widehat{\mathbf{h}}_n - \Delta\mathbf{h}_n\right)^{\mathrm{H}} \middle| \widehat{\mathbf{h}}_n \right\},
\end{aligned} \tag{3.16}$$

where: (a) holds according to **Assumption 1**; (b) holds by using the **Assumption 2** and **4**; (c) holds with the **Assumption 3**.

Using the results in Equation (3.12), (3.13), and (3.14), we have the last component in the RHS of Equation (3.16) given by

$$\begin{aligned} &\mathrm{E}\left\{\left(\hat{\mathbf{h}}_n-\Delta\mathbf{h}_n\right)\left(\hat{\mathbf{h}}_n-\Delta\mathbf{h}_n\right)^{\mathrm{H}}\middle|\,\hat{\mathbf{h}}_n\right\} \\ =&(1-\alpha)^2\hat{\mathbf{h}}_n\hat{\mathbf{h}}_n^{\mathrm{H}}+\alpha\sigma_h^2\mathbf{I}_{N_R}. \end{aligned} \tag{3.17}$$

By substituting (3.17) into (3.16), we have the correlation of the received vector conditioned on channel estimate matrix given by

$$\begin{aligned} &\mathrm{E}\left\{\mathbf{y}\mathbf{y}^{\mathrm{H}}|\hat{\mathbf{H}}\right\} \\ =&\sigma_z^2\mathbf{I}_{N_R}+\sum_{n=1}^{N_T}\mathrm{E}\left\{x_n x_n^*\right\}\left((1-\alpha)^2\hat{\mathbf{h}}_n\hat{\mathbf{h}}_n^{\mathrm{H}}+\alpha\sigma_h^2\mathbf{I}_{N_R}\right) \\ =&(1-\alpha)^2\hat{\mathbf{H}}\mathrm{E}\left\{\mathbf{x}\mathbf{x}^{\mathrm{H}}\right\}\hat{\mathbf{H}}^{\mathrm{H}}+\sigma_n^2\mathbf{I}_{N_R}. \end{aligned} \tag{3.18}$$

With the conditional mean $\mathrm{E}\left\{\mathbf{y}|\hat{\mathbf{H}}\right\}$ and correlation $\mathrm{E}\left\{\mathbf{y}\mathbf{y}^{\mathrm{H}}|\hat{\mathbf{H}}\right\}$, we obtain the conditional covariance matrix $\mathrm{cov}\left\{\mathbf{y},\mathbf{y}\,\middle|\,\hat{\mathbf{H}}\right\}$ given by

$$\begin{aligned} \mathrm{cov}\left\{\mathbf{y},\mathbf{y}\,\middle|\,\hat{\mathbf{H}}\right\} =&\mathrm{E}\left\{\mathbf{y}\mathbf{y}^{\mathrm{H}}\,\middle|\,\hat{\mathbf{H}}\right\}-\mathrm{E}\left\{\mathbf{y}\,\middle|\,\hat{\mathbf{H}}\right\}\cdot\left(\mathrm{E}\left\{\mathbf{y}\,\middle|\,\hat{\mathbf{H}}\right\}\right)^{\mathrm{H}} \\ =&(1-\alpha)^2\hat{\mathbf{H}}\mathbf{V}\hat{\mathbf{H}}^{\mathrm{H}}+\sigma_n^2\mathbf{I}_{N_R}. \end{aligned} \tag{3.19}$$

From the received signal model in Equation (3.1), we write the MMSE estimation, $\hat{x}_n$, as

$$\hat{x}_n=\mathbf{w}_n^{\mathrm{H}}\mathbf{y}+d_n. \tag{3.20}$$

The mean-square error conditioned on channel estimate matrix $\hat{\mathbf{H}}$ is then given by

$$\begin{aligned} J =&\mathrm{E}\left\{\left|\hat{x}_n-x_n\right|^2\middle|\,\hat{\mathbf{H}}\right\} \\ =&\mathbf{w}_n^{\mathrm{H}}\mathrm{E}\left\{\mathbf{y}\mathbf{y}^{\mathrm{H}}\,\middle|\,\hat{\mathbf{H}}\right\}\mathbf{w}_n+2\mathbf{w}_n^{\mathrm{H}}\mathrm{E}\left\{\mathbf{y}\,\middle|\,\hat{\mathbf{H}}\right\}d_n^* \\ &-2\mathbf{w}_n^{\mathrm{H}}\mathrm{E}\left\{\mathbf{y}x_n^*\,\middle|\,\hat{\mathbf{H}}\right\}+\left|d_n\right|^2-2d_n\mu_n^*+\mathrm{E}\left\{x_n x_n^*\right\}. \end{aligned} \tag{3.21}$$

By letting the partial derivative of J with respective to $\mathbf{w}_n$ and d_n be zeros, we have

$$\frac{\partial J}{\partial \mathbf{w}_n} = 2\mathrm{E}\left\{\mathbf{y}\mathbf{y}^{\mathrm{H}}|\hat{\mathbf{H}}\right\}\mathbf{w}_n + 2\mathrm{E}\left\{\mathbf{y}|\hat{\mathbf{H}}\right\}d_n^* - 2\mathrm{E}\left\{\mathbf{y}x_n^*|\hat{\mathbf{H}}\right\} = \mathbf{0}, \tag{3.22}$$

$$\frac{\partial J}{\partial d_n} = 2\mathbf{w}_n^{\mathrm{H}}\mathrm{E}\left\{\mathbf{y}|\hat{\mathbf{H}}\right\} + 2d_n - 2\mu_n = 0. \tag{3.23}$$

The solutions to Equations (3.22) and (3.23) are given by

$$d_n = \mu_n - \mathbf{w}_n^{\mathrm{H}}\mathrm{E}\left\{\mathbf{y}|\hat{\mathbf{H}}\right\}, \tag{3.24}$$

$$\mathbf{w}_n = \left(\mathrm{cov}\left\{\mathbf{y},\mathbf{y}|\hat{\mathbf{H}}\right\}\right)^{-1}\mathrm{cov}\left(\mathbf{y},x_n|\hat{\mathbf{H}}\right). \tag{3.25}$$

Substituting the results in Equations (3.15) and (3.19) to Equation (3.20), we arrive at Equation (3.11), hence completing the proof of ***Theorem* 2**. □

When x_n is under detection, replacing μ_n and v_n in Equation (3.11) with $\mu_n = 0$, $v_n = 1$, we then obtain the new expression of the estimation given by

$$\hat{x}_n = \kappa_n(1-\alpha)\,\mathbf{e}_n^{\mathrm{H}}\hat{\mathbf{H}}^{\mathrm{H}}\left((1-\alpha)^2\hat{\mathbf{H}}\mathbf{V}\hat{\mathbf{H}}^{\mathrm{H}} + \sigma_n^2\mathbf{I}_{N_R}\right)^{-1}\left(\mathbf{y} - (1-\alpha)\,\hat{\mathbf{H}}\mu\right) + \kappa_n(1-\alpha)^2\rho_n\mu_n, \tag{3.26}$$

where

$$\kappa_n = \left(1 + (1-\alpha)^2(1-v_n)\,\rho_n\right)^{-1},$$

$$\rho_n = \mathbf{e}_n^{\mathrm{H}}\hat{\mathbf{H}}^{\mathrm{H}}\left((1-\alpha)^2\hat{\mathbf{H}}\mathbf{V}\hat{\mathbf{H}}^{\mathrm{H}} + \sigma_n^2\mathbf{I}_{N_R}\right)^{-1}\hat{\mathbf{H}}\mathbf{e}_n,$$

3.2.2 LLR Computation

The scalar $\kappa_n(1-\alpha)$ is presented at both the signal and noise components at the right hand side (RHS) of Equation (3.26); therefore, it can be ignored for the derivation of the conditional probability density function $p(x_n|\hat{x}_n)$. Equation (3.26) then becomes

$$\hat{x}_n = \mathbf{e}_n^{\mathrm{H}}\hat{\mathbf{H}}^{\mathrm{H}}\left((1-\alpha)^2\hat{\mathbf{H}}\mathbf{V}\hat{\mathbf{H}}^{\mathrm{H}} + \sigma_n^2\mathbf{I}_{N_R}\right)^{-1}\left(\mathbf{y} - (1-\alpha)\,\hat{\mathbf{H}}\mu\right) + (1-\alpha)\rho_n\mu_n. \tag{3.27}$$

With the common assumption that the output from the equalizer follows Gaussian distribution, we rewrite the MMSE estimation in Equation (3.27) into

$$\hat{x}_n = \beta_n x_n + \zeta_n, \tag{3.28}$$

where the equivalent channel gain, β_n, is given by

$$\beta_n = (1-\alpha)\,\rho_n, \tag{3.29}$$

and the covariance of the residual interference plus noise, i.e., $\hat{v}_n = \mathrm{E}\left\{\zeta_n \zeta_n^{\mathrm{H}}\right\}$, is given by

$$\hat{v}_n = \rho_n \left(1 - (1-\alpha)^2 \rho_n v_n\right). \tag{3.30}$$

Therefore, according to Equation (3.28), the conditional probability density function $p\left(\hat{x}_n | x_n\right)$ is given by

$$p\left(\hat{x}_n \,|x_n = s_i\right) = \frac{1}{\pi \hat{v}_n} \exp\left(-\frac{\left|\hat{x}_n - (1-\alpha)\,\rho_n s_i\right|^2}{\hat{v}_n}\right). \tag{3.31}$$

The conditional probability density function $p\left(\hat{x}_n | x_n = s_i\right)$ will be used in Equation (3.7) for the computation of the *a posteriori* LLR of the associated bits.

3.2.3 Discussion

1) A Priori Information is Unavailable

In this case, the *a priori* LLR of the associated bits is set to 0; hence, the *a priori* mean vector $\boldsymbol{\mu} = \mathbf{0}$, and the covariance matrix $\mathbf{V} = \mathbf{I}_{N_T}$. With the zero mean vector and the unit covariance matrix in Equation (3.11), we arrive at

$$\hat{x}_n = (1-\alpha)\, e_n^{\mathrm{H}} \hat{\mathbf{H}}^{\mathrm{H}} \left((1-\alpha)^2 \hat{\mathbf{H}} \hat{\mathbf{H}}^{\mathrm{H}} + \left(\sigma_z^2 + \alpha N_T \sigma_h^2\right) \mathbf{I}_{N_R}\right)^{-1} \mathbf{y}. \tag{3.32}$$

The expression in Equation (3.32) is the same as the linear MMSE detection scheme conditioned on channel estimate derived in [68], and the MMSE-based compensation scheme in [71]. That is to say, their results can be seen as a special case in our scheme.

2) Turbo Equation with Perfect Channel Knowledge

In this case, $\alpha = 0$, and the expression in Equation (3.11) reduces to

$$\hat{x}_n = \mathbf{e}_n^{\mathrm{H}}\mathbf{V}\mathbf{H}^{\mathrm{H}}\left(\mathbf{H}\mathbf{V}\mathbf{H}^{\mathrm{H}} + \sigma_z^2\mathbf{I}_{N_R}\right)^{-1}\left(\mathbf{y} - \mathbf{H}\boldsymbol{\mu}\right) + \mu_n, \tag{3.33}$$

which is the widely used MMSE-based turbo equalization with perfect channel estimation [39, 41].

3) Turbo Equalization under Imperfect Channel with Small Channel Estimation Error

When the channel estimation error is much smaller compared to the real channel data, i.e., $\sigma_e^2 \ll \sigma_h^2$, we derive $\alpha \approx 0$, $\alpha\sigma_h^2 \approx \sigma_e^2$. Therefore, Equation (3.11) becomes

$$\hat{x}_n = \mathbf{e}_n^{\mathrm{H}}\mathbf{V}\hat{\mathbf{H}}^{\mathrm{H}}\left(\hat{\mathbf{H}}\mathbf{V}\hat{\mathbf{H}}^{\mathrm{H}} + \left(\sigma_z^2 + \sigma_e^2 \cdot \mathrm{tr}\left(\mathrm{E}\left\{\mathbf{x}\mathbf{x}^{\mathrm{H}}\right\}\right)\right)\mathbf{I}_{N_R}\right)^{-1}\left(\mathbf{y} - \hat{\mathbf{H}}\mu\right) + \mu_n, \tag{3.34}$$

which is the expression derived in [21, 39].

It is worth noting that in the case of $\sigma_e^2 \ll \sigma_h^2$, from Equation (3.12) and (3.13), we have $\mathrm{E}\left\{\Delta\mathbf{h}_n|\hat{\mathbf{h}}_n\right\} \approx \mathbf{0}$ and $\mathrm{cov}\left\{\Delta\mathbf{h}_n, \Delta\mathbf{h}_n|\hat{\mathbf{h}}_n\right\} = \sigma_e^2\mathbf{I}_{N_R}$. Alternatively, we can approximately view that the channel estimation error is independent from the channel estimate matrix, which is the assumption used in [21,39]. Therefore, we release the conditions used in [21,39] and develop the MMSE estimation conditioned on channel estimate in a general channel estimation error condition.

4) Conventional MMSE-based Turbo Equalization

In this case, by setting $\alpha = 0$, we have

$$\hat{x}_n = \mathbf{e}_n^{\mathrm{H}}\mathbf{V}\hat{\mathbf{H}}^{\mathrm{H}}\left(\hat{\mathbf{H}}\mathbf{V}\hat{\mathbf{H}}^{\mathrm{H}} + \sigma_z^2\mathbf{I}_{N_R}\right)^{-1}\left(\mathbf{y} - \hat{\mathbf{H}}\mu\right) + \mu_n, \tag{3.35}$$

where we can see that the channel estimate is directly applied to the equalization process.

Therefore, we can see the derivation provided in Equation (3.11) is a general expression where existing MMSE-based detection schemes can be seen as special cases.

3.3 SQRD-Based Turbo Equalization Conditioned on Channel Estimate

3.3.1 The Re-development of SQRD-based Turbo Equalization

We rewrite the MMSE estimation in Equation (3.11) in matrix format, given by

$$\hat{\mathbf{x}} = (1-\alpha)\,\mathbf{V}\hat{\mathbf{H}}^{\mathrm{H}}\left((1-\alpha)^2\hat{\mathbf{H}}\mathbf{V}\hat{\mathbf{H}}^{\mathrm{H}} + \sigma_n^2\mathbf{I}_{N_R}\right)^{-1}\left(\mathbf{y} - (1-\alpha)\,\hat{\mathbf{H}}\boldsymbol{\mu}\right) + \boldsymbol{\mu}. \tag{3.36}$$

Based on Equation (3.36), we define a new received vector as

$$\hat{\mathbf{y}} = (1-\alpha)\,\mathbf{V}\hat{\mathbf{H}}^{\mathrm{H}}\left((1-\alpha)^2\hat{\mathbf{H}}\mathbf{V}\hat{\mathbf{H}}^{\mathrm{H}} + \sigma_n^2\mathbf{I}_{N_R}\right)^{-1}\mathbf{y}.$$

We then rewrite $\hat{\mathbf{y}}$ as [2]

$$\begin{aligned}\hat{\mathbf{y}} &= (1-\alpha)\,\mathbf{V}^{\frac{1}{2}}\left((1-\alpha)^2\mathbf{V}^{\frac{1}{2}}\hat{\mathbf{H}}^{\mathrm{H}}\hat{\mathbf{H}}\mathbf{V}^{\frac{1}{2}} + \sigma_n^2\mathbf{I}_{N_T}\right)^{-1}\mathbf{V}^{\frac{1}{2}}\hat{\mathbf{H}}^{\mathrm{H}}\mathbf{y} \\ &= (1-\alpha)\left((1-\alpha)^2\hat{\mathbf{H}}^{\mathrm{H}}\hat{\mathbf{H}} + \sigma_n^2\mathbf{V}^{-1}\right)^{-1}\hat{\mathbf{H}}^{\mathrm{H}}\mathbf{y}.\end{aligned} \tag{3.37}$$

Define the new extended channel matrix as

$$\underline{\mathbf{H}} = \begin{bmatrix} (1-\alpha)\,\hat{\mathbf{H}} \\ \sigma_n\mathbf{V}^{-\frac{1}{2}} \end{bmatrix},$$

we can rewrite the new received vector $\hat{\mathbf{y}}$ as

$$\hat{\mathbf{y}} = (1-\alpha)\left(\underline{\mathbf{H}}^{\mathrm{H}}\underline{\mathbf{H}}\right)^{-1}\hat{\mathbf{H}}^{\mathrm{H}}\mathbf{y}. \tag{3.38}$$

It is worth noting that the format of Equation (3.38) is typically introduced in SQRD-based equalization, for example, [22, 39, 41, 62]. Performing QR decomposition on the new extended channel matrix $\underline{\mathbf{H}}$, i.e.,

$$\underline{\mathbf{H}} = \mathbf{QR} = \begin{bmatrix} \mathbf{P} \\ \\ \mathbf{T} \end{bmatrix}\mathbf{R}, \tag{3.39}$$

[2] Using $\mathbf{G}^{\mathrm{H}}\left(\mathbf{G}\mathbf{G}^{\mathrm{H}} + \sigma_n^2\mathbf{I}\right)^{-1} = \left(\mathbf{G}^{\mathrm{H}}\mathbf{G} + \sigma_n^2\mathbf{I}\right)^{-1}\mathbf{G}^{\mathrm{H}}$, and letting $\mathbf{G} = (1-\alpha)\,\hat{\mathbf{H}}\mathbf{V}^{\frac{1}{2}}$, the derivation is obtained

we have the following governing conditions satisfied:

$$\mathbf{PR} = (1-\alpha)\,\hat{\mathbf{H}}, \tag{3.40}$$

$$\mathbf{TR} = \sigma_n \mathbf{V}^{-\frac{1}{2}}, \tag{3.41}$$

$$\mathbf{P}^{\mathrm{H}}\mathbf{P} + \mathbf{T}^{\mathrm{H}}\mathbf{T} = \mathbf{I}_{N_T}, \tag{3.42}$$

$$(1-\alpha)\,\mathbf{P}^{\mathrm{H}}\hat{\mathbf{H}} + \sigma_n \mathbf{T}^{\mathrm{H}}\mathbf{V}^{-\frac{1}{2}} = \mathbf{R}. \tag{3.43}$$

$\mathbf{R}$ and $\mathbf{T}$ are upper and triangular matrices, respectively. Let $\widetilde{\mathbf{y}} = \mathbf{P}^{\mathrm{H}}\mathbf{y}$, we have

$$\begin{aligned}\widetilde{\mathbf{y}} =& \mathbf{P}^{\mathrm{H}}\hat{\mathbf{H}}\mathbf{x} - \mathbf{P}^{\mathrm{H}}\Delta\mathbf{H}\mathbf{x} + \mathbf{P}^{\mathrm{H}}\mathbf{z} \\ =& \mathbf{R}\mathbf{x} - \sigma_n \mathbf{T}^{\mathrm{H}}\mathbf{V}^{-\frac{1}{2}}\mathbf{x} + \alpha\mathbf{P}^{\mathrm{H}}\hat{\mathbf{H}}\mathbf{x} - \mathbf{P}^{\mathrm{H}}\Delta\mathbf{H}\mathbf{x} + \mathbf{P}^{\mathrm{H}}\mathbf{z},\end{aligned} \tag{3.44}$$

where Equation (3.43) is used for the above derivation.

As $\mathbf{R}$ is an upper triangular matrix, we introduce a back-forward method for data detection. Specially, we extract $\widetilde{y}_n = \mathbf{e}_n^{\mathrm{H}}\widetilde{\mathbf{y}}$ from Equation (3.44), i.e.,

$$\begin{aligned}\widetilde{y}_n =& \mathbf{e}_n^{\mathrm{H}}\left(\mathbf{R} - \sigma_n \mathbf{T}^{\mathrm{H}}\mathbf{V}^{-\frac{1}{2}}\right)\mathbf{e}_n x_n + \sum_{l=n+1}^{N_T-1}\mathbf{R}_{n,l}x_l \\ &- \sigma_n \sum_{l=0}^{n-1}\underline{\mathbf{T}}_{n,l}x_l + \alpha\mathbf{e}_n^{\mathrm{H}}\mathbf{P}^{\mathrm{H}}\hat{\mathbf{H}}\mathbf{x} - \mathbf{e}_n^{\mathrm{H}}\mathbf{P}^{\mathrm{H}}\Delta\mathbf{H}\mathbf{x} + \mathbf{e}_n^{\mathrm{H}}\mathbf{P}^{\mathrm{H}}\mathbf{z}.\end{aligned} \tag{3.45}$$

$\widetilde{y}_n$ will be used for the estimation of x_n. The RHS of Equation (3.45) contains six components: the first component denotes the expected signal component; the second component contributes to the interference from the previous detected data symbols; the third component indicates the interference from undetected data symbols; the fourth and fifth components are the interference introduced by imperfect channel estimation; the last component is attributed to the noise.

We compute the conditional mean $\widetilde{\mu}_n = E\left\{\widetilde{y}_n | x_n, \hat{\mathbf{H}}\right\}$ and covariance $\widetilde{v}_n = \mathrm{cov}\left\{\hat{y}_n, \hat{y}_n | x_n, \hat{\mathbf{H}}\right\}$, given by

$$\widetilde{\mu}_n = \mathbf{e}_n^{\mathrm{H}}\left(\mathbf{R} - \sigma_n \mathbf{T}^{\mathrm{H}}\mathbf{V}^{-\frac{1}{2}}\right)\mathbf{e}_n x_n + \sum_{l=n+1}^{N_T-1}\mathbf{R}_{n,l}\mu_l^{\mathrm{P}} - \sigma_n \sum_{l=0}^{n-1}\underline{\mathbf{T}}_{n,l}\mu_l, \tag{3.46}$$

$$
\begin{aligned}
\widetilde{v}_n =& (1-\alpha)^2 \mathbf{e}_n^{\mathrm{H}} \mathbf{P}^{\mathrm{H}} \hat{\mathbf{H}} \left(\mathbf{V} - v_n \mathbf{e}_n \mathbf{e}_n^{\mathrm{H}} \right) \hat{\mathbf{H}}^{\mathrm{H}} \mathbf{P} \mathbf{e}_n + \sigma_n^2 \mathbf{e}_n^{\mathrm{H}} \mathbf{P}^{\mathrm{H}} \mathbf{P} \mathbf{e}_n \\
=& \mathbf{e}_n^{\mathrm{H}} \left(\mathbf{R} - \sigma_n \mathbf{T}^{\mathrm{H}} \mathbf{V}^{-\frac{1}{2}} \right) \left(\mathbf{V} - v_n \mathbf{e}_n \mathbf{e}_n^{\mathrm{H}} \right) \cdot \left(\mathbf{R} - \sigma_n \mathbf{T}^{\mathrm{H}} \mathbf{V}^{-\frac{1}{2}} \right)^{\mathrm{H}} \mathbf{e}_n + \sigma_n^2 \mathbf{e}_n^{\mathrm{H}} \mathbf{P}^{\mathrm{H}} \mathbf{P} \mathbf{e}_n \\
=& \mathbf{e}_n^{\mathrm{H}} \mathbf{R} \left(\mathbf{V} - v_n \mathbf{e}_n \mathbf{e}_n^{\mathrm{H}} \right) \mathbf{R}^{\mathrm{H}} \mathbf{e}_n + \sigma_n^2 \mathbf{e}_n^{\mathrm{H}} \mathbf{P}^{\mathrm{H}} \mathbf{P} \mathbf{e}_n + \sigma_n^2 \mathbf{e}_n^{\mathrm{H}} \mathbf{T}^{\mathrm{H}} \mathbf{V}^{-\frac{1}{2}} \left(\mathbf{V} - v_n \mathbf{e}_n \mathbf{e}_n^{\mathrm{H}} \right) \mathbf{V}^{-\frac{1}{2}} \mathbf{T} \mathbf{e}_n \\
=& \sum_{l=n+1}^{N_T-1} \left| \mathbf{R}_{n,l} \right|^2 v_l^{\mathrm{P}} + \sigma_n^2 - \sigma_n^2 \mathbf{e}_n^{\mathrm{H}} \mathbf{T}^{\mathrm{H}} \mathbf{T} \mathbf{e}_n \\
=& \sum_{l=n+1}^{N_T-1} \left| \mathbf{R}_{n,l} \right|^2 v_l^{\mathrm{P}} + \sigma_n^2 - \sigma_n^2 |\mathbf{T}_{n,n}|^2,
\end{aligned}
\tag{3.47}
$$

where μ_l^{P} and v_l^{P} denote the *a posteriori* mean and covariance of the already detected data symbols. From Equation (3.45), we perform soft interference cancelation, i.e., using the *a priori* information from channel decoder for the undetected data symbols, and using the *a posteriori* information from the previous detected data symbols, leading to

$$
\breve{y}_n = \widetilde{y}_n - \sum_{l=n+1}^{N_T-1} \mathbf{R}_{n,l} \mu_l^{\mathrm{P}} + \sigma_n \sum_{l=0}^{n-1} \underline{\mathbf{T}}_{n,l} \mu_l, \tag{3.48}
$$

and we can easily derive the conditional mean $\breve{\mu}_n = E\left\{ \breve{y}_n \middle| x_n, \hat{\mathbf{H}} \right\}$ and covariance $\breve{v}_n = \mathrm{cov}\left\{ \breve{y}_n, \breve{y}_n | x_n, \hat{\mathbf{H}} \right\}$, given by

$$
\breve{\mu}_n = \mathbf{e}_n^{\mathrm{H}} \left(\mathbf{R} - \sigma_n \mathbf{T}^{\mathrm{H}} \mathbf{V} \right) \mathbf{e}_n x_n, \tag{3.49}
$$

$$
\breve{v}_n = \sum_{l=n+1}^{N_T-1} \left| \mathbf{R}_{n,l} \right|^2 v_l^{\mathrm{P}} + \sigma_n^2 - \sigma_n^2 |\mathbf{T}_{n,n}|^2. \tag{3.50}
$$

Based on Equations (3.49) and (3.50), and following the Gaussian output assumption, we derive the conditional probability density function $p\left(\breve{y}_n | x_n = s_i \right)$ given by

$$
p\left(\breve{y}_n | x_n = s_i \right) = \frac{1}{\pi \breve{v}_n} \exp\left(- \frac{\left| \breve{y}_n - \breve{\mu}_n \right|^2}{\breve{v}_n} \right). \tag{3.51}
$$

Replacing $p\left(\hat{x}_n | x_n = s_i \right)$ in Equation (3.7) with $p\left(\breve{y}_n | x_n = s_i \right)$, and following the process from Equation (3.7) to (3.9), we derive the *a posteriori* LLR $L^O\left(q_{n,m} \right)$ and the *extrinsic* LLR $L^E\left(q_{n,m} \right)$ of the associated bits. $L^E\left(q_{n,m} \right)$ will be used for the channel decoder.

As we introduced a successive interference cancelation manner for data detection, the *a posteriori* LLR of the associated bits is required to compute the *a posteriori* mean and covariance for later data detection, which is presented in Equations (3.48), (3.49), and (3.50). By collecting the *a posteriori* LLR of all the bits associated with x_n, i.e., $\mathbf{L}_n^O = \left[L^O\left(q_{n,0}\right), L^O\left(q_{n,1}\right), \cdots, L^O\left(q_{n,\log_2 M-1}\right)\right]$, we derive the *a posteriori* probability of x_n given by

$$P\left(x_n = s_i|\mathbf{L}_n^O\right) = \prod_{m=0}^{\log_2 M-1} P\left(q_{n,m} = b_{i,m}|L^O\left(q_{n,m}\right)\right), \tag{3.52}$$

where

$$P\left(q_{n,m} = b|L^O\left(q_{n,m}\right)\right) = \frac{\exp\left((2b-1)\,L^O\left(q_{n,m}\right)/2\right)}{\exp\left(L^O\left(q_{n,m}\right)/2\right) + \exp\left(L^O\left(q_{n,m}\right)/2\right)}.$$

In addition, the *a posteriori* mean μ_n^{P} and covariance v_n^{P} are computed as

$$\mu_n^{\mathrm{P}} = \sum_{\forall s_i, s_i \in \chi} s_i \cdot P\left(x_n = s_i|\mathbf{L}_n^O\right), \tag{3.53}$$

$$v_n^{\mathrm{P}} = \sum_{\forall s_i, s_i \in \chi} |s_i|^2 \cdot P\left(x_n = s_i|\mathbf{L}_n^O\right) - |\mu_n^{\mathrm{P}}|^2. \tag{3.54}$$

μ_n^{P} and v_n^{P} are used in Equation (3.48) for soft interference cancelation of the detected symbols, and Equation (3.50) for computing the conditional covariance $\breve{v}_n$.

We hereby present the SQRD-based turbo equalization algorithm in **Algorithm 3.1**.

3.3.2 Sorted QR Decomposition Algorithm

As the SQRD-based turbo equalization introduces a successive detection manner, we need to obtain the optimal detection order to avoid error propagation issue. That is to say, $|\mathbf{R}_{N_T-1,N_T-1}|$ is maximized, followed by the maximization of $|\mathbf{R}_{N_T-2,N_T-2}|$, and so on. However, in the Gram-Schmidt procedure of the QR decomposition algorithm, the computation of the diagonal elements of $\mathbf{R}$ is in an opposite order, from $|\mathbf{R}_{0,0}|$ to $|\mathbf{R}_{N_T-1,N_T-1}|$. The alternative solution to this difficulty is to minimize $|\mathbf{R}_{k,k}|$ in the order it is computed. This is

Table 3.1: **Algorithm 3.1:** SQRD-based Turbo Equalization

Input: Channel estimate $\hat{\mathbf{H}}$, channel estimation error σ_e^2, noise covariance σ_z^2, the *a priori* LLRs of the associated bits $L^{\mathrm{I}}(q_{n,m})$;

Output: The *extrinsic* LLRs of the associated bits $L^{\mathrm{E}}(q_{n,m})$.

Initialization:

1. Compute the *a priori* mean vector $\boldsymbol{\mu}$, and the *a priori* covariance matrix $\mathbf{V}$ through Equations (3.2) to (3.5);
2. Compute $\alpha = \frac{\sigma_e^2}{\sigma_h^2+\sigma_e^2}$, $\sigma_n^2 = \sigma_z^2 + \alpha\sigma_h^2 \cdot tr\left(\mathrm{E}\left\{\mathbf{x}\mathbf{x}^{\mathrm{H}}\right\}\right)$, and construct the new extended channel matrix $\underline{\mathbf{H}}$;
3. Perform sorted QR decomposition on $\underline{\mathbf{H}}$ to obtain $\mathbf{P}$, $\mathbf{T}$, $\mathbf{R}$, and the detection order $\mathbf{p}$;
4. Rearrange $\boldsymbol{\mu}$, $\mathbf{V}$ according to $\mathbf{p}$;

Preprocessing:

5. Compute $\widetilde{\mathbf{y}} = \mathbf{P}^{\mathrm{H}}\mathbf{y}$, and obtain $\widetilde{y}_n$;

Successive Interference Cancelation:

6. for $n = N_T - 1 : -1 : 0$;
7. Compute $\check{y}_n$ according to Equation (3.48);
8. Compute the conditional probability density function $p\left(\check{y}_n | x_n = s_i\right)$ based on Equation (3.51);

   LLR Computation:

9. Compute the *a posteriori* and *extrinsic* LLRs, i.e., $L^{\mathrm{O}}(q_{n,m})$ and $L^{\mathrm{E}}(q_{n,m})$ through Equations (3.7) to (3.9);
10. Compute the *a posteriori* mean μ_n^{P} and covariance v_n^{P} according to Equations (3.53) and (3.54).
11. end

Return $L^{\mathrm{E}}(q_{n,m})$.

motivated by the fact that the data symbol detected latter affects fewer other users, hence a small SINR is expected. Therefore, the sorted QR decomposition algorithm is summarized in **Algorithm 3.2**.

3.3.3 Schemes Generalization and Discussion

1) SQRD-based Turbo Detection under Perfect Channel Estimation

With perfect channel knowledge, i.e., $\sigma_e^2 = 0$ and $\alpha = 0$, the extended channel matrix is equivalent to that derived in [41,60]. In addition, if the *a priori* information is unavailable, i.e., $\mathbf{V} = \mathbf{I}_{N_T}$, then the extended channel matrix is reduced to the one used in [62] where SQRD based detection scheme is first introduced. Therefore, the ideal cases discussed in [41,60] and [62] can be seen as special cases in our proposition.

2) SQRD-based Turbo Detection under Imperfect Channel with Small Channel Estimation Error

In this case, the expression in Equation (3.35) is reduced to the expression used in [39] where the assumption of the independence of the channel estimation error and the channel estimate is used. Therefore, the SQRD-based turbo equalization scheme studied in [39] can be seen as a special case as well in our proposition.

3)Conventional SQRD-based Turbo Detection under Imperfect Channel Estimation

In this case, the channel estimate is directly applied to SQRD-based detection process. This is equivalent to set $\alpha = 0$ but the channel estimate error is present.

To summarize, we can see the derived SQRD-based turbo equalization conditioned on channel estimate is a general expression for SQRD-based detection scheme and existing works in [39,41,62] are special cases.

Table 3.2: **Algorithm 3.2:** Sorted QR Decomposition

Input: The extended channel matrix $\underline{\mathbf{H}}$, N_R and N_T;

Output: The unity matrix $\mathbf{Q} = \begin{bmatrix} \mathbf{P} \\ \\ \mathbf{T} \end{bmatrix}$, upper triangular matrix $\mathbf{R}$ and detection order $\mathbf{p}$.

Initialization:

1. $\mathbf{Q} = \underline{\mathbf{H}}$, $\mathbf{R} = \mathbf{0}$, and $\mathbf{p} = [0, 1, \cdots, N_T - 1]^{\mathrm{T}}$;
2. for $n = 0 : 1 : (N_T - 1)$
3. Compute the norm of the n-th column vector in $\mathbf{Q}$ as $\mathbf{Norm}_n = \mathbf{e}_n^{\mathrm{H}}\mathbf{Q}^{\mathrm{H}}\mathbf{Q}\mathbf{e}_n$;
4. end;

Gram-Schmidt Procedure:

5. for $n = 0 : 1 : (N_T - 1)$
6. find $k = \arg\min_{m=n+1,\cdots,N_T-1}\mathbf{Norm}_m$;
7. Exchange the n-th and k-th columns in $\mathbf{R}$, $\mathbf{p}$, and $\mathbf{Norm}$, and the first $N_R + n$ rows in $\mathbf{Q}$, and exchange the elements $\mathbf{Q}_{N_R+n,n}$ and $\mathbf{Q}_{N_R+k,k}$;
8. Compute $\mathbf{R}_{n,n} = \sqrt{\mathbf{Norm}_n}$, and update the n-th column in $\mathbf{Q}$ as $\mathbf{Q}_{:,n} = \mathbf{Q}_{:,n}/\mathbf{R}_{n,n}$;
9. for $m = n + 1 : 1 : (N_T - 1)$
10. Compute the $\mathbf{R}_{n,m} = \mathbf{e}_n^{\mathrm{H}}\mathbf{Q}^{\mathrm{H}}\mathbf{Q}\mathbf{e}_m$, and update the m-th column vector in $\mathbf{Q}$ as $\mathbf{Q}_{:,m} = \mathbf{Q}_{:,m} - \mathbf{R}_{n,m}\mathbf{Q}_{:,n}$, and $\mathbf{Norm}_m = \mathbf{Norm}_m - |\mathbf{R}_{n,m}|^2$;
11. end
12. end

Return $\mathbf{P}$, $\mathbf{T}$, $\mathbf{R}$, and $\mathbf{p}$.

3.3.4 Computational Complexity Analysis

We use the number of floating-point operations (flops) to evaluate the computational complexity [41]. The main computations include the computations in SQRD algorithm in **Algorithm 2** and the turbo equalization process in **Algorithm 1**. From **Algorithm 2**, we can derive the complexity of the SQRD algorithm is in the order of $O\left(N_T^3 + N_T^2 N_R\right)$. For SQRD-based turbo equalization, the preprocessing requires $O\left(N_T N_R\right)$ flops. For the estimation of each transmitted symbol in the turbo equalization, it mainly includes soft interference cancelation (which requires $O\left(N_T N_R\right)$ flops), computing the LLRs (which consumes $O\left(M 2^M\right)$ flops). Therefore, the overall computational complexity of the SQRD-based turbo equalization is in the order of $O\left(N_T^3 + N_T^2 N_R + N_T M 2^M\right)$. Compared to complexity of the MMSE-based turbo equalization which is in the order of $O\left(N_T^2 N_R + N_T N_R^3 + N_T M 2^M\right)$, we can see the SQRD-based turbo equalization conditioned on channel estimate has lower complexity when $N_T \leq N_R$.

3.4 Performance Evaluation and Discussions

We use Monte-Carlo simulations to study the performance of the proposed MMSE-based turbo equalization with imperfect channel estimation. We define $\lambda = \sigma_e^2 / \sigma_n^2$ to describe the channel estimation error level, and let λ = 0, 0.1, 0.2, and 0.5. λ= 0 indicates the ideal case where perfect channel knowledge is available at the receiver. The other values from λ= 0.1 to λ= 0.5 indicate the channel estimation error becoming increasingly severe. A 4×4 MIMO system and Turbo code with the code rate of 1/2 are considered. To facilitate the comparison, we denotes the mentioned schemes discussed in previous Sections as follows.

- **Scheme 1: Conventional MMSE-based Turbo Equalization.** This is the case where the channel estimate is directly applied to the equalization process. In this

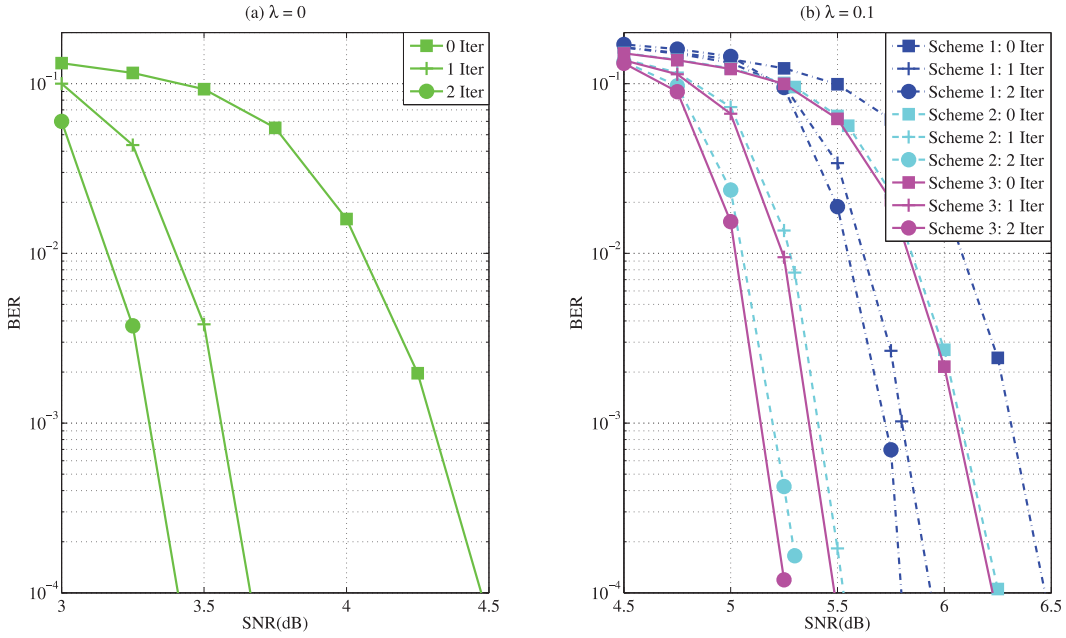


Figure 3.2: BER performance of the MMSE-based turbo equalization schemes, 1/2 Turbo code, QPSK, 4×4 MIMO: $\lambda = 0$ and $\lambda = 0.1$.

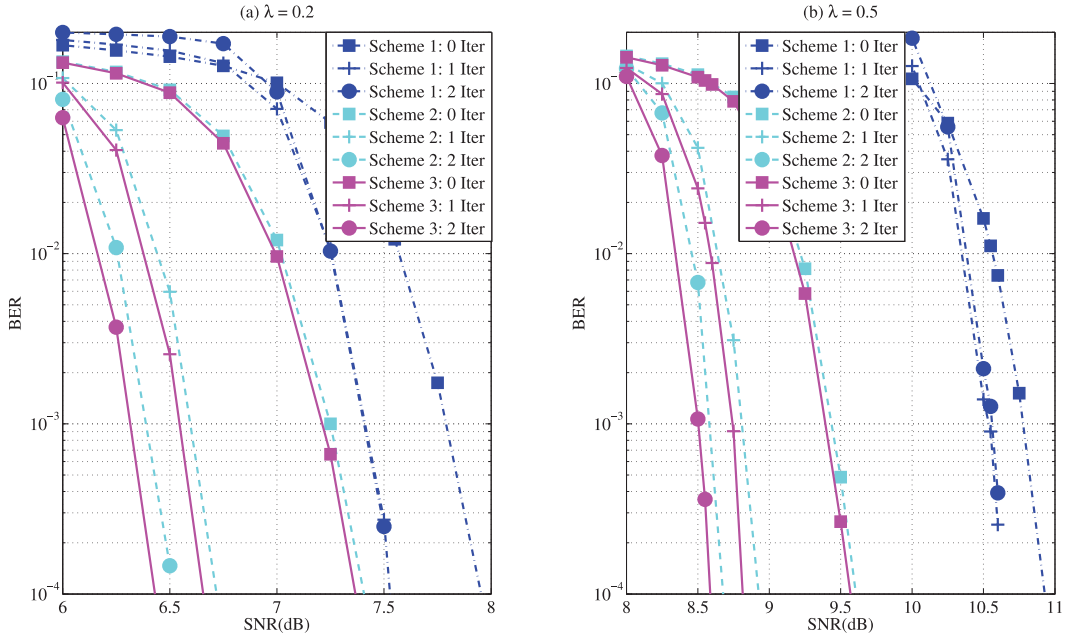


Figure 3.3: BER performance of the MMSE-based turbo equalization schemes, 1/2 Turbo code, QPSK, 4×4 MIMO: $\lambda = 0.2$ and $\lambda = 0.5$.

case, $\alpha = 0$, and $\sigma_n^2 = \sigma_z^2$;

- **Scheme 2: MMSE-based Turbo Equalization with Channel Estimation Error Independent From Channel Estimate.** This corresponds to the turbo equalization scheme in [21, 39]. In this scheme, $\alpha = 0$, and $\sigma_n^2 = \sigma_z^2 + \sigma_e^2 \cdot \mathrm{tr}\left(\mathrm{E}\left\{\mathbf{x}\mathbf{x}^{\mathrm{H}}\right\}\right)$. We have shown that when channel estimation error is sufficiently low, this scheme performs almost the same as the proposed MMSE-based turbo equalization conditioned on channel estimate;

- **Scheme 3: Proposed MMSE-based Turbo Equalization Conditioned on Channel Estimate.** This corresponds the derived general MMSE-based turbo equalization conditioned on channel estimate. In this case, $\alpha = \frac{\sigma_e^2}{\sigma_h^2 + \sigma_e^2}$, $\sigma_n^2 = \sigma_z^2 + \alpha\sigma_h^2 \cdot \mathrm{tr}\left(\mathrm{E}\left\{\mathbf{x}\mathbf{x}^{\mathrm{H}}\right\}\right)$;

- **Scheme 4: Conventional MMSE-SQRD Based Turbo Equalization.** In this scheme, the channel estimate is directly applied to the SQRD-based turbo equalization even if channel estimation error is presented. In this case, $\alpha = 0$, and $\sigma_n^2 = \sigma_z^2$, are used for the development;

- **Scheme 5: MMSE-SQRD Based Turbo Equalization with Channel Estimation Error Independent From Channel Estimate.** This corresponds to the turbo equalization scheme in [39]. In this case, $\alpha = 0$, and $\sigma_n^2 = \sigma_z^2 + \sigma_e^2 \cdot \mathrm{tr}\left(\mathrm{E}\left\{\mathbf{x}\mathbf{x}^{\mathrm{H}}\right\}\right)$, are used in the development. When channel estimation error is sufficiently low, this scheme performs almost the same as the proposed MMSE-SQRD based turbo equalization conditioned on channel estimate as we show in Section 3.3.3;

- **Scheme 6: MMSE-SQRD Based Turbo Equalization Conditioned on Channel Estimation.** This corresponds to the proposed general expression of the MMSE-SQRD based turbo equalization scheme where $\alpha = \frac{\sigma_e^2}{\sigma_h^2 + \sigma_e^2}$, $\sigma_n^2 = \sigma_z^2 + \alpha\sigma_h^2 \cdot \mathrm{tr}\left(\mathrm{E}\left\{\mathbf{x}\mathbf{x}^{\mathrm{H}}\right\}\right)$.

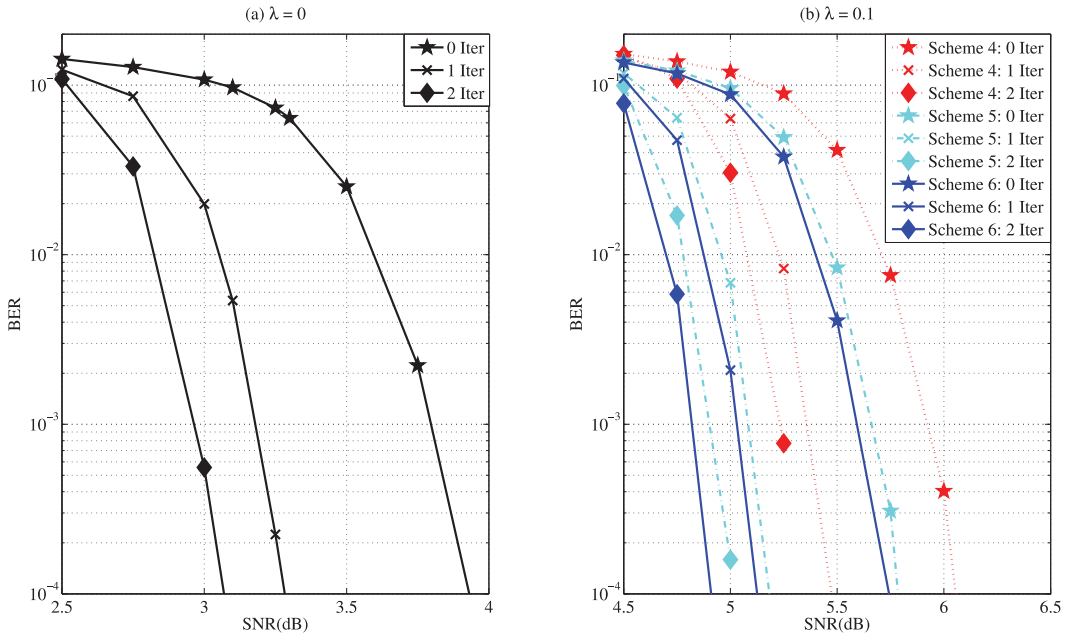


Figure 3.4: BER performance of the MMSE-SQRD based turbo equalization schemes, 1/2 Turbo code, QPSK, 4×4 MIMO: $\lambda = 0$ and $\lambda = 0.1$.

It is worth pointing out that when $\lambda = 0$, i.e., perfect channel estimation is available, **Scheme 1** and **2** are equivalent to **Scheme 3**, and **Scheme 4** and **5** are equivalent to **Scheme 6**. This is because $\alpha = 0$ in this ideal case. Besides, the initial estimation of all schemes degrade to the linear detection schemes.

Figure 3.2 and 3.3 present the BER performance comparison of the discussed MMSE-based turbo equalization schemes where QPSK modulation is adopted in the evaluation. In Figure 3.2(a), we present the results with perfect channel estimation as a benchmark for comparison. As we can see in Figure 3.2 and 3.3, when channel estimation error becomes increasingly severe, the system BER performance of all schemes has shown much obvious degradation. In addition, the performance of the **Scheme 1** becomes worse when the channel estimation error increases, while **Scheme 2** and **Scheme 3** have shown much enhanced system BER performance comparatively. These results reveal that convention-

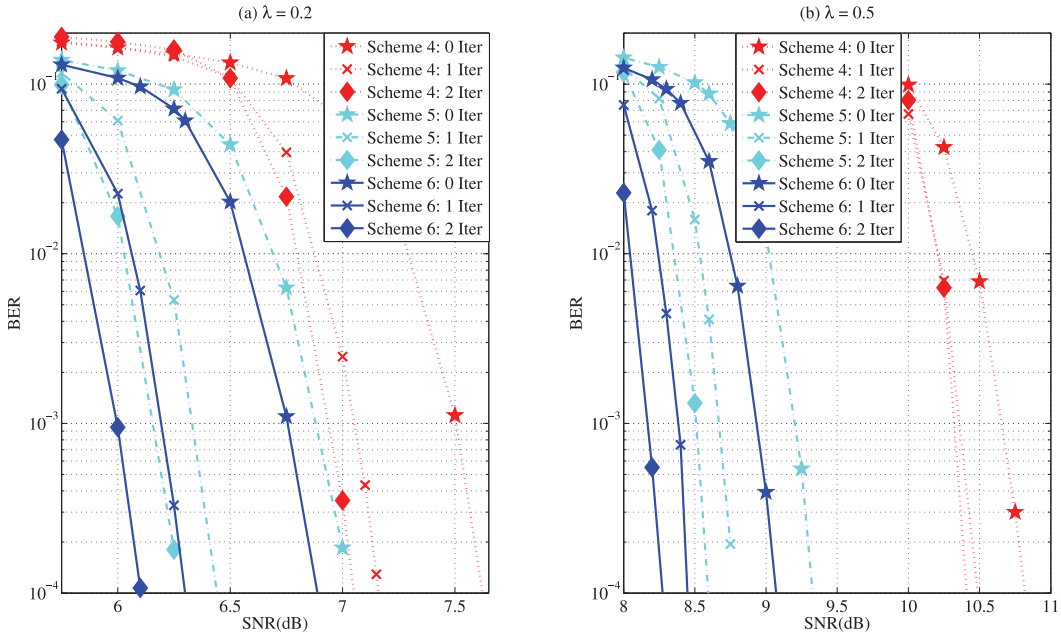


Figure 3.5: BER performance of the MMSE-SQRD based turbo equalization schemes, 1/2 Turbo code, QPSK, 4×4 MIMO: $\lambda = 0.2$ and $\lambda = 0.5$.

al MMSE-based turbo equalization scheme where channel estimate is directly applied to the equalization process performs very bad when channel estimation error is presented. Moreover, in the working SNR region for this 4×4 MIMO system with QPSK modulation adopted, the MMSE-based turbo equalizer can be designed with the assumption that the channel estimation error is independent from the channel estimate.

The BER performance of the discussed MMSE-SQRD based turbo equalization schemes are presented in Figure 3.4 and 3.5, where QPSK modulation is adopted. Again, with the increase of channel estimation error, BER performance degradation appears in all turbo detection schemes. Specifically, we note that when channel estimation error is severe, for example, $\lambda = 0.5$, **Scheme 4**, i.e., the conventional MMSE-SQRD based turbo equalization scheme performs worst, while **Scheme 5** and **Scheme 6** gain about 2.4dB and 2.7dB in the initial iteration at the BER level 10^{-4}, respectively, and these corresponding gains are

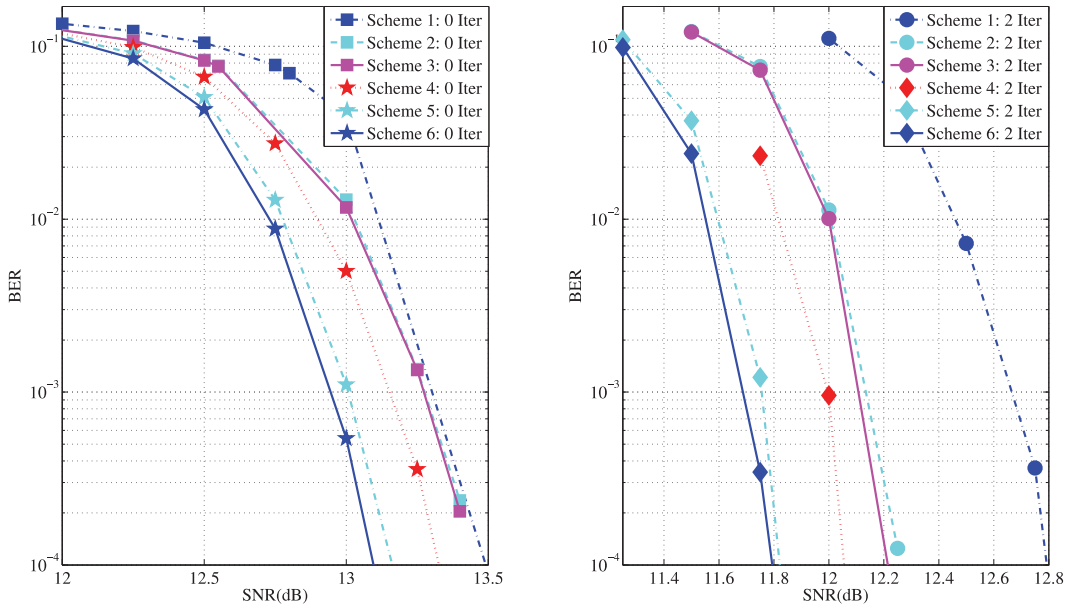


Figure 3.6: BER performance comparison, 1/2 Turbo code, 16QAM, 4×4 MIMO: $\lambda = 0.1$

increased to 2.9dB and 3.2dB after 2 iterations, respectively. Besides, It is worth noting that in this particular 4×4 MIMO system with QPSK modulation adopted, **Scheme 6** shows much improved BER performance compared to **Scheme 5**. This is a little different from the results of the MMSE-based turbo equalization schemes. The difference can be attributed to the SIC introduced in MMSE-SQRD based turbo equalization. Although the development of the MMSE-SQRD based turbo equalization scheme is based on the expression provided by MMSE-based turbo equalization, SIC can enlarge the gain of the previous detection. As a result, latter detected symbols will benefit from the detection gain obtained in previous symbol detection.

In Figure 3.6, we show the BER performance of all discussed detection schemes when $\lambda = 0.1$ and 16QAM modulation is adopted. From 3.6, we can see when channel estimation error is not severe, the linear detection schemes perform almost the same, within 0.5dB at BER level 10^{-4}. However, with iterations, we can see that:

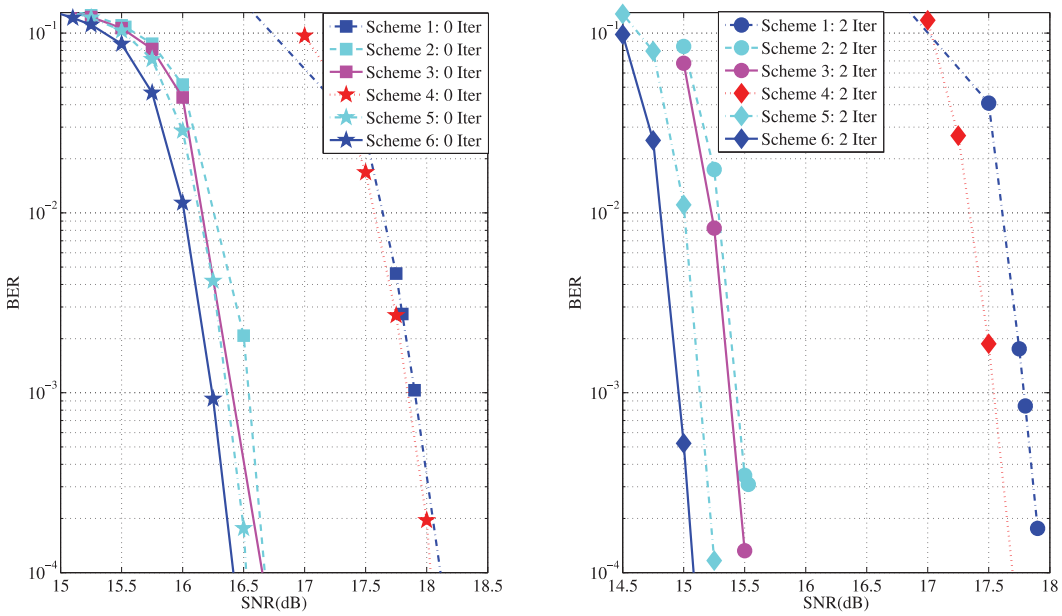


Figure 3.7: BER performance comparison, 1/2 Turbo code, 16QAM, 4×4 MIMO: $\lambda = 0.5$

- The BER performance of **Scheme 1** performs the worst at all iterations;
- The performance of the **Scheme 2** and **Scheme 3** are almost the same; the performance of the **Scheme 5** and **Scheme 6** are very close;
- The MMSE-SQRD based turbo equalization schemes have shown much better performance after two iterations. To be specific, compared to the **Scheme 1**, the SNR reduction at BER level of 10^{-4} for **Scheme 4**, **Scheme 5**, and **Scheme 6**, are about 0.75dB, 0.98dB, and 1dB, respectively.

These results indicate that in this case and working SNR region, the MMSE-based and MMSE-SQRD based turbo equalization schemes will degrade to the conventional MMSE-based and MMSE-SQRD based turbo equalization schemes.

With comparatively severe channel estimation error conditions where $\lambda = 0.5$, the BER performance of the 16QAM modulated symbols in a 4×4 MIMO system is shown in Figure 3.7. In this case, it is clear that when channel estimation error becomes significant,

the working SNR region will be high. However, we can see that in this server channel estimation error conditions, both conventional MMSE-based and MMSE-SQRD based turbo equalization schemes have shown significant performance loss compared to the case of perfect channel estimation. However, the design of the turbo equalization schemes which consider channel estimation error in the development can greatly improve the overall system performance comparatively. Specifically, compared to the performance of the **Scheme 1**, at the BER level of 10^{-4}, **Scheme 2** and **Scheme 3** achieve about 1.5dB gain in the linear detection, whereas **Scheme 5** and **Scheme 6** obtain 1.7dB. With iterations, the corresponding gains for **Scheme 2** and **Scheme 3** increase to 2.5dB, and **Scheme 5** and **Scheme 6** achieves about 2.75dB and 2.9dB SNR reduction, respectively. It is worth noting the performance gain of the linear MMSE-based detection scheme has been revealed in [68, 70, 71]. However, we have shown that with iterations, we can achieve much better enhanced system BER performance for both the MMSE-based and MMSE-SQRD based turbo equalization schemes.

3.5 Summary

In this chapter, we have investigate the turbo equalization schemes conditioned on the channel estimate. We have shown the derivations of both the proposed MMSE-based and MMSE-SQRD based turbo detection schemes are general expressions, while existing MMSE-based MMSE-SQRD based turbo equalization schemes can be seen as special cases. We then conduct numerical simulations to evaluate the system BER performance in a 4×4 MIMO system with both QPSK and 16QAM modulations adopted. Compared to the perfect channel estimation conditions, we have shown significant performance degradation for all detection schemes when channel estimation error is presented. However, within all those detection schemes, the conventional MMSE-based and MMSE-SQRD based turbo equaliza-

tion schemes have shown significant performance loss. The results clearly indicate that the development of the turbo equalization schemes with channel estimation error in consideration will greatly help with the improvement of the system BER performance. We have also shown that in proper working SNR region, the proposed MMSE-based and MMSE-SQRD based turbo equalization schemes conditioned on channel estimate will behave similarly to the corresponding turbo equalization schemes with the assumption that channel estimation error is independent from channel estimate. However, we also show in MMSE-SQRD based turbo equalization schemes, due to the SIC manner in detection, the proposed MMSE-SQRD based turbo equalization scheme conditioned on channel estimate has shown better system performance compared to the MMSE-based turbo equalization scheme with the assumption that channel estimation error is independent from channel estimate.

Chapter 4

A New SQRD-based Soft Interference Cancelation Scheme in Multi-user MIMO SC-FDMA Systems

[1] In this chapter, we investigate SQRD-based data detection and soft interference cancelation in multiuser multiple-input multiple-output (MIMO) single-carrier frequency division multiple access (SC-FDMA) system. The SC-FDMA system has shown a block transmission property, while the symbol detection is independent over each subcarrier for a OFDM system. We propose a new SQRD-based multi-user detection scheme by utilizing the *a*

[1]The related work has been published in:

- *IEEE Communications Letters*, vol. 21, no. 4, pp. 821–824, April 2017.
- *Proceedings of the 24th IEEE Newfoundland Electrical and Computer Engineering Conference* (NECEC15), St. John's, NL, Canada, November 2015.

priori information from the channel decoder in the decomposition process through the use of a new extended channel matrix. The received signal vector for the current user's data detection can be categorized into the expected signal, the interference, and the noise components, where the interference includes those from both the detected and the undetected users. Soft interference cancelation is then performed on the data. The residual interference from the detected users can be easily obtained due to the use of successive interference cancelation, whereas the interference from the undetected users can be obtained by using the *a priori* information from channel decoder. We further consider on how to extract the expected signal from the interference component, thus further improving the SINR conditions at the receiver. Simulation results show that the proposed scheme significantly improves the BER performance compared to the original SQRD-based multiuser detection scheme and the MMSE-based turbo receivers.

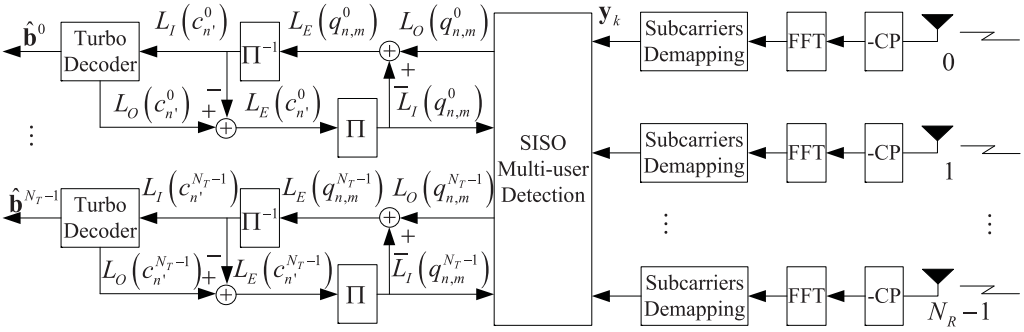
The rest of this chapter is organized as follows. In Section 4.1, we present the system model of the multi-user MIMO SC-FDMA system. Following that, we provide the detail derivation of the proposed new SQRD-based soft interference cancelation scheme in Section 4.2. We then present the numerical simulation results in Section 4.3. Finally, the conclusions are drawn in Section 4.4.

4.1 System Model

We study the uplink transmission for a multi-user SC-FDMA system. The base station (BS) is equipped with N_R receive antennas and serves N_U users, each equipped with a single omni-directional antenna. The block diagram of transmitter and receiver architecture is shown in Figure 4.1.

At the transmitter side, for user l, the information bit stream (of length K), $\mathbf{b}^l = \left[b_0^l, \cdots, b_{K-1}^l\right]^{\mathrm{T}}$ is first fed into the Turbo encoder to generate an output stream of length L,

(a) Block diagram of the transmitter



(b) Block diagram of the receiver

Figure 4.1: Schematic diagram of the transmitter and receiver architecture in MIMO SC-FDMA systems

$\mathbf{c}^l = \left[c^l_0, \cdots, c^l_{L-1}\right]^{\mathrm{T}}$, where the effective code rate $R = K/L$. The output from the interleaver, $\mathbf{q}^l = \left[q^l_0, \cdots, q^l_{L-1}\right]^{\mathrm{T}}$ is QAM modulated to produce the symbol block $\mathbf{s}^l = \left[s^l_0, \cdots, s^l_{N-1}\right]^{\mathrm{T}}$, where $s^l_n = \mathscr{Q}\left(\left[q^l_{n,0}, \cdots, q^l_{n,log_2 M-1}\right]\right)$, and M is the cardinality of the QAM symbol set χ. After N-point DFT operation, we obtain the frequency domain symbol vector $\mathbf{x}^l = \left[x^l_0, \cdots, x^l_{N-1}\right]^{\mathrm{T}}$, where $\mathbf{x}^l = \mathbf{F}_N \mathbf{s}^l$, and the symbols over the k-th subcarrier for all users can then be denoted as $\mathbf{x}_k = \left[x^0_k, \cdots, x^{N_U-1}_k\right]^{\mathrm{T}}$. For data symbols from all users over all subcarriers, denoted by $\mathbf{x} = \left[\mathbf{x}^{\mathrm{T}}_0, \cdots, \mathbf{x}^{\mathrm{T}}_{N-1}\right]^{\mathrm{T}}$, we have

$$\mathbf{x} = \left(\mathbf{F}_N \otimes \mathbf{I}_{N_U}\right) \mathbf{s}, \tag{4.1}$$

where the time domain symbol vector for all users are $\mathbf{s} = \left[\mathbf{s}^{\mathrm{T}}_0, \cdots, \mathbf{s}^{\mathrm{T}}_{N-1}\right]^{\mathrm{T}}$ and $\mathbf{s}_n = \left[s^0_n, \cdots, s^{N_U-1}_n\right]^{\mathrm{T}}$.

The received signal over the k-th subcarrier in the frequency domain is

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{z}_k, \tag{4.2}$$

where $\mathbf{z}_k$ is the additive Gaussian noise satisfying $\mathrm{E}\{\mathbf{z}_k\} = \mathbf{0}$ and $\mathrm{E}\left\{\mathbf{z}_k \mathbf{z}^{\mathrm{H}}_k\right\} = \sigma^2_z \mathbf{I}_{N_R}$. $\mathbf{H}_k$

represents the frequency domain channel matrix on the k-th subcarrier. An iterative receiver architecture, as shown in Figure 4.1(b), is considered in this work, where the *extrinsic* information is exchanged between SISO decoder and SISO detector.

4.2 Sorted QR Decomposition Based Soft Interference Cancelation Scheme

4.2.1 SISO MMSE Detection

From Equation (4.2), collecting the received symbols over all subcarriers, we have

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z} = \mathbf{H}\left(\mathbf{F}_N \otimes \mathbf{I}_{N_U}\right)\mathbf{s} + \mathbf{z}, \tag{4.3}$$

where $\mathbf{y} = \left[\mathbf{y}_0^{\mathrm{T}}, \cdots, \mathbf{y}_{N-1}^{\mathrm{T}}\right]^{\mathrm{T}}$, $\mathbf{z} = \left[\mathbf{z}_0^{\mathrm{T}}, \cdots, \mathbf{z}_{N-1}^{\mathrm{T}}\right]^{\mathrm{T}}$, and $\mathbf{H}$ is the frequency domain channel matrix, given by

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_0 & & & \\ & \mathbf{H}_1 & & \\ & & \ddots & \\ & & & \mathbf{H}_{N-1} \end{bmatrix}. \tag{4.4}$$

With the *a priori* information from the channel decoder, we have the mean vector $\boldsymbol{\mu} = \left[\boldsymbol{\mu}_0^{\mathrm{T}}, \cdots, \boldsymbol{\mu}_{N-1}^{\mathrm{T}}\right]^{\mathrm{T}}$, where $\boldsymbol{\mu}_n = \left[\mu_n^0, \cdots, \mu_n^{N_U-1}\right]^{\mathrm{T}}$, and the covariance matrix $\mathbf{C} = \mathrm{diag}\left\{\left[\mathbf{v}_0^{\mathrm{T}}, \cdots, \mathbf{v}_{N-1}^{\mathrm{T}}\right]^{\mathrm{T}}\right\}$, where $\mathbf{v}_n = \left[v_n^0, \cdots, v_n^{N_U-1}\right]^{\mathrm{T}}$, with

$$\mu_n^l = \textstyle\sum_{\forall s, s\in\chi} s \cdot P\left(s_n^l = s\right), \tag{4.5}$$

$$v_n^l = \textstyle\sum_{\forall s, s\in\chi} |s|^2 \cdot P\left(s_n^l = s\right) - \left|\mu_n^l\right|^2, \tag{4.6}$$

and $P\left(s_n^l = s\right) = \prod_{m=0}^{\log_2 M-1} P\left(q_{n,m}^l = b\right)$. Assuming the same covariance for symbols from the same user, the MMSE estimation is then given by

$$\begin{aligned}\hat{\mathbf{s}} = &\left(\mathbf{F}_N^{\mathrm{H}} \otimes \mathbf{I}_{N_U}\right) \mathbf{H}^{\mathrm{H}} \left(\mathbf{H}\mathbf{V}\mathbf{H}^{\mathrm{H}} + \sigma_z^2 \mathbf{I}\right)^{-1} \\ &\cdot \left(\mathbf{y} - \mathbf{H}\left(\mathbf{F}_N \otimes \mathbf{I}_{N_U}\right)\boldsymbol{\mu}\right) + \boldsymbol{\rho}\boldsymbol{\mu},\end{aligned} \tag{4.7}$$

where $\mathbf{V} = \mathbf{I}_N \otimes \bar{\mathbf{V}}$, $\bar{\mathbf{V}} = \mathrm{diag}\left\{\left[\bar{v}^0, \cdots, \bar{v}^{N_U-1}\right]\right\}$, $\bar{v}^l = \frac{1}{N}\sum_{n=0}^{N-1} v_n^l$; $\boldsymbol{\rho} = \mathrm{diag}\left\{\left[\boldsymbol{\rho}_0^{\mathrm{T}}, \cdots, \boldsymbol{\rho}_{N-1}^{\mathrm{T}}\right]^{\mathrm{T}}\right\}$, $\boldsymbol{\rho}_n = \left[\rho_n^0, \cdots, \rho_n^{N_U-1}\right]^{\mathrm{T}}$, and ρ_n^l is given by

$$\rho_n^l = \rho^l = \frac{1}{N}\sum_{k=0}^{N-1} \mathbf{e}_k^{\mathrm{H}} \mathbf{H}_k^{\mathrm{H}} \left(\mathbf{H}_k \bar{\mathbf{V}} \mathbf{H}_k^{\mathrm{H}} + \sigma_z^2 \mathbf{I}_{N_R}\right)^{-1} \mathbf{H}_k \mathbf{e}_k. \tag{4.8}$$

4.2.2 QR decomposition Based Soft Interference Cancelation

Inspired by the work in [22], we propose a new sorted QR decomposition based scheme for soft interference cancelation. We first define a new extended channel matrix as

$$\underline{\mathbf{H}}_k = \begin{bmatrix} \sigma_z^{-1}\mathbf{H}_k \bar{\mathbf{V}}^{1/2} \\ \mathbf{I}_{N_U} \end{bmatrix}. \tag{4.9}$$

The new extended channel matrix includes the *a priori* covariance matrix, which is not considered in [22]. In iterative procedure, $\bar{\mathbf{V}}$, as well as $\underline{\mathbf{H}}_k$ will be updated based on the renewed *a priori* information from channel decoder.

The extended channel matrix can be then decomposed as

$$\underline{\mathbf{H}}_k = \mathbf{Q}_k \mathbf{R}_k = \begin{bmatrix} \mathbf{P}_k \\ \mathbf{T}_k \end{bmatrix} \mathbf{R}_k, \tag{4.10}$$

with the following governing conditions satisfied

$$\mathbf{Q}_k^{\mathrm{H}} \underline{\mathbf{H}}_k = \sigma_z^{-1} \mathbf{P}_k^{\mathrm{H}} \mathbf{H}_k \bar{\mathbf{V}}^{1/2} + \mathbf{T}_k^{\mathrm{H}} = \mathbf{R}_k, \tag{4.11}$$

$$\mathbf{Q}_k^{\mathrm{H}} \mathbf{Q}_k = \mathbf{P}_k^{\mathrm{H}} \mathbf{P}_k + \mathbf{T}_k^{\mathrm{H}} \mathbf{T}_k = \mathbf{I}_{N_U}, \tag{4.12}$$

$$\mathbf{P}_k \mathbf{R}_k = \sigma_z^{-1} \mathbf{H}_k \bar{\mathbf{V}}^{1/2}, \tag{4.13}$$

$$\mathbf{T}_k \mathbf{R}_k = \mathbf{I}_{N_U}, \tag{4.14}$$

where $\mathbf{R}_k$ and $\mathbf{T}_k$ are upper triangular matrices, respectively.

Let $\tilde{\mathbf{y}}_k = \sigma_z^{-1} \mathbf{P}_k^{\mathrm{H}} \mathbf{y}_k$, and through Equation (4.11), we have

$$\tilde{\mathbf{y}}_k = \mathbf{R}_k \bar{\mathbf{V}}^{-1/2} \mathbf{x}_k - \mathbf{T}_k^{\mathrm{H}} \bar{\mathbf{V}}^{-1/2} \mathbf{x}_k + \sigma_z^{-1} \mathbf{P}_k^{\mathrm{H}} \mathbf{z}_k. \tag{4.15}$$

To simplify the matrix representation, we define $\underline{\mathbf{R}}_k = \mathbf{R}_k \bar{\mathbf{V}}^{-1/2}$, $\underline{\mathbf{T}}_k = \mathbf{T}_k^{\mathrm{H}} \bar{\mathbf{V}}^{-1/2}$, and $\underline{\mathbf{P}}_k = \sigma_z^{-1} \mathbf{P}_k^{\mathrm{H}}$. It is easy to derive that $\underline{\mathbf{R}}_k$ and $\underline{\mathbf{T}}_k$ are upper and lower triangular matrices, respectively. We identify the l-th received signal at the k-th subcarrier, $\tilde{y}_{k,l} = \mathbf{e}_l^{\mathrm{H}} \tilde{\mathbf{y}}_k$, i.e.,

$$\begin{aligned} \tilde{y}_{k,l} = & \left(\underline{\mathbf{R}}_k^{(l,l)} - \underline{\mathbf{T}}_k^{(l,l)} \right) x_k^l + \sum_{\ell=l+1}^{N_U-1} \underline{\mathbf{R}}_k^{(l,\ell)} x_k^\ell \\ & - \sum_{\ell=0}^{l-1} \underline{\mathbf{T}}_k^{(l,\ell)} x_k^\ell + \sum_{\ell=0}^{N_R-1} \underline{\mathbf{P}}_k^{(l,\ell)} z_k^\ell, \end{aligned} \tag{4.16}$$

where $\underline{\mathbf{R}}_k^{(l,\ell)}$, $\underline{\mathbf{T}}_k^{(l,\ell)}$, and $\underline{\mathbf{P}}_k^{(l,\ell)}$ are the l-th row, ℓ-th column element in $\underline{\mathbf{R}}_k$, $\underline{\mathbf{T}}_k$, and $\underline{\mathbf{P}}_k$, respectively.

By collectively considering the received symbols of user l over all sub-carriers defining as $\tilde{\mathbf{y}}^l = \left[\tilde{y}_{0,l}, \cdots, \tilde{y}_{N-1,l} \right]^{\mathrm{T}}$, we have

$$\begin{aligned} \tilde{\mathbf{y}}^l = & \left(\underline{\mathbf{R}}^{(l,l)} - \underline{\mathbf{T}}^{(l,l)} \right) \mathbf{F}_N \mathbf{s}^l + \sum_{\ell=l+1}^{N_U-1} \underline{\mathbf{R}}^{(l,\ell)} \mathbf{F}_N \mathbf{s}^\ell \\ & - \sum_{\ell=0}^{l-1} \underline{\mathbf{T}}^{(l,\ell)} \mathbf{F}_N \mathbf{s}^\ell + \sum_{\ell=0}^{N_R-1} \underline{\mathbf{P}}^{(l,\ell)} \mathbf{z}^\ell \end{aligned} \tag{4.17}$$

where the RHS of the equation corresponds to the expected signal, and the interferences from the detected users, undetected users, and noise, respectively. $\mathbf{z}^\ell = \left[z_{0,\ell}, \cdots, z_{N-1,\ell} \right]^{\mathrm{T}}$ denotes the AWGN noise from the ℓ-th receive antenna. $\underline{\mathbf{R}}^{(l,\ell)}$, $\underline{\mathbf{T}}^{(l,\ell)}$, and $\underline{\mathbf{P}}^{(l,\ell)}$ are diagonal matrices: $\underline{\mathbf{R}}^{(l,\ell)} = \mathrm{diag}\left\{ \left[\underline{\mathbf{R}}_0^{(l,\ell)}, \underline{\mathbf{R}}_1^{(l,\ell)}, \cdots, \underline{\mathbf{R}}_{N-1}^{(l,\ell)} \right]^{\mathrm{T}} \right\}$, $\underline{\mathbf{T}}^{(l,\ell)} = \mathrm{diag}\left\{ \left[\underline{\mathbf{T}}_0^{(l,\ell)}, \underline{\mathbf{T}}_1^{(l,\ell)}, \cdots, \underline{\mathbf{T}}_{N-1}^{(l,\ell)} \right]^{\mathrm{T}} \right\}$, and $\underline{\mathbf{P}}^{(l,\ell)} = \mathrm{diag}\left\{ \left[\underline{\mathbf{P}}_0^{(l,\ell)}, \underline{\mathbf{P}}_1^{(l,\ell)}, \cdots, \underline{\mathbf{P}}_{N-1}^{(l,\ell)} \right]^{\mathrm{T}} \right\}$, respectively.

Soft interference cancelation is then performed and the resulting frequency domain sig-

nal, $\widehat{\mathbf{y}}^l$, becomes

$$\begin{aligned}\widehat{\mathbf{y}}^l = &\left(\underline{\mathbf{R}}^{(l,l)} - \underline{\mathbf{T}}^{(l,l)}\right)\mathbf{F}_N\mathbf{s}^l + \sum_{\ell=l+1}^{N_U-1}\underline{\mathbf{R}}^{(l,\ell)}\mathbf{F}_N\left(\mathbf{s}^\ell - \boldsymbol{\mu}^\ell\right)\\ &- \sum_{\ell=0}^{l-1}\underline{\mathbf{T}}^{(l,\ell)}\mathbf{F}_N\left(\mathbf{s}^\ell - \boldsymbol{\mu}^\ell\right) + \sum_{\ell=0}^{N_R-1}\underline{\mathbf{P}}^{(l,\ell)}\mathbf{z}^\ell,\end{aligned} \tag{4.18}$$

where $\boldsymbol{\mu}^l = \left[\mu_0^l, \mu_1^l, \cdots, \mu_{N-1}^l\right]^{\mathrm{T}}$ represents the time domain mean vector of the *a priori* information for user ℓ. We define the last three components in Equation (4.18), which correspond to the residual interference and noise, as $\widehat{\mathbf{z}}^l$, where $\widehat{\mathbf{z}}^l = \sum_{\ell=l+1}^{N_U-1}\underline{\mathbf{R}}^{(l,\ell)}\mathbf{F}_N\left(\mathbf{s}^\ell - \boldsymbol{\mu}^\ell\right) - \sum_{\ell=0}^{l-1}\underline{\mathbf{T}}^{(l,\ell)}\mathbf{F}_N\left(\mathbf{s}^\ell - \boldsymbol{\mu}^\ell\right) + \sum_{\ell=0}^{N_R-1}\underline{\mathbf{P}}^{(l,\ell)}\mathbf{z}^\ell$. It is clear that $\widehat{\mathbf{z}}^l$ comes from three sources: the detected users, the undetected users, and the noise, respectively. It is worth noting that the first component in RHS of Equation (4.18) denotes the expected signal component, and it is clear that we extract additional expected signal component from the interference component identified in [22]. Besides, compared with the work in [22], we also conduct interference cancelation for undetected users. In an iterative procedure, the *a priori* information for the undetected users will be used; hence the residual interference from the undetected users will be reduced. These two improvement leads to the enhancement of SINR.

It is easy to figure out that $\mathrm{E}\left\{\hat{\mathbf{z}}^l\right\} = \mathbf{0}$, and the covariance $\boldsymbol{A}_l = \mathrm{E}\left\{\widehat{\mathbf{z}}^l \cdot \left(\widehat{\mathbf{z}}^l\right)^{\mathrm{H}}\right\}$ is given by

$$\begin{aligned}\boldsymbol{A}_l = &\sum_{\ell=l+1}^{N_U-1}\underline{\mathbf{R}}^{(l,\ell)}\mathbf{F}_N\mathbf{V}^\ell\mathbf{F}_N^{\mathrm{H}}\left(\underline{\mathbf{R}}^{(l,\ell)}\right)^{\mathrm{H}}\\ &+ \sum_{\ell=0}^{l-1}\underline{\mathbf{T}}^{(l,\ell)}\mathbf{F}_N\mathbf{V}^\ell\mathbf{F}_N^{\mathrm{H}}\left(\underline{\mathbf{T}}^{(l,\ell)}\right)^{\mathrm{H}} + \sigma_z^2\sum_{\ell=0}^{N_R-1}\underline{\mathbf{P}}^{(l,\ell)}\left(\underline{\mathbf{P}}^{(l,\ell)}\right)^{\mathrm{H}},\end{aligned} \tag{4.19}$$

where $\mathbf{V}^\ell = \mathrm{diag}\left\{\left[v_0^\ell, \cdots, v_{N-1}^\ell\right]\right\}$. Assuming that the covariance for symbols from the same user is the same, denoted by $\bar{v}^\ell$, Equation (4.19) can be re-written as

$$\begin{aligned}\boldsymbol{A}_l = &\sum_{\ell=l+1}^{N_U-1}\underline{\mathbf{R}}^{(l,\ell)}\left(\underline{\mathbf{R}}^{(l,\ell)}\right)^{\mathrm{H}}\bar{v}^\ell + \sum_{\ell=0}^{l-1}\underline{\mathbf{T}}^{(l,\ell)}\left(\underline{\mathbf{T}}^{(l,\ell)}\right)^{\mathrm{H}}\bar{v}^\ell\\ &+ \sigma_z^2\sum_{\ell=0}^{N_R-1}\underline{\mathbf{P}}^{(l,\ell)}\left(\underline{\mathbf{P}}^{(l,\ell)}\right)^{\mathrm{H}}.\end{aligned} \tag{4.20}$$

Utilizing the properties of QR decomposition given by Equation (4.12), we have $\boldsymbol{A}_l = \mathrm{diag}\left\{\left[\gamma_{0,l}, \gamma_{1,l}, \cdots, \gamma_{N-1,l}\right]^{\mathrm{T}}\right\}$ where $\gamma_{k,l} = \sum_{\ell=l+1}^{N_U-1}\left|\underline{\mathbf{R}}_k^{l,\ell}\right|^2\bar{v}^\ell + 1 - \left|\mathbf{T}_k^{l,l}\right|^2$.

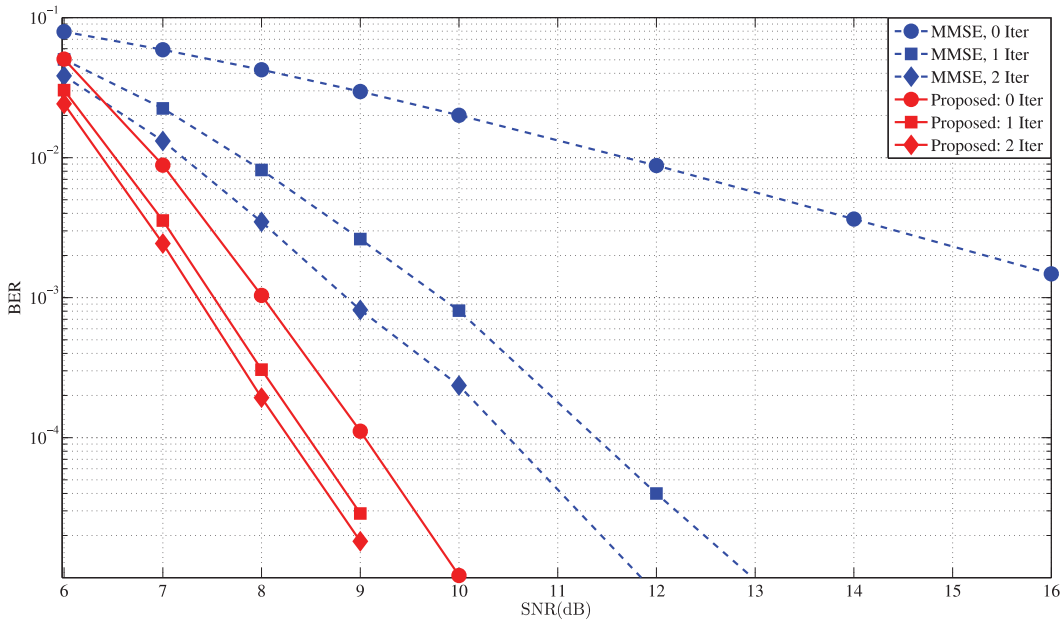


Figure 4.2: BER performance comparison, QPSK, 2/3 Turbo code

With the known expectation and covariance of the interference and noise component, Equation (4.18) can be simplified as $\hat{\mathbf{y}}^l = \left(\underline{\mathbf{R}}^{(l,l)} - \underline{\mathbf{T}}^{(l,l)}\right)\mathbf{F}_N\mathbf{s}^l + \hat{\mathbf{z}}^l$. Based on Equation (4.18), the MMSE estimation of the symbols from user l now becomes

$$\hat{\mathbf{s}}^l = \mathbf{F}_N^{\mathrm{H}}\boldsymbol{B}_l^{\mathrm{H}}\left(\boldsymbol{B}_l\boldsymbol{B}_l^{\mathrm{H}}\bar{v}^l + \boldsymbol{A}_l\right)^{-1}\left(\widehat{\mathbf{y}}^l - \boldsymbol{b}_l\mathbf{F}_N\boldsymbol{\mu}^l\right) + \boldsymbol{\lambda}^l\boldsymbol{\mu}^l, \tag{4.21}$$

where $\boldsymbol{B}_l = \underline{\mathbf{R}}^{(l,l)} - \underline{\mathbf{T}}^{(l,l)}$ is a diagonal matrix with the k-th diagonal element $B_{k,l} = \underline{\mathbf{R}}_k^{l,l} - \underline{\mathbf{T}}_k^{l,l}$; $\boldsymbol{\lambda}^l = \mathrm{diag}\left\{\left[\lambda_0^l, \cdots, \lambda_{N-1}^l\right]\right\}$, where $\lambda_n^l = \frac{1}{N}\sum_{k=0}^{N-1}\frac{|\boldsymbol{B}_{k,l}|^2}{|\boldsymbol{B}_{k,l}|^2\bar{v}^l + \gamma_{k,l}}$.

We rewrite $\hat{s}_n^l$ from Equation (4.21) as

$$\hat{s}_n^l = \beta_n^l s_n^l + \varsigma_n^l. \tag{4.22}$$

Similarly, we have

$$\beta_n^l = \lambda_n^l, \tag{4.23}$$

$$\mathrm{E}\left\{\varsigma_n^l\right\} = 0, \quad \mathrm{E}\left\{\varsigma_n^l\left(\varsigma_n^l\right)^{\mathrm{H}}\right\} = \lambda_n^l\left(1 - \lambda_n^l\bar{v}^l\right). \tag{4.24}$$

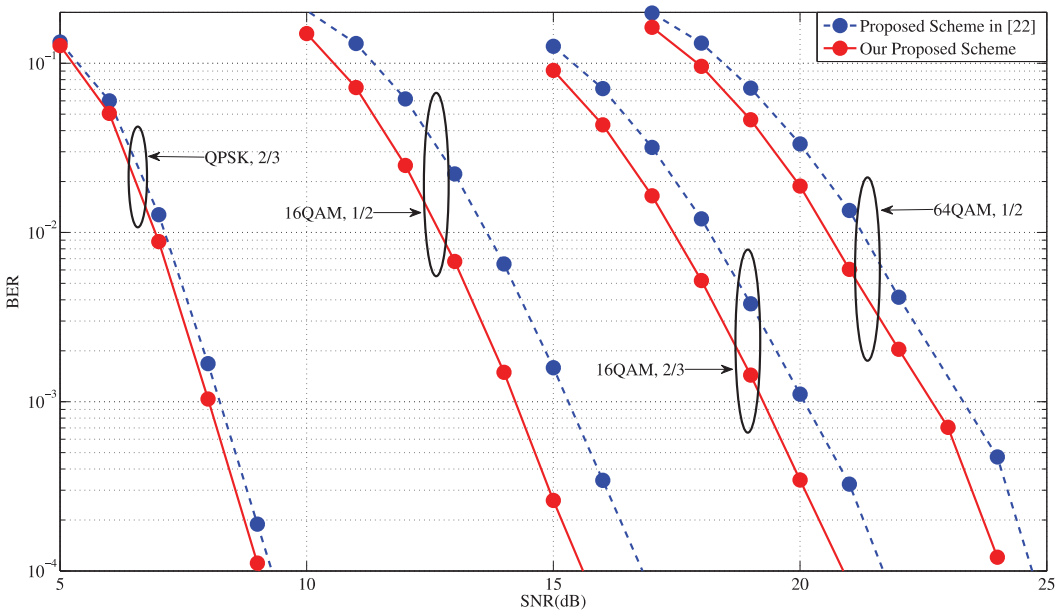


Figure 4.3: BER performance comparison, 0 iteration

With the signal model in (4.22), we can derive the conditional probability density function $p\left(\hat{s}_n^l | s_n^l = s\right)$, which will be used to compute the *extrinsic* LLR $L_E\left(q_{n,m}^l\right)$. After the detection of all symbols from user l, as shown in Figure 4.1(b), $L_E\left(q_{n,m}^l\right)$ is deinterleaved, and fed into turbo decoder for decoding.

4.2.3 Computational Complexity Analysis

In our scheme, the size of the extended channel matrix is the same as that in [22]; therefore, the computation complexity of the QR decomposition is the same as that in [22]. We introduce soft interference cancelation for the undetected users; hence, extra computation is put on the current user's detection. However, at each iteration, with the known *a priori* information, the previous mentioned procedure can be pre-processed. Alternatively, the overall computation of the proposed scheme is almost the same as that in [22], which is of low complexity compared to the traditional MMSE-based schemes.

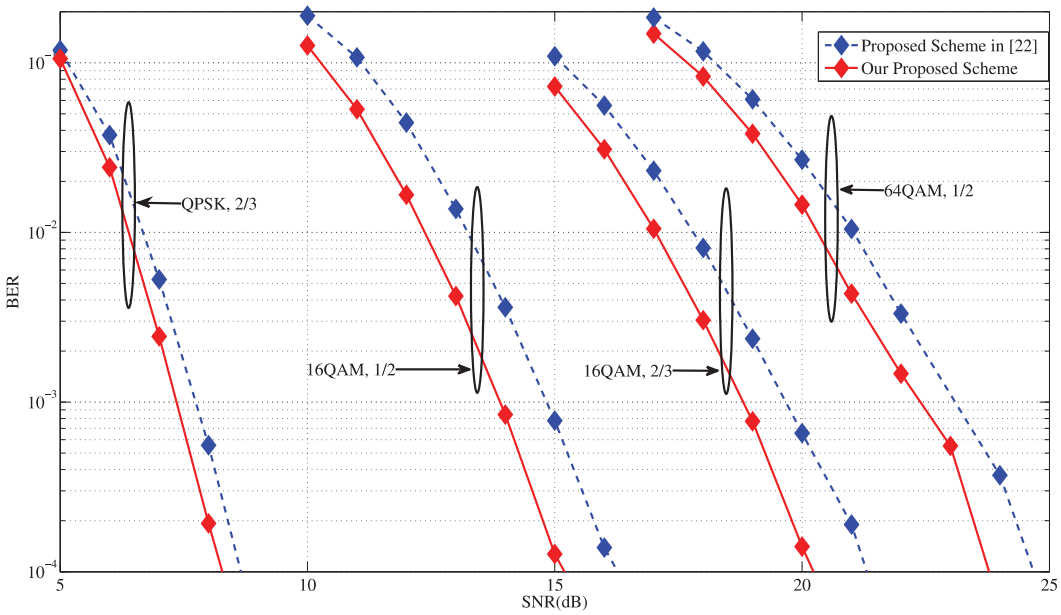


Figure 4.4: BER performance comparison, 2 iteration

4.3 Performance Evaluation and Discussions

We now evaluate the performance of the proposed scheme in fading channels. The system bandwidth is 5MHz with 300 subcarriers. The Extended Vehicular A channel (EVA) from [22] is adopted. A 4×4 MIMO SC-FDMA system is considered for performance analysis.

Figure 4.2 shows the BER performance comparison of the proposed scheme with the traditional MMSE scheme. QPSK modulation and turbo code with a code rate 2/3 are considered. Apparently, the proposed scheme outperforms the traditional MMSE scheme in every iteration and the performance gain is remarkable. These results are consistent with the conclusions drawn from literature where a SQRD-based detection scheme is investigated.

In Figure 4.3, we compare the BER performance of the proposed scheme with that of the scheme proposed in [22] for the initial iteration for different modulation schemes and code rates. Because the *a priori* information from the turbo decoder is not yet available, the extended channel matrix in our proposed scheme essentially degrades to that in [22].

However, as explained in the previous section, because of the ability to extract additional expected signal component from the interference component identified in [22], the SINR of our proposed scheme is improved, thus leading to the apparent BER performance improvement.

We further investigate how iteration helps to improve system performance and compare the results from both schemes after two iterations, as shown in Figure 4.4. The performance enhancement is obvious. With iterations, the *a priori* information from the turbo decoder is available for the detection process. Therefore, the SQRD processing can use the covariance matrix updated with the renewed *a priori* information to output an optimal detection order that limits the error propagation issues in successive interference cancelation. Then, the successive interference cancelation process performs interference cancelation for the undetected users, and the *a priori* information is utilized to update the *a priori* mean and the residual interference from the undetected users. The advantage in utilizing the *a priori* in iterative processing as well as the capability to extracts additional expected signal component contribute the obvious BER performance gain compared to that of the scheme proposed in [22].

4.4 Summary

In this chapter, we have investigated the sorted QR decomposition based soft interference cancelation in multi-user MIMO SC-FDMA systems. We have proposed an improved SQRD-based multi-user detection scheme which is featured by 1) employing the *a priori* information from the channel decoder in the sorted QR decomposition procedure; 2) performing soft successive interference cancelation with the *a priori* information for undetected users data; and 3) extracting more precious expected signal component by separating the expected signal from the interference component. Simulation results show that the pro-

posed scheme outperforms the MMSE-based turbo receivers and the original SQRD-based multi-user detection scheme in terms of BER performance for various scenarios.

Chapter 5

Stair Matrix and its Applications to Massive MIMO Uplink Data Detection

[1] In this chapter, we propose to develop an iterative method to achieve near linear MMSE detection scheme use the stair matrix. A stair matrix is a special tridiagonal matrix where the off-diagonal elements in either the even or the odd rows are zeros [91]. In the development of the iterative method using stair matrix, we need to address two fundamental issues: the probability that the convergence conditions are satisfied, and the convergence rate. The previous one tells whether the stair matrix is applicable or not (applicability) in massive MIMO, while the latter one reveals the advantages over the use of a diagonal matrix. We

[1]The related work has been published in or submitted to:

- *IEEE Wireless Communications Letters*, 2017, under review.
- *IEEE Transactions on Communications*, accepted, DOI: 10.1109/TCOMM.2017.2789211.
- *Proceedings of the International Conference on Communications*, (ICC'17), Paris, France, May 2017.

address these issues in massive MIMO, and the contributions are summarized as follows:

- We show that when N_B grows to infinity, the probability that the convergence conditions are met approaches 1. As the number of antennas at a base station in Massive MIMO systems can be hundreds, this conclusion demonstrates the applicability of the stair matrix in massive MIMO systems. In finite N_B region (or low r region), we show that by using the stair matrix, the probability that the convergence conditions are met can be greatly improved, and the cumulative distribution function of the maximum eigenvalue of the convergence matrix indicates that the convergence rate can be speeded up by using the stair matrix;

- We propose an iterative method with the use of the stair matrix, and demonstrate the proposed scheme has the same level of computational complexity compared to the existing iterative methods where the diagonal matrix is applied;

- We apply the stair matrix in Neumann series expansion to approach matrix inverse, and demonstrate that the mean-square error of the truncated expansion can be greatly reduced compared to the use of diagonal matrix;

- The proposed iterative method is to approximate linear MMSE estimation vector, and we show that the residual estimation error, between the iterative estimation and the linear MMSE estimation, is much less than that of the Jacobi method where the diagonal matrix is applied;

- We perform numerical simulations to evaluate the system BER performance, and show that the proposed iterative method achieves significant performance improvement over the proposals where diagonal matrix is adopted in the development;

- We apply the proposed scheme to a massive MIMO system with extended vehicular A

(EVA) channel, and show significant BER performance enhancement over the reported near linear MMSE detection schemes.

The rest of this chapter is organized as follows. Section 5.1 provides the system model, including the massive MIMO structure and the preliminary work of linear ZF/MMSE detection. In Section 5.2, the introduction to stair matrix and its applicability in massive MIMO will be presented. The implementation of stair matrix in massive MIMO data detection with iterative method is presented in Section 5.3. In Section 5.4, we conduct the numerical simulations and present the results and discussion. Finally, the conclusions are summarized in Section 5.5.

5.1 System Model

We consider the massive MIMO uplink with N_B antennas at a base station to simultaneously serve N_U single-antenna user equipment. The N_U bitstream from each user is first encoded, then interleaved, and fed into digital modulator. The modulated symbols are transmitted into the massive MIMO channel, and the received signal vector at base station can be expressed as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z}, \tag{5.1}$$

where $\mathbf{y} = [y_1, y_2, \cdots, y_{N_B}]^{\mathrm{T}}$ is a complex-valued $N_B \times 1$ vector, with y_m denoting the received signal from the m-th receiving antenna. $\mathbf{x} = [x_1, x_2, \cdots, x_{N_U}]^{\mathrm{T}}$ with the transmitted symbol of user u denoted by x_u. The transmitted symbols are unit power normalized and independent from each other, i.e., $\mathrm{E}\left\{\mathbf{x}\mathbf{x}^{\mathrm{H}}\right\} = \mathbf{I}_{N_U}$. $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \cdots, \mathbf{h}_{N_U}]$ denotes the channel matrix with $\mathbf{h}_u \in \mathbb{C}^{N_B \times 1}$ where each entry is independent and identically distributed (i.i.d.), modeled as the flat Rayleigh fading channel [2, 29, 69, 84]. $\mathbf{z} = [z_1, z_2, \cdots, z_{N_B}]^{\mathrm{T}}$ is the noise vector, satisfying $\mathrm{E}\left\{\mathbf{z}\mathbf{z}^{\mathrm{H}}\right\} = \sigma_z^2 \mathbf{I}_{N_B}$ with each entry modeled as zero-mean complex Gaus-

sian circularly symmetric (ZMCGCS) random variable. It is worth noting that in frequency selective fading channels, by applying the orthogonal frequency division multiplexing and single-carrier frequency division multiple access (OFDM/SC-FDMA) techniques, the signal model expressed in Equation (5.1) is established over each subcarrier.

5.1.1 Linear MMSE Data Detection

The multi-user data detector at the base station is to compute the *a posteriori* log likelihood ratio (LLR) of the bits associated with the modulated symbols. After the knowledge of the channel matrix (note that the channel matrix is obtained through channel estimator, where time domain and/or frequency domain training pilots are used for the channel estimation [95, 96]), the well-known linear MMSE data detection can be given by

$$\hat{\mathbf{x}} = \left(\mathbf{H}^{\mathrm{H}}\mathbf{H} + \sigma_z^2\mathbf{I}_{N_U}\right)^{-1}\mathbf{H}^{\mathrm{H}}\mathbf{y} = \mathbf{W}^{-1}\mathbf{y}^{\mathrm{MF}}, \tag{5.2}$$

where $\mathbf{y}^{\mathrm{MF}} = \mathbf{H}^{\mathrm{H}}\mathbf{y}$ can be seen as the matched-filter output, and the MMSE equalization matrix $\mathbf{W}$ can be expressed as

$$\mathbf{W} = \mathbf{G} + \sigma_z^2\mathbf{I}_{N_U}, \tag{5.3}$$

where $\mathbf{G} = \mathbf{H}^{\mathrm{H}}\mathbf{H}$ is the Gram matrix. It is worth noting that in high signal-to-noise ratio (SNR) region, if the Gram matrix is invertible [2], Equation (5.2) can be reduced to

$$\hat{\mathbf{x}} = \mathbf{G}^{-1}\mathbf{y}^{\mathrm{MF}}, \tag{5.4}$$

which is the linear ZF data detection scheme, where the noise component is not considered in the equalization process.

To obtain the *a posteriori* LLR of the bits associated with the modulated symbols, we

[2] According to the random matrix theory [2, 97], in massive MIMO, the probability that $\mathbf{G}$ is invertible is high.

write the estimation in Equation (5.2) as

$$\hat{x}_u = \mathbf{e}_u^{\mathrm{H}} \hat{\mathbf{x}} = \rho_u x_u + \xi_u, \tag{5.5}$$

where the equivalent channel gain ρ_u and the *a posteriori* noise-plus-interference (NPI) ξ_u can be given by

$$\rho_u = \mathbf{e}_u^{\mathrm{H}} \mathbf{W}^{-1} \mathbf{G} \mathbf{e}_u, \tag{5.6}$$

$$\xi_u = \mathbf{e}_u^{\mathrm{H}} \mathbf{W}^{-1} \mathbf{G} \left(\mathbf{x} - x_u \mathbf{e}_u \right) + \mathbf{e}_u^{\mathrm{H}} \mathbf{W}^{-1} \mathbf{H}^{\mathrm{H}} \mathbf{z}. \tag{5.7}$$

The covariance of the NPI $v_u^2 = \mathrm{cov}\left(\xi_u, \xi_u\right)$ is given by

$$\begin{aligned} v_u^2 &= \mathbf{e}_u^{\mathrm{H}} \mathbf{W}^{-1} \mathbf{G} \mathbf{G} \mathbf{W}^{-1} \mathbf{e}_u + \sigma_z^2 \mathbf{e}_u^{\mathrm{H}} \mathbf{W}^{-1} \mathbf{G} \mathbf{W}^{-1} \mathbf{e}_u - \rho_u^2 \\ &= \rho_u - \rho_u^2. \end{aligned} \tag{5.8}$$

Given Equations (5.5), (5.6), and (5.8), we derive the *max-log* approximated LLR of the bits associated with x_u, given by

$$L\left(b_{u,k}\right) = \gamma_u \left(\min_{s \in \chi_k^0} \left| \frac{\hat{x}_u}{\rho_u} - s \right|^2 - \min_{s' \in \chi_k^1} \left| \frac{\hat{x}_u}{\rho_u} - s' \right|^2 \right), \tag{5.9}$$

where $b_{u,k}$ is the k-th mapping bit associated with x_u; $\gamma_u = \rho_u^2 / v_u^2$ is the *a posteriori* signal-to-noise-plus-interference ratio (SINR); $\chi_k^b \triangleq \{ s \,|\, s \in \chi, q_k = b \}$ denotes the subset of χ, where the k-th mapping bit associated with the constellation symbol s, i.e. q_k, is b; χ is the constellation symbols set. After data detection of all users, the LLRs are fed into the soft-input channel decoder for the decoding process.

5.1.2 Neumann Series Expansion

In the previous subsection, we note that the matrix inversion operations are involved in linear MMSE/ZF data detection. The matrix inverse is computational costly especially when the matrix size is large. One of the promising practical solutions to address the

matrix inversion issue is to employ the Neumann series expansion [23, 98]. The complete Neumann series expansion of the matrix inverse $\mathbf{W}^{-1}$ is given by [23, 98]

$$\mathbf{W}^{-1} = \sum_{l=0}^{\infty} \left(\mathbf{X}^{-1}\left(\mathbf{X}-\mathbf{W}\right)\right)^{l}\mathbf{X}^{-1}, \tag{5.10}$$

where $\mathbf{X} \neq \mathbf{W}$, and the following conditions are satisfied:

$$\lim_{l\to\infty}\left(\mathbf{I}-\mathbf{X}^{-1}\mathbf{W}\right)^{l} = \mathbf{0}. \tag{5.11}$$

When the high orders are ignored, the truncated Neumann series expansion can be expressed as

$$\mathbf{W}_{L}^{-1} = \sum_{l=0}^{L-1} \left(\mathbf{X}^{-1}\left(\mathbf{X}-\mathbf{W}\right)\right)^{l}\mathbf{X}^{-1}. \tag{5.12}$$

The significance of the Neumann series expansion is to approach matrix inversion by using matrix multiplications with known $\mathbf{X}^{-1}$. Specifically, if the inverse of the selected matrix $\mathbf{X}^{-1}$ is easy to obtain, then the truncated expansion $\mathbf{W}_{L}^{-1}$ can be used to approximate $\mathbf{W}^{-1}$. Generally, when we select the matrix $\mathbf{X}$ that is close to $\mathbf{W}$, a few L orders expansion $\mathbf{W}_{L}^{-1}$ in Equation (5.12) can be close to $\mathbf{W}^{-1}$. Fortunately, in massive MIMO systems, the gram matrix $\mathbf{G}$ is diagonally dominant; hence the diagonal matrix, i.e., $\mathbf{D} = \mathrm{diag}\left\{\left[\mathbf{W}_{(0,0)}, \mathbf{W}_{(1,1)}, \cdots, \mathbf{W}_{(N_U-1,N_U-1)}\right]^{\mathrm{T}}\right\}$ can be selected as $\mathbf{X}$, then the L order expansion $\mathbf{W}_{L}^{-1}$ is given by

$$\mathbf{W}_{L}^{-1} = \sum_{l=0}^{L-1} \left(\mathbf{D}^{-1}\left(\mathbf{D}-\mathbf{W}\right)\right)^{l}\mathbf{D}^{-1}. \tag{5.13}$$

In [23], the authors have provided the upper bound of the residual estimation error [3] using $\mathbf{W}_{L}^{-1}$ to approximate $\mathbf{W}^{-1}$, i.e.,

$$\left\|\left(\mathbf{W}^{-1}-\mathbf{W}_{L}^{-1}\right)\mathbf{y}^{\mathrm{MF}}\right\|_{2} \leqslant \left\|\mathbf{I}-\mathbf{D}^{-1}\mathbf{W}\right\|_{\mathrm{F}}^{L}\left\|\hat{\mathbf{x}}\right\|_{2}, \tag{5.14}$$

[3] The residual estimation error defined in [23] is to evaluate the mean-square error between the estimation using the L order expansion and the exact linear MMSE estimation, since the previous one is to approximate the latter one.

where $\|\mathbf{A}\|_{\mathrm{F}}$ and $\|\mathbf{a}\|_2$ are the Frobenius norm of a matrix $\mathbf{A}$ and the ℓ_2-norm of a vector $\mathbf{a}$. From Equation (5.14), we can see that the upper bound of residual estimation error decreases as the increase of the expansion order and N_B. In other words, if the number of antennas at the base station is sufficiently large, even with a small order expansion, the residual estimation error will be small. Particularly, when N_B is sufficiently large and the expansion order $L \leqslant 2$, the computation required for the Neumann series expansion will be much reduced, compared to the matrix inverse operations. These two observations support the use of the diagonal matrix in Neumann series expansion for massive MIMO systems.

5.1.3 Jacobi Method

In Neumann series expansion, if the expansion order is greater than 2, the matrix multiplication operations are involved; hence, the computational complexity is comparable with that of the matrix inverse operations. On the other hand, as we can see in Equation (5.14), if N_B is not sufficiently large, with an expansion order that is less than 2, the residual estimation error is still considerable. These two factors limit the applications of diagonal matrix in Neumann series expansion.

To avoid the matrix multiplication operations, but maintain a reasonable orders of expansion, we can use the iterative methods. To be specific, we first rewrite the MMSE estimation in Equation (5.2) as

$$\mathbf{W}\hat{\mathbf{x}} = \mathbf{y}^{\mathrm{MF}}. \tag{5.15}$$

By transforming the matrix inverse problem into the format of Equation (5.15), we can adopt the iterative methods to solve linear equations. Generally, the iterative methods follow the following process:

(1) Provide an initial estimation;

(2) Follow an iterative structure to obtain the next estimation;

(3) When the estimation converges, output the final estimation.

In the Jacobi method, we have the initial estimation as

$$\mathbf{x}^{(0)} = \mathbf{D}^{-1}\mathbf{y}^{\mathrm{MF}}, \tag{5.16}$$

which is the common selection in most of the existing literature. The iterative structure is given by

$$\begin{aligned}\mathbf{x}^{(i+1)} &= \mathbf{D}^{-1}\left((\mathbf{D}-\mathbf{W})\,\mathbf{x}^{(i)} + \mathbf{y}^{\mathrm{MF}}\right) \\ &= \mathbf{x}^{(i)} - \mathbf{D}^{-1}\mathbf{W}\mathbf{x}^{(i)} + \mathbf{D}^{-1}\mathbf{y}^{\mathrm{MF}},\end{aligned} \tag{5.17}$$

where $\mathbf{x}^{(i)}$ denotes the i-th estimation. According to the iterative structure in Equation (5.17), and using the initial estimation given by Equation (5.16), we can derive the i-th estimation given by

$$\mathbf{x}^{(i)} = \sum_{l=0}^{i}\left(\mathbf{D}^{-1}\left(\mathbf{D}-\mathbf{W}\right)\right)^{l}\mathbf{D}^{-1}\mathbf{y}^{\mathrm{MF}}. \tag{5.18}$$

That is to say, by selecting the initial estimation given by (5.16), after i iterations following Jacobi iterative structure, we have the same estimation results as the $(i+1)$-th order expansion in the Neumann series. Therefore, the convergence conditions, the residual estimation error, and the estimation results are the same as those in the previous subsection. However, as we can see from Equation (5.16) to Equation (5.17), only matrix-vector product operations are involved; therefore, the Jacobi method has low complexity compared to the Neumann series expansion with the same iterations (or orders in Neumann series expansion).

5.2 Stair Matrix and its Applicability to Massive MIMO Systems

In this section, we will first introduce the stair matrix and its properties. And then, we will demonstrate the applicability of the stair matrix to massive MIMO systems.

5.2.1 Stair Matrix and its Properties

To begin with, we have the following definitions.

Definition 1 [91]: In an $N \times N$ matrix $\mathbf{A}$, if its entry $\mathbf{A}_{(m,n)} = \mathbf{e}_m^H \mathbf{A} \mathbf{e}_n$, $m, n = 1, 2, \cdots, N$, satisfies $\mathbf{A}_{(m,n)} = 0$ where $n \notin \{m-1, m, m+1\}$, we then call it as a tridiagonal matrix, denoted by $\mathbf{A} = \text{tridiag}\big(\mathbf{A}_{(m,m-1)}, \mathbf{A}_{(m,m)}, \mathbf{A}_{(m,m+1)}\big)$.

A typical tridiagonal matrix $\mathbf{A}$ can be given as

$$\begin{bmatrix} \mathbf{A}_{(1,1)} & \mathbf{A}_{(1,2)} & & & & & \\ \mathbf{A}_{(2,1)} & \mathbf{A}_{(2,2)} & \mathbf{A}_{(2,3)} & & & & \\ & \ddots & \ddots & & \ddots & & \\ & & \mathbf{A}_{(N-1,N-2)} & \mathbf{A}_{(N-1,N-1)} & \mathbf{A}_{(N-1,N)} \\ & & & \mathbf{A}_{(N,N-1)} & \mathbf{A}_{(N,N)} \end{bmatrix}$$

Definition 2 [91, 92]: If a tridiagonal matrix satisfies one of the following conditions:

(I) $\mathbf{A}_{(m,m-1)} = 0$, $\mathbf{A}_{(m,m+1)} = 0$, where $m = 2k-1$, $k = 1, 2, \cdots, \lfloor (N+1)/2 \rfloor$. Alternatively, the non-diagonal elements in the odd rows of tridiagonal matrix are zeros;

(II) $\mathbf{A}_{(m,m-1)} = 0$, $\mathbf{A}_{(m,m+1)} = 0$, where $m = 2k$, $k = 1, 2, \lfloor N/2 \rfloor$. In other words, the non-diagonal elements in the even rows of tridiagonal matrix are zeros;

we then call it as a stair matrix, denoted by $\mathbf{A} = \text{stair}\big(\mathbf{A}_{(m,m-1)}, \mathbf{A}_{(m,m)}, \mathbf{A}_{(m,m+1)}\big)$.

In accordance, a stair matrix is of type I if the condition (I) is satisfied and is of type II if the condition (II) is satisfied. For example, a 5×5 stair matrix has the following forms:

$$\mathbf{A} = \begin{bmatrix} \times & & & & \\ \times & \times & \times & & \\ & & \times & & \\ & & \times & \times & \times \\ & & & & \times \end{bmatrix} \quad \text{or} \quad \mathbf{A} = \begin{bmatrix} \times & \times & & & \\ & \times & & & \\ & \times & \times & \times & \\ & & & \times & \\ & & & \times & \times \end{bmatrix}.$$

The previous one is of type I and the latter one is of type II. Next, we provide the following properties of the stair matrix in ***Corollary*** 1 and 2.

Corollary 1. *Let* $\mathbf{A}$ *be a stair matrix. Then* $\mathbf{A}^H$ *is also a stair matrix. In addition, if* $\mathbf{A}$ *is of type I, then* $\mathbf{A}^H$ *is of type II, and vice verse.*

Proof. Using the definition, it is straightforward to obtain *Corollary* 1. □

Corollary 1 shows that the properties of the stair matrix of type I and type II are almost the same; therefore, we only consider the stair matrix of type I hereafter except for specification.

Corollary 2. *Let* $\mathbf{A}$ *be a stair matrix.* $\mathbf{A}$ *is nonsingular if and only if* $\mathbf{A}_{m,m}$, $m = 1, 2, \cdots, N$, *is nonzero. Furthermore, the inverse of* $\mathbf{A}$, *i.e.,* $\mathbf{A}^{-1}$ *is also a stair matrix of the same type, given by* $\mathbf{A}^{-1} = \mathbf{D}^{-1}(2\mathbf{D} - \mathbf{A})\mathbf{D}^{-1}$, *where* $\mathbf{D} = diag\left(\left[\mathbf{A}_{(1,1)}, \mathbf{A}_{(2,2)}, \cdots, \mathbf{A}_{(N,N)}\right]\right)$ *is the diagonal matrix extracted from* $\mathbf{A}$.

Proof. Since $\det(\mathbf{A}) = \prod_{m=1}^{N} \mathbf{A}_{(m,m)}$, we can see that $\mathbf{A}$ is nonsingular if and only if $\mathbf{A}_{(m,m)}$, $m = 1, 2, \cdots, N$, is nonzero.

Following the matrix multiplications, we can obtain that $\mathbf{D}^{-1}(2\mathbf{D} - \mathbf{A})\mathbf{D}^{-1}\mathbf{A} = \mathbf{I}_N$. Moreover, we can easily derive that $\mathbf{A}^{-1}$ is also a stair matrix and of the same type as $\mathbf{A}$. □

From *Corollary* 2, we have the **Algorithm 5.1** to obtain $\mathbf{A}^{-1}$. It is clear from **Algorithm 5.1** that the complexity to obtain the inverse of a stair matrix is $O(N)$, which is the same order of the computation of $\mathbf{D}^{-1}$.

5.2.2 Using Stair Matrix in Neumann Series Expansion

As the linear MMSE detection scheme is known to have better system performance compared to the linear ZF detection scheme, we show the use of the stair matrix in approaching

Table 5.1: **Algorithm 5.1:** Compute the Inverse of a Stair Matrix

Input: The Stair Matrix $\mathbf{A} = \text{stair}\left(\mathbf{A}_{(m,m-1)}, \mathbf{A}_{(m,m)}, \mathbf{A}_{(m,m+1)}\right)$
Output: $\mathbf{A}^{-1} = \mathbf{B} = \text{stair}\left(\mathbf{B}_{(m,m-1)}, \mathbf{B}_{(m,m)}, \mathbf{B}_{(m,m+1)}\right)$
1.for $m = 1:1:N$
2. $\quad \mathbf{B}_{(m,m)} = 1 \big/ \mathbf{A}_{(m,m)}$
3. end
4. for $m = 2:2:2\lfloor N/2 \rfloor$
5. $\quad \mathbf{B}_{(m,m-1)} = -\mathbf{A}_{(m,m-1)}\mathbf{B}_{(m,m)}\mathbf{B}_{(m-1,m-1)}$;
6. $\quad \mathbf{B}_{(m,m+1)} = -\mathbf{A}_{(m,m+1)}\mathbf{B}_{(m,m)}\mathbf{B}_{(m+1,m+1)}$;
7. end
Return B.

linear ZF estimation in this Section. In Section 5.4, we show the corresponding results of using stair matrix in approximating linear MMSE estimation.

To begin with, we extract the stair matrix, $\mathbf{S} = \text{stair}(\mathbf{G}_{u,u-1}, \mathbf{G}_{u,u}, \mathbf{G}_{u,u+1})$, from the Gram matrix, given by

$$\mathbf{S}_{(u,v)} = \begin{cases} \mathbf{G}_{(u,v)}, & u \in \mathbb{U}_1, v = u; \\ \mathbf{G}_{(u,v)}, & u \in \mathbb{U}_2, v \in \{u-1, u, u+1\}; \\ 0, & \text{otherwise}, \end{cases}$$

where $\mathbb{U} \triangleq \{n | n \in \mathbb{N}, n \leqslant N_U + 1\}$, denoting the positive integer numbers less than N_U; $\mathbb{U}_1$ and $\mathbb{U}_2$ are subsets of $\mathbb{U}$, defined as $\mathbb{U}_1 \triangleq \{n | n \in \mathbb{U}, n = 2k-1, k \in \mathbb{N}\}$, denoting the odd numbers less than N_U, and $\mathbb{U}_2 \triangleq \{n | n \in \mathbb{U}, n = 2k, k \in \mathbb{N}\}$, denoting the even numbers less than N_U, respectively.

Applying the stair matrix in Neumann series expansion, similar to Equation (5.10), we

have

$$\mathbf{G}^{-1} = \sum_{k=0}^{\infty} \left(\mathbf{I} - \mathbf{S}^{-1}\mathbf{G}\right)^k \mathbf{S}^{-1}, \tag{5.19}$$

where $\mathbf{X}$ is replaced with the stair matrix $\mathbf{S}$ and the Gram matrix is considered. The convergence condition for Equation (5.19) is

$$\lim_{k\to\infty} \left(\mathbf{I} - \mathbf{S}^{-1}\mathbf{G}\right)^k = \mathbf{0}, \tag{5.20}$$

or equivalently

$$\rho\left(\mathbf{B}\right) = |\lambda_0| < 1, \tag{5.21}$$

i.e., the maximum eigenvalue of the convergence matrix $\mathbf{B}$ is less than one. $\mathbf{B} = \mathbf{I} - \mathbf{S}^{-1}\mathbf{G}$, and $\rho\left(\mathbf{B}\right)$ is the spectral radius of the matrix $\mathbf{B}$. $|\lambda_0| \geqslant |\lambda_1| \geqslant \cdots \geqslant |\lambda_{N_U-1}|$ denote the N_U eigenvalues of $\mathbf{B}$.

The convergence condition is critical for the application of the stair matrix in massive MIMO systems. In order to investigate the maximum eigenvalue of $\mathbf{B}$, we suppose N_U is odd [4], and derive each entry in $\mathbf{B}$ given by

$$\mathbf{B}_{(u,v)} = \begin{cases} -\frac{\mathbf{G}_{(u,v)}}{\mathbf{G}_{(u,u)}}, & u \in \mathbb{U}_1, v \neq u; \\ 0, & u \in \mathbb{U}_1, v = u; \\ -\frac{\mathbf{G}_{(u,v)}}{\mathbf{G}_{(u,u)}} + \frac{\mathbf{G}_{(u,u-1)}\cdot\mathbf{G}_{(u-1,v)}}{\mathbf{G}_{(u,u)}\mathbf{G}_{(u-1,u-1)}} + \frac{\mathbf{G}_{(u,u+1)}\cdot\mathbf{G}_{(u+1,v)}}{\mathbf{G}_{(u,u)}\mathbf{G}_{(u+1,u+1)}}, & u \in \mathbb{U}_2, v \neq u; \\ \frac{\mathbf{G}_{(u,u-1)}\cdot\mathbf{G}_{(u-1,u)}}{\mathbf{G}_{(u,u)}\mathbf{G}_{(u-1,u-1)}} + \frac{\mathbf{G}_{(u,u+1)}\cdot\mathbf{G}_{(u+1,u)}}{\mathbf{G}_{(u,u)}\mathbf{G}_{(u+1,u+1)}}, & u \in \mathbb{U}_2, v = u. \end{cases} \tag{5.22}$$

where **Algorithm 5.1** is used to compute the matrix inverse of the stair matrix.

We have the following theorem:

Theorem 3. $\mathbf{B}_{(u,v)}$ *is given by Equation* (5.22)*, and* $|\lambda_0| \geqslant |\lambda_1| \geqslant \cdots \geqslant |\lambda_{N_U-1}|$ *denote the* N_U *eigenvalues of* $\mathbf{B}$*. We have*

$$\lim_{N_B\to\infty} \Pr\left\{|\lambda_0| < 1\right\} = 1. \tag{5.23}$$

[4] When N_U is even, the difference in the expression of $\mathbf{B}$ is only present in the last row. However, the general result is also expected.

Proof. See Appendix B. □

Theorem 3 shows that when the number of antennas at base station is sufficiently large, the convergence conditions in Equation (5.20) can be satisfied; therefore, using the stair matrix in Neumann series expansion can be established. As we show in Section V, for the system with 25 active users under service, $N_B = 150$ is sufficient to meet the convergence conditions.

Hence we demonstrate the applicability of the stair matrix in massive MIMO systems.

5.2.3 Residual Estimation Error

We now investigate the residual estimation error by using the truncated Neumann series expansion. According to Equation (5.12), we have

$$\mathbf{G}_L^{-1} = \sum_{l=0}^{L-1} \left(\mathbf{S}^{-1}\left(\mathbf{S}-\mathbf{G}\right)\right)^l \mathbf{S}^{-1}. \tag{5.24}$$

Replacing $\mathbf{G}^{-1}$ with $\mathbf{G}_L^{-1}$ in Equation (5.4), we have

$$\hat{\mathbf{x}}^{(L)} = \mathbf{G}_L^{-1}\mathbf{y}^{\mathrm{MF}}. \tag{5.25}$$

Therefore, the residual estimation error[5] $J = \left\|\hat{\mathbf{x}}^{(L)} - \hat{\mathbf{x}}\right\|_2$, is bounded as

$$\begin{aligned} J &= \left\|\left(\mathbf{G}^{-1} - \mathbf{G}_L^{-1}\right)\mathbf{y}^{\mathrm{MF}}\right\|_2 \\ &= \left\|\sum_{l=L}^{\infty}\left(\mathbf{S}^{-1}\left(\mathbf{S}-\mathbf{G}\right)\right)^l \mathbf{S}^{-1}\mathbf{y}^{\mathrm{MF}}\right\|_2 \\ &= \left\|\left(\mathbf{S}^{-1}\left(\mathbf{S}-\mathbf{G}\right)\right)^L \mathbf{G}^{-1}\mathbf{y}^{\mathrm{MF}}\right\|_2 \\ &\leqslant \|\mathbf{B}\|_{\mathrm{F}}^L \left\|\hat{\mathbf{x}}\right\|_2, \end{aligned} \tag{5.26}$$

[5]The residual estimation error is consistent with the definition in [23], which is to evaluate the difference between the linear MMSE estimation and the approximate estimation using truncated Neumann series expansion.

where the inequality holds since $\|\mathbf{Ax}\|_2 \leqslant \|\mathbf{A}\|_{\mathrm{F}} \|\mathbf{x}\|_2$. As shown in Appendix B, when $N_B \to \infty$, $\Pr\left\{\|\mathbf{B}\|_{\mathrm{F}}^2 < 1\right\} \to 1$ and $\mathrm{E}\left\{\|\mathbf{B}\|_{\mathrm{F}}^2\right\} \to 0$. That is to say, the residual estimation error will approach 0 as indicated by inequality (5.26). Inequality (5.26) also indicates that increasing the truncation order in Neumann series expansion, the upper bound of the residual estimation error can be reduced. This evidence, together with the high probability that the convergence conditions will be met, supports the applications of the stair matrix to massive MIMO systems.

5.3 Implementation of the Stair Matrix in Iterative Method

Due to the involvement of matrix multiplications, the truncation order in Neumann series expansion is limited to three; otherwise, the computational complexity is comparable with matrix inversion algorithm. Besides, we note that in existing work, the computation of the LLR is obtained by utilizing the NPI after the first truncation order in Neumann series expansion (or first iteration in iterative method). This implementation, however, causes significant performance loss when N_B is not sufficiently large (or $r = N_B/N_U$ is not large, for example, $r < 8$). In this Section, we address these issues by developing an iterative method using stair matrix.

5.3.1 Stair Matrix in Iterative Method

Compared to the linear ZF detection, linear MMSE detection achieves a better balance in consideration of interference and noise. Therefore, we will introduce an iterative method to approach the linear MMSE detection.

To start with, we extract the stair matrix $\mathbf{S} = \mathrm{stair}\left(\mathbf{W}_{(u,u-1)}, \mathbf{W}_{(u,u)}, \mathbf{W}_{(u,u+1)}\right)$. It is worth noting that compared to the stair matrix we discussed in previous section, the diagonal elements in the new stair matrix has increased by σ_z^2 according to Equation (5.3),

which brings negligible computational cost. According to Equation (5.17), we have

$$\begin{aligned}\mathbf{x}^{(i+1)} &= \mathbf{S}^{-1}\left((\mathbf{S}-\mathbf{W})\,\mathbf{x}^{(i)}+\mathbf{y}^{\mathrm{MF}}\right)\\ &= \mathbf{x}^{(i)}-\mathbf{S}^{-1}\mathbf{W}\mathbf{x}^{(i)}+\mathbf{S}^{-1}\mathbf{y}^{\mathrm{MF}},\end{aligned} \tag{5.27}$$

where $\mathbf{x}^{(i)}$ is the i-th estimation.

In accordance, if the initial estimation $\mathbf{x}^{(0)}$ is selected as

$$\mathbf{x}^{(0)} = \mathbf{S}^{-1}\mathbf{y}^{\mathrm{MF}}, \tag{5.28}$$

following the iterative process in Equation (5.27), we can derive

$$\mathbf{x}^{(i)} = \sum_{l=0}^{i}\left(\mathbf{S}^{-1}\left(\mathbf{S}-\mathbf{W}\right)\right)^{l}\mathbf{S}^{-1}\mathbf{y}^{\mathrm{MF}}, \tag{5.29}$$

which indicates that the iterative method in Equation (5.27) can be seen as the truncated Neumann series expansion based method by selecting $\mathbf{X} = \mathbf{S}$. However, in Equation (5.27), only a matrix-vector product is involved, hence the overall computational complexity is of the order $O\left(KN_U^2\right)$, where K denotes the number of iterations.

5.3.2 Computation of the LLR

After the estimation of transmitted vector $\mathbf{x}$, we need to compute the LLRs of the associated bits for the soft-input channel decoder. After K iterations, the equivalent channel gain $\rho_u^{(K)}$ and the covariance of the NPI $\left|v_u^{(K)}\right|^2$ can be respectively given by

$$\rho_u^{(K)} = \mathbf{e}_u^{\mathrm{H}}\mathbf{W}_K^{-1}\mathbf{G}\mathbf{e}_u, \tag{5.30}$$

$$\left|v_u^{(K)}\right|^2 = \mathbf{e}_u^{\mathrm{H}}\mathbf{W}_K^{-1}\mathbf{G}\mathbf{G}\mathbf{W}_K^{-1}\mathbf{e}_u + \sigma_z^2\mathbf{e}_u^{\mathrm{H}}\mathbf{W}_K^{-1}\mathbf{G}\mathbf{W}_K^{-1}\mathbf{e}_u - \left|\rho_u^{(K)}\right|^2 \tag{5.31}$$

Apparently, Equations (5.30) and (5.31) requires matrix multiplications if $K \geq 2$. Therefore, in [23, 84–86], $\mathbf{D}^{-1}$, where $\mathbf{D} = \mathrm{diag}\left\{\left[\mathbf{W}_{(0,0)}, \mathbf{W}_{(1,1)}, \cdots, \mathbf{W}_{(N_U-1,N_U-1)}\right]^{\mathrm{T}}\right\}$ is considered for the simplification. This approximation, however, as we will show in the next section, has caused a significant performance loss.

As we can see from Equation (5.8), the exact *a posteriori* covariance of the NPI in linear MMSE estimation can be derived if the equivalent channel gain is obtained. However, in [23], the authors have claimed that this relationship is not supported in the truncated Neumann series expansion. The main reason for that claim is attributed to the fact that $\mathbf{W}_K^{-1}$ is far away from $\mathbf{W}^{-1}$ with small K. In the previous section, we introduced the iterative method for detection, and the iteration numbers can be sufficiently large since the computational complexity in one iteration is of the order $O\left(N_U^2\right)$. With sufficiently large iterations, $\mathbf{W}_K^{-1}$ can be quite close to $\mathbf{W}^{-1}$ (we will show this in the next section); hence, we can used Equation (5.8) to derive the covariance of the NPI. The next question is how to maintain low computational complexity to obtain the equivalent channel gain.

We rewrite the equivalent channel gain in Equation (5.8) as $\rho_u = \mathbf{e}_u^{\mathrm{H}}\mathbf{W}^{-1}\mathbf{G}\mathbf{e}_u = 1 - \sigma_z^2\mathbf{e}_u^{\mathrm{H}}\mathbf{W}^{-1}\mathbf{e}_u$. In addition, we replace $\mathbf{W}^{-1}$ with $\mathbf{W}_K^{-1}$, leading to

$$\rho_u^{(K)} = 1 - \sigma_z^2\mathbf{e}_u^{\mathrm{H}}\mathbf{W}_K^{-1}\mathbf{e}_u. \tag{5.32}$$

That is to say, we need obtain the diagonal elements in $\mathbf{W}_K^{-1}$ to compute $\rho_u^{(K)}$.

If N_B and r are sufficiently large, the Gram matrix $\mathbf{G}$ and $\mathbf{W}$ will become diagonal dominant; therefore, $\mathbf{D}^{-1}$ can be a good approximation of $\mathbf{W}^{-1}$, and we have the approximation to $\rho_u^{(K)}$ given by

$$\rho_u^{(K)} \approx 1 - \sigma_z^2\mathbf{D}_{(u,u)}^{-1}, \tag{5.33}$$

and $\left|v_u^{(K)}\right|^2$ is approximated as

$$\left|v_u^{(K)}\right|^2 \approx \rho_u^{(K)}\left(1 - \rho_u^{(K)}\right). \tag{5.34}$$

As a consequence, the *a posteriori* SINR is approximated as

$$\gamma_u^{(K)} \approx \frac{\left|\rho_u^{(K)}\right|^2}{\left|v_u^{(K)}\right|^2} = \frac{\rho_u^{(K)}}{1 - \rho_u^{(K)}}. \tag{5.35}$$

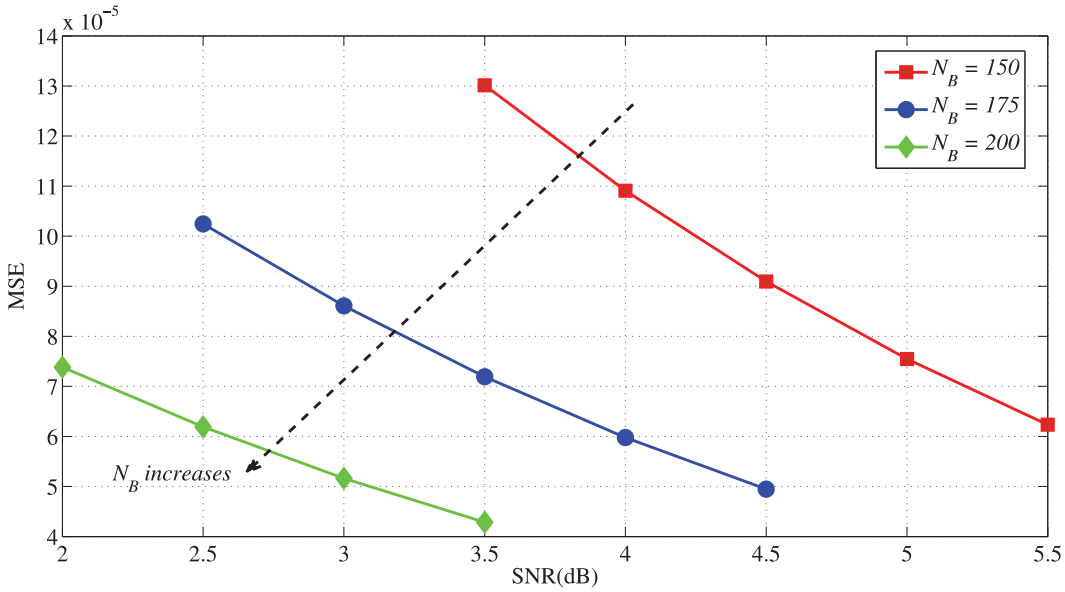


Figure 5.1: The normalized mean-square error for the approximation in Equation (5.33), $N_U = 25$

$\rho_u^{(K)}$ and $\gamma_u^{(K)}$ are used in Equation (5.9) to compute $L\left(b_{u,k}\right)$.

The accuracy of the approximation in Equation (5.33) is important for the overall system performance. In order to show the accuracy of Equation (5.33), we define

$$MSE = \mathrm{E}\left\{\frac{1}{N_U}\sum_{u=0}^{N_U-1}\left|\rho_u^{(K)} - \rho_u\right|^2\right\}, \tag{5.36}$$

to show the normalized mean-square error between the approximation $\rho_u^{(K)}$ given by Equation (5.33) and the exact equivalent channel gain ρ_u given by Equation (5.6). The results are shown in Figure 5.1. The results are obtained by randomly generating 2000000 channel realization, and ρ_u is computed by Equation (5.6), while the approximation is given by Equation (5.33). It is clear from Figure 5.1 that the approximation in Equation (5.33) is quite close to the exact ρ_u as MSE is at very low level. Besides, Figure 5.1 also indicates that with the increase of N_B or average SNR at receiver, the approximation will be closer. These results indicate our approximation given by Equation (5.33) works well in approach-

Table 5.2: **Algorithm 5.2:** Proposed Iterative Method Using Stair Matrix

---

Input: $\mathbf{y}$, $\mathbf{H}$, σ_z^2, and Iteration number K;

Output: LLRs of the associated bits $L\left(b_{u,k}\right)$.

Initialization:

1. $\mathbf{G} = \mathbf{H}^{\mathrm{H}}\mathbf{H}$, $\mathbf{W} = \mathbf{G} + \sigma_z^2\mathbf{I}_{N_U}$, $\mathbf{y}^{\mathrm{MF}} = \mathbf{H}^{\mathrm{H}}\mathbf{y}$;
2. $\mathbf{S} = \mathrm{stair}\left(\mathbf{W}_{(u,u-1)}, \mathbf{W}_{(u,u)}, \mathbf{W}_{(u,u+1)}\right)$;
3. Compute $\mathbf{S}^{-1}$ through **Algorithm 5.1**, and $\mathbf{D}^{-1} = \mathrm{diag}\left(\mathbf{S}^{-1}\right)$;
4. Initial estimation: $\mathbf{x}^{(0)} = \mathbf{S}^{-1}\mathbf{y}^{\mathrm{MF}}$;

Iteration:

5. for $i = 1:1:K$
6. $\quad \mathbf{x}^{(i+1)} = \mathbf{S}^{-1}\left(\left(\mathbf{S}-\mathbf{W}\right)\mathbf{x}^{(i)} + \mathbf{y}^{\mathrm{MF}}\right)$;
7. end

LLR Computation:

8. $\rho_u^{(K)} = 1 - \sigma_z^2\mathbf{D}_{(u,u)}^{-1}$, $\gamma_u^{(K)} = \frac{\rho_u^{(K)}}{1-\rho_u^{(K)}}$;
9. $L\left(b_{u,k}\right) = \gamma_u^{(K)}\left(\min\limits_{s\in\chi_k^0}\left|\frac{\hat{x}_u^{(K)}}{\rho_u^{(K)}} - s\right|^2 - \min\limits_{s'\in\chi_k^1}\left|\frac{\hat{x}_u^{(K)}}{\rho_u^{(K)}} - s'\right|^2\right)$.

Return $L\left(b_{u,k}\right)$.

---

ing the linear MMSE detection method. We will also validate this result in next section in BER evaluation.

It is worth pointing out that although we utilize the diagonal matrix to estimate the equivalent channel gain, the computation of $\gamma_u^{(K)}$ in Equation (5.35) indicates that we try to approach the SINR in linear MMSE detection to derive the LLRs of the associated bits. This is quite different from the existing work [23, 84–86], where the SINR after the first iteration (or the first truncation order in Neumann series expansion method) is adopted. In

fact, as the iterations increase, the covariance of the NPI will decrease, and our proposed approximation method is more efficient and accurate. In numerical simulations, we also validate that our approximation in (5.33) and (5.35) outperforms the approximation in existing work.

To summarize, we present **Algorithm 5.2** for the proposed iterative method using stair matrix.

5.3.3 Computational Complexity Analysis

We consider the number of real number multiplications/divisions to evaluate the computational complexity. In initialization steps, the computation of $\mathbf{W}$ and $\mathbf{y}^{\mathrm{MF}}$ requires $2N_BN_U^2$ and $4N_BN_U$ real number multiplications, respectively. According to **Algorithm 5.1**, the computation of $\mathbf{S}^{-1}$ requires $3\left(N_U-1\right)$ real number multiplications and N_U real number divisions. The initial estimation, provided in Step 4 of **Algorithm 5.2**, requires $\frac{N_U+1}{2}\times 2+\frac{N_U-1}{2}\times(8+2)=6N_U-4$ real number multiplications. Therefore, the total computational complexity in initialization steps is $2N_BN_U^2+4N_BN_U+10N_U-7$.

The iteration steps in **Algorithm 5.2** involves matrix-vector production. The computation of $(\mathbf{S}-\mathbf{W})\,\mathbf{x}^{(i)}$ requires $\frac{N_U+1}{2}\times 4\left(N_U-1\right)+\frac{N_U-1}{2}\times 4\left(N_U-3\right)=4\left(N_U-1\right)^2$ real number multiplications. The resultant vector is multiplied by a stair matrix, and additional $6N_U-4$ real number multiplications are required. Therefore, the total computational complexity in iteration steps is $K\left(4N_U^2-2N_U\right)$. That is to say, the computational complexity of the iterative process is of $O\left(N_U^2\right)$, which is the same as the existing iterative methods where the diagonal matrix is applied.

Last, to obtain $L\left(b_{k,u}\right)$, we need the computation of $\rho_u^{(K)}$, and the proposed approximation method only requires the diagonal elements in $\mathbf{D}$, which is obtained in step 3. Compared to the existing work in [23, 84–86], our proposed scheme saves computational

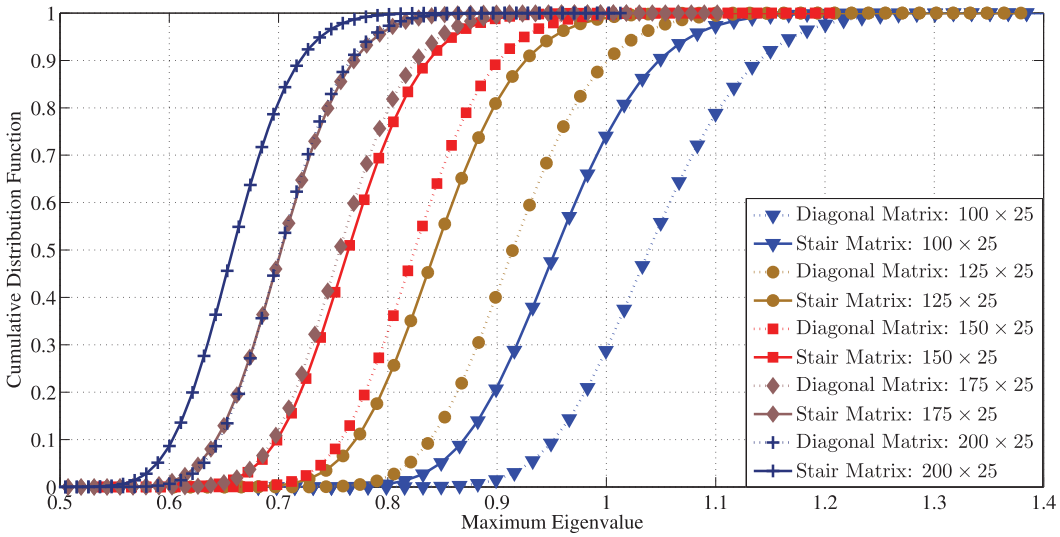


Figure 5.2: Cumulative distribution function of the maximum eigenvalue $N_U = 25$.

complexity in this stage.

To summarize, the overall computational complexity is the same level of the existing work [23, 84–86]. However, as we will see in next Section, the stair matrix outperforms the diagonal matrix in the evaluated performance metrics.

5.4 Numerical Simulations and Performance Evaluation

5.4.1 Convergence Conditions

We first investigate the convergence condition using the stair matrix. Using the Monte-Carlo method, we generate 2e7 random channel matrices $\mathbf{H}$. For each $\mathbf{H}$, we extract the diagonal matrix $\mathbf{D}$ and the stair matrix $\mathbf{S}$, and compute the maximum eigenvalues of the matrix $\mathbf{I} - \mathbf{D}^{-1}\mathbf{G}$, and $\mathbf{I} - \mathbf{S}^{-1}\mathbf{G}$, respectively. Using numerical simulations, we eventually obtain the cumulative distribution function (CDF) of the maximum eigenvalues, given by Figure 5.2. In Figure 5.2, we evaluate the scenario that 25 users are in service and we increase the number of antennas at base station from 100 to 200. The following observations can be

found:

- With the increase of antennas at base station, the probability that the convergence conditions are met, i.e., $\Pr\left\{\rho\left(\mathbf{I}-\mathbf{S}^{-1}\mathbf{G}\right)<1\right\}$ and $\Pr\left\{\rho\left(\mathbf{I}-\mathbf{D}^{-1}\mathbf{G}\right)<1\right\}$ will increase. Specifically, for the usage of the diagonal matrix, the probability that the convergence conditions are met, is increase from 0.29 when $N_B = 100$, to 1 when $N_B = 200$. In accordance, for the usage of the stair matrix, $\Pr\left\{\rho\left(\mathbf{I}-\mathbf{S}^{-1}\mathbf{G}\right)<1\right\}$ is increased from 0.74 when $N_B = 100$, to 1 when $N_B = 200$;

- In low $r = N_B/N_U \leq 5$ region, the usage of the stair matrix can increase the probability that the convergence conditions are met. For example, when $N_B = 100$, $\Pr\left\{\rho\left(\mathbf{I}-\mathbf{D}^{-1}\mathbf{G}\right)<1\right\}$ is only 0.29, while $\Pr\left\{\rho\left(\mathbf{I}-\mathbf{S}^{-1}\mathbf{G}\right)<1\right\}$ becomes 0.76. This indicates that in some low r region, the diagonal matrix is not applicable while the stair matrix can be used;

- In any system configuration, $\Pr\left\{\rho\left(\mathbf{I}-\mathbf{S}^{-1}\mathbf{G}\right)<a\right\} \geqslant \Pr\left\{\rho\left(\mathbf{I}-\mathbf{D}^{-1}\mathbf{G}\right)<a\right\}$, $a \in (0,1)$. As the maximum eigenvalue determines the convergence rate, we can conclude that by using the stair matrix, the convergence rate is more likely faster compared to the use of the diagonal matrix.

The above observations validate the applicability of the usage of the stair matrix and diagonal matrix in massive MIMO systems. Besides, the results reveal that by using stair matrix, we can increase the probability that the convergence conditions are met in low r region compared to the usage of the diagonal matrix. In high r region, it is also shown that the probability that the convergence conditions are satisfied is almost 1 using both the stair matrix and the diagonal matrix. This is consistent with the results in [23] where the applicability of the diagonal matrix is demonstrated. It is also revealed in [80, 97] that in high r region, the probability that the maximum eigenvalue less than 1 is near 1.

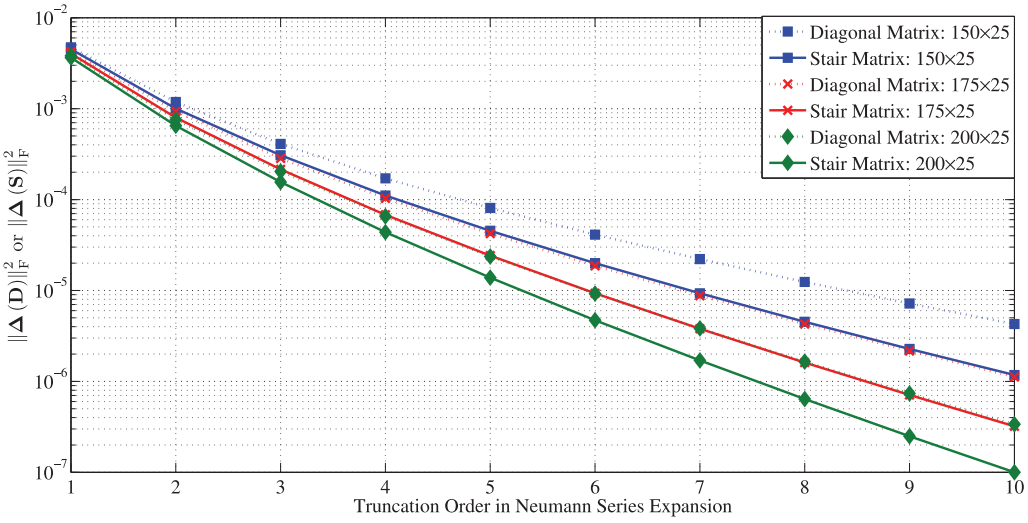


Figure 5.3: Normalized mean-square error for the matrix inverse approximation

Furthermore, we also find that by using the stair matrix, the convergence rate is more likely accelerated than with the use of the diagonal matrix.

5.4.2 Matrix Inverse

We now investigate the performance of the stair matrix in Neumann series expansion to approach the matrix inverse[6]. We define $\mathbf{\Delta}(\mathbf{S}) = \left(\mathbf{I} - \sum_{l=0}^{L-1}\left(\mathbf{I}-\mathbf{S}^{-1}\mathbf{G}\right)^l\mathbf{S}^{-1}\mathbf{G}\right)$, where $\mathbf{S} = \text{stair}\left(\mathbf{G}_{u,u-1}, \mathbf{G}_{u,u}, \mathbf{G}_{u,u+1}\right)$, to show the error matrix between the identity matrix and the corresponding matrix multiplication when the approximate matrix inverse is adopted. Accordingly, $\mathbf{\Delta}(\mathbf{D}) = \left(\mathbf{I} - \sum_{l=0}^{L-1}\left(\mathbf{I}-\mathbf{D}^{-1}\mathbf{G}\right)^l\mathbf{D}^{-1}\mathbf{G}\right)$ where $\mathbf{D} = \text{diag}\left(\left[\mathbf{G}_{(0,0)}, \mathbf{G}_{(1,1)}, \cdots, \mathbf{G}_{(N_U-1,N_U-1)}\right]^{\text{T}}\right)$. In addition, we have $\text{E}\{\frac{1}{N_U^2}\|\mathbf{\Delta}(\mathbf{S})\|_{\text{F}}^2\}$ and $\text{E}\{\frac{1}{N_U^2}\|\mathbf{\Delta}(\mathbf{D})\|_{\text{F}}^2\}$ to indicate the normalized mean-square error[7] for the approximation using the stair matrix and the diagonal matrix, respectively. With different truncation order, we present the results in Figure 5.3. The

[6]In implementation, we propose the iterative method as shown in section IV. However, the results of the iterative method can be seen as the Neumann series expansion.

[7]The normalized mean-square error here is consistent with the definition in [81]. This term is defined to evaluate the difference between the approximate and exact matrix inverse.

following observations can be found:

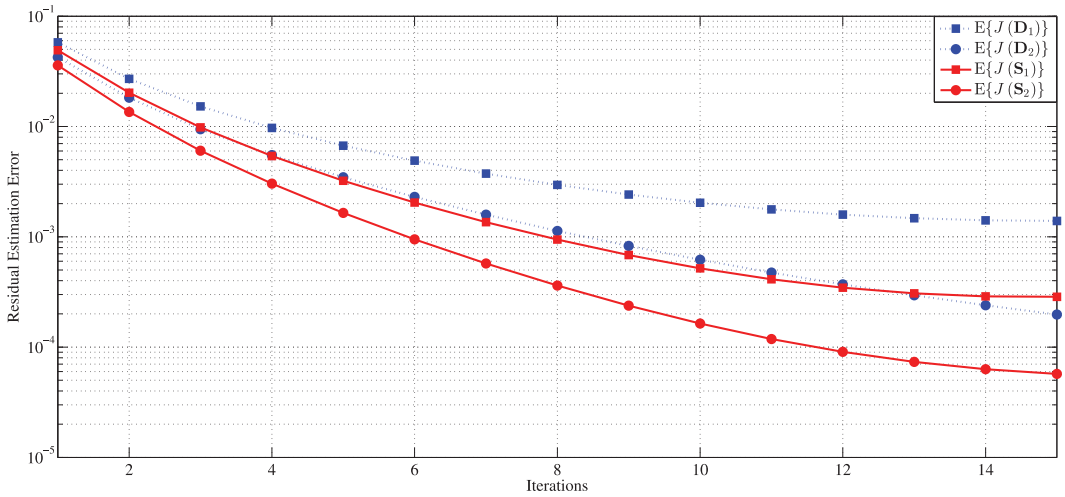
- With the increase of the truncation order, the normalized mean-square error is decreased. This indicates that the more truncation orders used in Neumann series expansion, the closer of the resulting approximation is to the actual matrix inverse;
- By using the stair matrix, the normalized mean-square error is always less than that of using the diagonal matrix in the same system configuration. This indicates that the use of the stair matrix always achieves better approximation performance with the same truncation order compared to the use of the diagonal matrix;
- By using the stair matrix, fewer iterations are required to achieve the same level of the normalized mean-square error than using the diagonal matrix. As the truncation order is equivalent to the iteration number in the iterative method, the fewer iterations indicate lower computational complexity in implementation.

To summarize, we conclude that the use of the stair matrix outperforms the use of the diagonal matrix. As we showed in section IV.A, the truncation order is equivalent to the iterations in the iterative method; therefore, the results in Figure 5.3 help to interpret the convergence performance of the proposed iterative method.

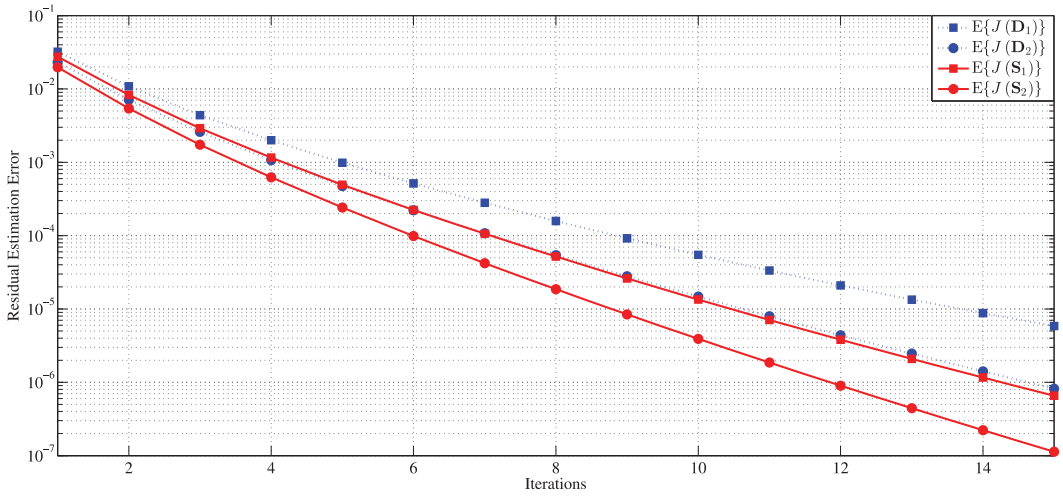
5.4.3 Residual Estimation Error

In the iterative method, the estimation is to approach the estimation vector in linear ZF/MMSE method. In section III.C, an upper bound of the residual estimation error[8] for the use of the stair matrix in approaching linear ZF detection is presented. In order

[8]We have shown in Section IV.A that the iterative estimation is equivalent to the estimation using the truncated Neumann series expansion; therefore, we use the term, residual estimation error, defined in [23] to evaluate the difference between the linear MMSE estimation and the iterative estimation.



(a)



(b)

Figure 5.4: Residual Estimation Error: (a) N_B = 150, N_U = 25, average SNR= 5dB; (b) N_B = 200, N_U = 25, average SNR= 3.5dB

to differentiate the residual estimation error for the use of stair matrix and the diagonal

matrix in linear ZF and MMSE detection, we define the following terms:

$$J(\mathbf{D}_1) = \left\| \left(\mathbf{D}_1^{-1} (\mathbf{D}_1 - \mathbf{G}) \right)^L \mathbf{G}^{-1} \mathbf{y}^{\mathrm{MF}} \right\|_2,$$

$$J(\mathbf{D}_2) = \left\| \left(\mathbf{D}_2^{-1} (\mathbf{D}_2 - \mathbf{W}) \right)^L \mathbf{W}^{-1} \mathbf{y}^{\mathrm{MF}} \right\|_2,$$

$$J(\mathbf{S}_1) = \left\| \left(\mathbf{S}_1^{-1} (\mathbf{S}_1 - \mathbf{G}) \right)^L \mathbf{G}^{-1} \mathbf{y}^{\mathrm{MF}} \right\|_2,$$

$$J(\mathbf{S}_2) = \left\| \left(\mathbf{S}_2^{-1} (\mathbf{S}_2 - \mathbf{W}) \right)^L \mathbf{W}^{-1} \mathbf{y}^{\mathrm{MF}} \right\|_2,$$

where $\mathbf{W} = \mathbf{G} + \sigma_z^2 \mathbf{I}_{N_U}$, and

$$\mathbf{D}_1 = \mathrm{diag}\left\{ \left[\mathbf{G}_{(0,0)}, \mathbf{G}_{(1,1)}, \cdots, \mathbf{G}_{(N_U-1,N_U-1)} \right]^{\mathrm{T}} \right\},$$

$$\mathbf{D}_2 = \mathrm{diag}\left\{ \left[\mathbf{W}_{(0,0)}, \mathbf{W}_{(1,1)}, \cdots, \mathbf{W}_{(N_U-1,N_U-1)} \right]^{\mathrm{T}} \right\},$$

$$\mathbf{S}_1 = \mathrm{stair}\left(\mathbf{G}_{m,m-1}, \mathbf{G}_{m,m}, \mathbf{G}_{m,m+1} \right),$$

$$\mathbf{S}_2 = \mathrm{stair}\left(\mathbf{W}_{m,m-1}, \mathbf{W}_{m,m}, \mathbf{W}_{m,m+1} \right),$$

According to Equation (5.26), we can see that $J(\mathbf{D}_1)$ and $J(\mathbf{D}_2)$ denote the residual estimation error for the use of the diagonal matrix in approaching linear ZF and MMSE detection, respectively. $J(\mathbf{S}_1)$ and $J(\mathbf{S}_2)$ denote the residual estimation error for the use of the stair matrix in approaching linea ZF and MMSE detection, respectively. For a given system configuration and average receiving SNR, we present the residual estimation error performance in Figure 5.4. The results are present with 2000000 randomly generated channel realizations, and the following observations are found:

- From Figure 5.4(a) and 5.4(b), $\mathrm{E}\{J(\mathbf{S}_1)\}$ is always less than $\mathrm{E}\{J(\mathbf{D}_1)\}$, and $\mathrm{E}\{J(\mathbf{S}_2)\}$ is always less than $\mathrm{E}\{J(\mathbf{D}_2)\}$ after the same iteration numbers. These results reflect that after the same iterations, using the stair matrix in the iterative method can approach both the linear ZF and MMSE estimation more closely compared to the use of the diagonal matrix;

- In Figure 5.4(a), we note that, for the use of the diagonal matrix, the residual estimation error decreases slowly and remains a comparatively high level even with large iteration numbers. However, by using the stair matrix, we can speed up the decreasing rate and achieve a comparatively lower estimation error level. These results are consistent with the previous numerical results where we demonstrate that the use of the diagonal matrix may not be applicable in low r ratio.

- From Figure 5.4(a) and Figure 5.4(b), we can see that, with the increase of the receiving antennas at the base station, the performance gain with the use of the stair matrix over the use of the diagonal matrix becomes small. These results are reasonable as N_B increases, $\mathbf{G}$ and $\mathbf{W}$ both become diagonal dominant. However, we can also achieve comparatively lower residual estimation error by using the stair matrix in the iterative method.

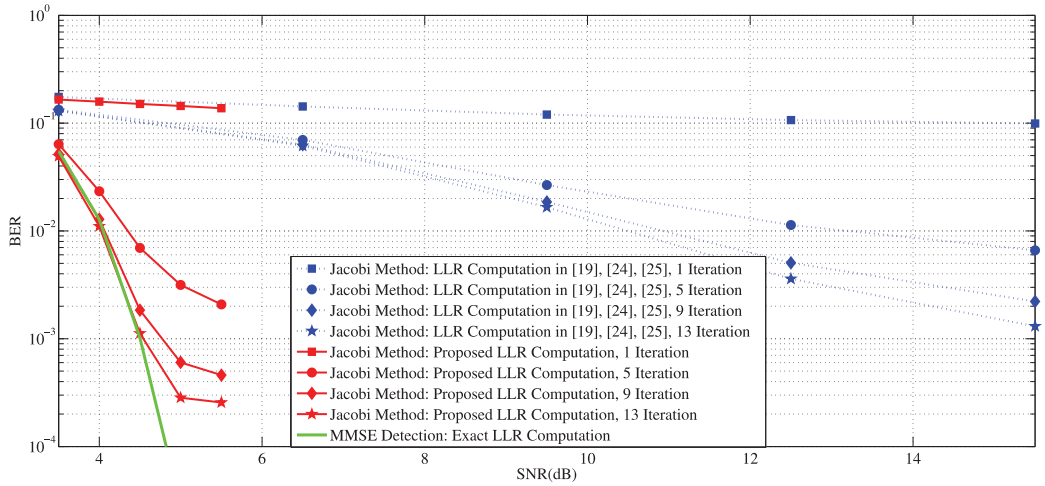
To summarize, we conclude that the use of the stair matrix outperforms the use of the diagonal matrix in terms of the residual estimation error. The performance gain is more significant in low r ratio, but still obvious in high r ratio.

5.4.4 BER Performance

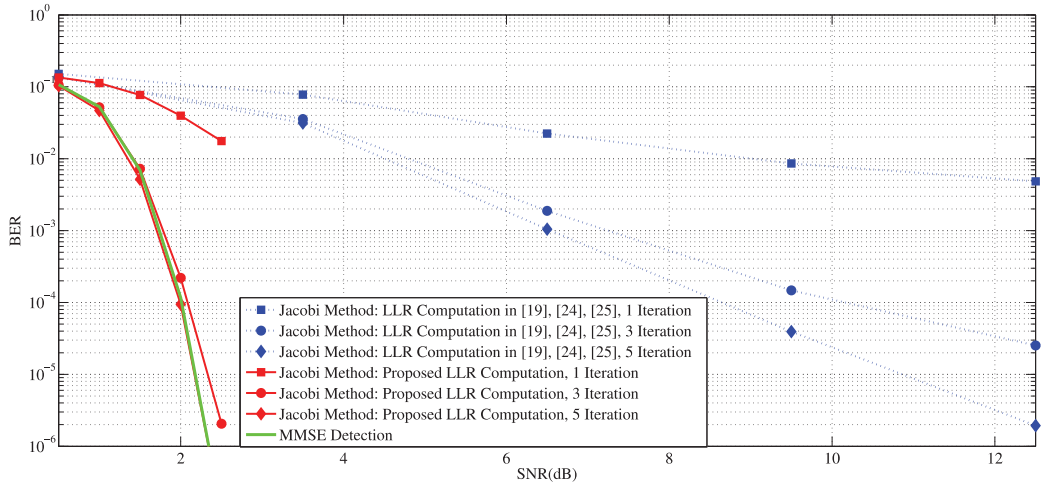
We now evaluate the system BER performance. In the system, the base station is simultaneously serving $N_U = 25$ users. For each user, a LDPC code with code length 64800, code rate 1/2 is considered for channel code scheme[9]. We consider 64QAM modulation, and a block independent channel is considered for the evaluation.

To begin with, we investigate the proposed LLR computation given by (5.35), and the equivalent channel gain ρ_u and the covariance of the NPI v_u are approximated by (5.33) and (5.34). For comparison, we provide the linear MMSE detection as a benchmark, where

[9]LDPC code has been an agreed standard for long code in 5G



(a)



(b)

Figure 5.5: BER performance: (a) $N_B = 150$, $N_U = 25$; (b) $N_B = 250$, $N_U = 25$. LDPC code: code length 64800, code rate 0.5. 64QAM modulation.

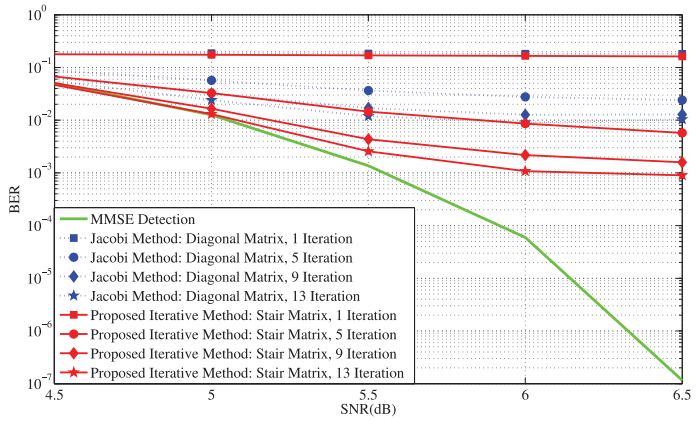
the LLR computation is given by Equation (5.9) with ρ_u and v_u given by Equation (5.6) and (5.8), respectively. The LLR computation in existing work such as [23, 84, 85] is to compute the covariance of the NPI after the first iteration. It is worth pointing out that the iterative methods in [84, 85] requires fewer iterations to approach the linear MMSE

detection; however, the LLR computation used in MMSE detection is not computed from the exact NPI of the MMSE detection, but the NPI after the first iteration. In Figure 5.5, we can see that the BER performance of the Jacobi method with the LLR computation in [23, 84, 85] is far away from the BER performance of the MMSE detection with the exact LLR computation. This is consistent with our previous analysis, where we pointed out that the covariance of the NPI will decrease with iterations. However, we note that the proposed LLR computation can greatly improve the BER performance of the Jacobi method by approximating the covariance of the NPI of the MMSE detection. Hereafter, we only utilize the proposed LLR computation for the BER performance comparison.

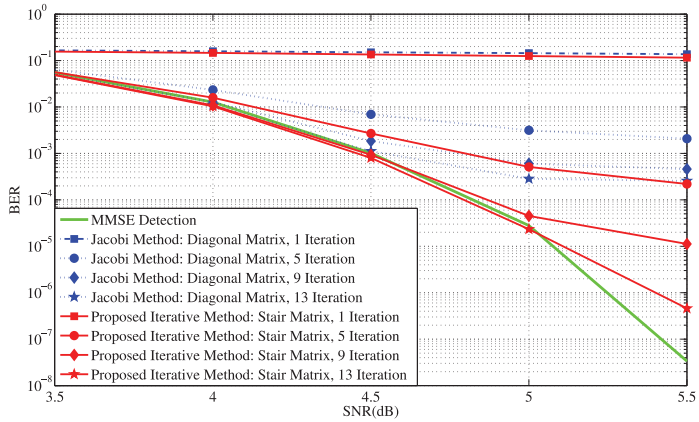
We now present the results with low $r = N_B/N_U$ region, and the results are presented in Figure 5.6. The following observations are found.

- From Figure 5.6(a), we note that the BER performance improvement with the proposed stair matrix compared to the diagonal matrix is obvious. However, the system performance is still far away from the MMSE detection even with sufficiently large iterations. Specially, for the use of the diagonal matrix, the performance levels off after 9 iterations; for the use of the stair matrix, the performance is greatly improved, but a leveled off performance still appears. These are attributed to the slow convergence rate and convergence conditions not 100 percent satisfied;

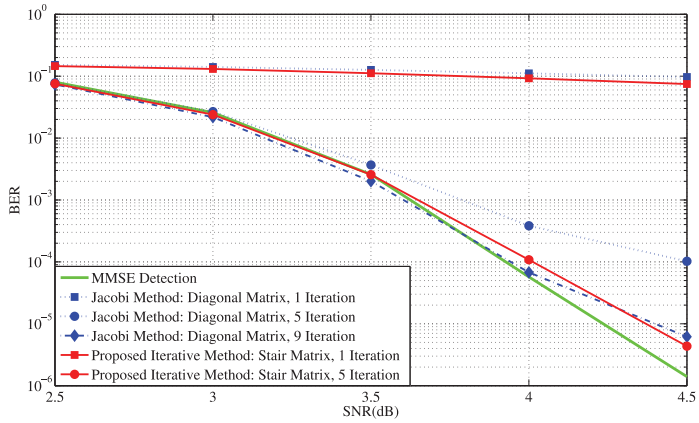
- From Figure 5.6(b) and Figure 5.6(c), we can see that the BER performance eventually converges to the performance of the MMSE detection. Specifically, in the system configuration $N_B = 150$, $N_U = 25$, at SNR= $5dB$, the BER performance of the proposed iterative method after 13 iterations is almost the same as the performance of the MMSE detection. In the system configuration $N_B = 175$, $N_U = 25$, at SNR= $4dB$, the BER performance of the Jacobi method after 9 iterations approaches the performance of the MMSE detection;



(a)



(b)



(c)

Figure 5.6: BER performance: (a) $N_B = 125$, $N_U = 25$; (b) $N_B = 150$, $N_U = 25$; (c) $N_B = 175$, $N_U = 25$. LDPC code: code length 64800, code rate 0.5. 64QAM modulation.

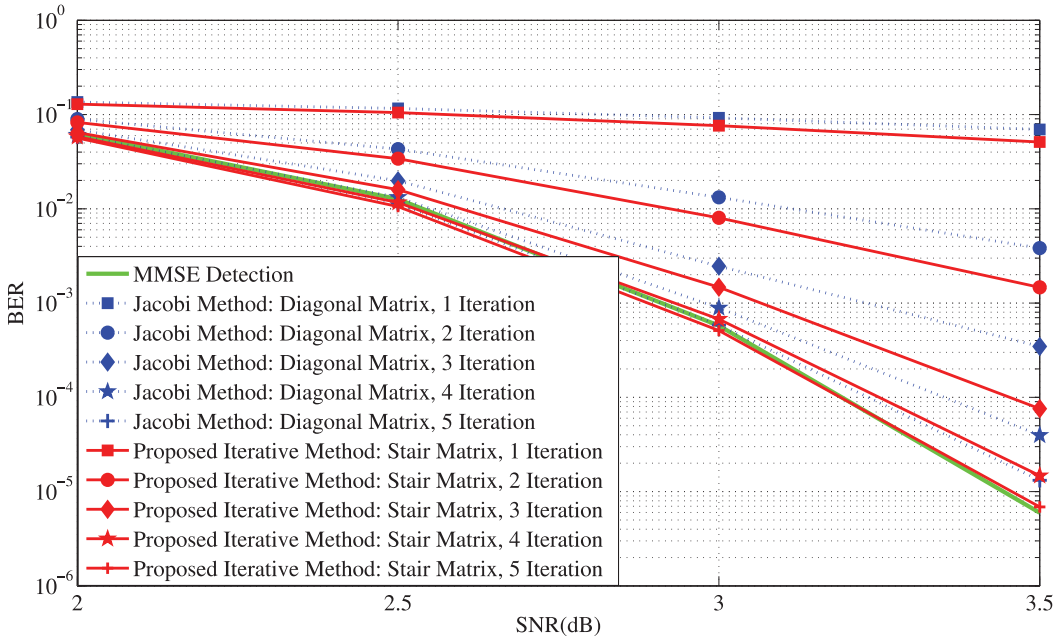
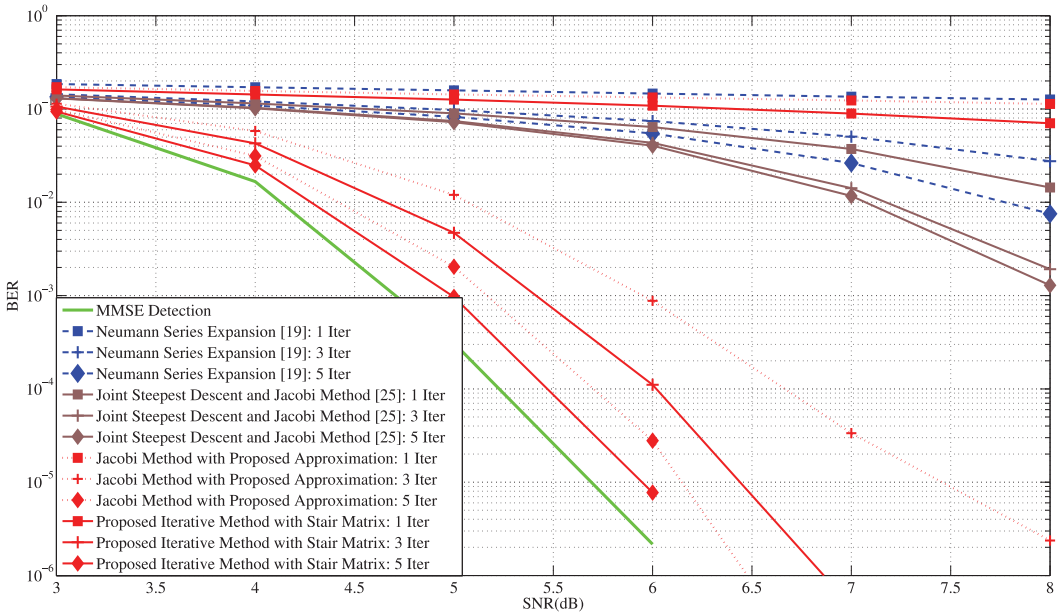


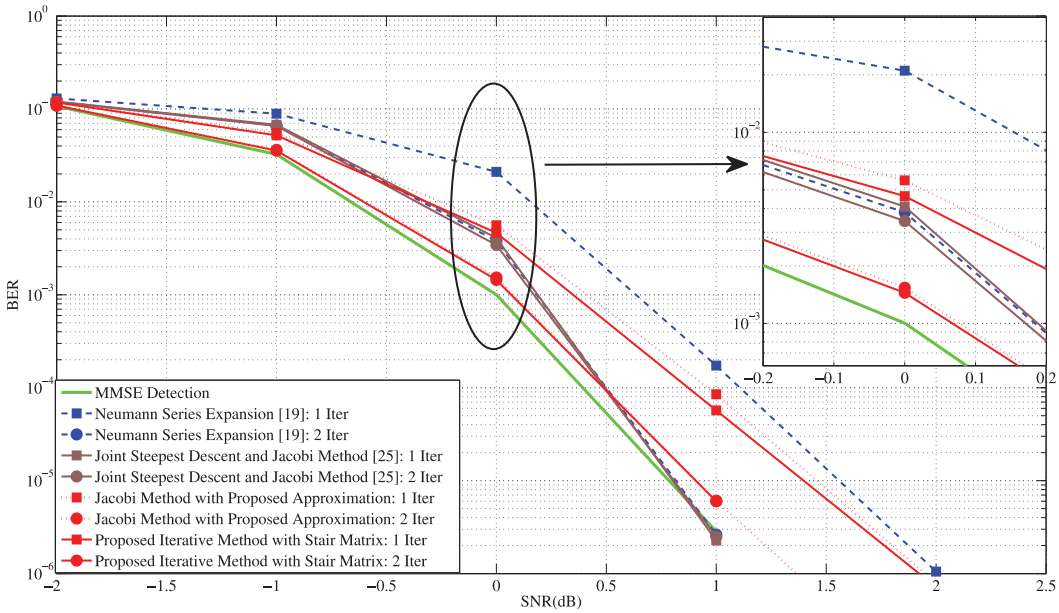
Figure 5.7: BER performance: $N_B = 200$, $N_U = 25$. LDPC code: code length 64800, code rate 0.5. 64QAM modulation.

- From Figure 5.6(a) to Figure 5.6(c), we can see that the convergence rate of the proposed iterative method is faster than that of the Jacobi method. These results are consistent with the previous analysis. With faster convergence rate, fewer iterations are required for the proposed iterative method, hence reducing the overall system computational complexity.

Next, we evaluate the BER performance in the system configuration with high $r = N_B/N_U$ region, and the results are shown in Figure 5.7. It is clear that both the uses of the diagonal matrix and stair matrix require few iterations to converge. However, as indicated by the cumulative distribution function of the maximum eigenvalue, $\Pr\left\{\rho\left(\mathbf{I}-\mathbf{S}^{-1}\mathbf{G}\right)<a\right\} \geqslant \Pr\left\{\rho\left(\mathbf{I}-\mathbf{D}^{-1}\mathbf{G}\right)<a\right\}$, $a \in (0,1)$, we can conclude that the convergence rate of the proposed iterative method using the stair matrix is faster than that of the Jacobi method using the diagonal matrix. The results validate these conclusions.



(a)



(b)

Figure 5.8: BER performance: (a) $N_B = 150$, $N_U = 25$; (b) $N_B = 400$, $N_U = 25$. LDPC code: code length 1152, code rate 0.5. 64QAM modulation.

In order to show the applications of the proposed scheme in a practical system, we evaluate the system BER performance in an OFDM system where the extended vehicular A (EVA) channels are used to generate the channel data for each user [39,99]. We assume the channel data remains constant during one OFDM symbol block, and varies independently from one block to another [51, 73]. For each user, the LDPC encoder with code length 1152, code rate 0.5, is adopted. We consider 64QAM modulation, and all users take up 192 subcarriers for data transmission. As we mentioned in Section II, the received signal model is established at each subcarrier. Therefore, the estimation is performed over individual subcarrier. We also provide the existing proposals in [23, 85] for comparison. The results are presented in Figure 5.8, with the following observations found:

- In Figure 5.8(a), the number of antennas at the base station is 150, and the number of user under service is 25. In this comparatively low r region, we can see that the performance of existing proposals is far away from that of the exact linear MMSE detection scheme. This is obvious as we pointed out previously that the NPI in the initial estimation is far away from the exact NPI in linear MMSE detection. However, both the Jacobi method with the diagonal matrix in the development and the proposed iterative method with the stair matrix can eventually approach the performance of the linear MMSE detection scheme. In addition, we can see that the BER performance using the stair matrix is always much better than that using the diagonal matrix, which is consistent with our previous demonstration;

- In Figure 5.8(b), N_B = 400 and N_U = 25. In this scenario, r = 16, which falls in the application region of Neumann series expansion based detection scheme [23]. We can see that after two iterations, the performance of the Neumann series expansion based detection scheme converges to linear MMSE detection even with the NPI in the initial estimation for LLR computation. These results are consistent with the observations

in [23]. In this comparatively high r region, it is clear that the BER performance of all detection methods eventually converge to that of the linear MMSE detection scheme.

The results in Figure 5.8 validate the performance enhancement of the proposed iterative method using the stair matrix in the development over the existing proposals where the diagonal matrix is adopted, especially in a comparatively low r region. This is consistent with our previous analysis. It is also foreseeable that applying the proposed scheme in SC-FDMA systems, the performance enhancement can be achieved since the received signal model in Equation (5.1) holds over each subcarrier as well.

5.5 Summary

In this chapter, we propose the application of the stair matrix in massive MIMO systems. To begin with, we demonstrate that with sufficiently large number of antennas at the base station, the probability that the convergence conditions are met with the use of the stair matrix approaches 1. We then propose an iterative method to reduce the computational complexity and show that the overall computational complexity is of the same level as the existing iterative methods where the diagonal matrix is applied. Furthermore, we evaluate the performance of the stair matrix in terms of the probability that the convergence conditions are met, the normalized mean-square error in Neumann series expansion to approach the matrix inverse, the residual estimation error of the iterative method to approach the linear ZF/MMSE estimation, and the system BER performance. Numerical simulations show that performance enhancement by using the stair matrix over the diagonal matrix is present in all performance metrics.

Chapter 6

Low Complexity and Fast Processing Algorithms for V2I Massive MIMO Uplink Detection

[1] Using 5G techniques to support vehicular communications has been recently studied [6–8, 10–12, 33, 35, 36, 101–103]. Particularly, one typical application scenario in future 5G network is to provide high level service experience for end users on the move [101]. The

[1]The related work has been published in or submitted to

- *IEEE Transactions on Vehicular Technology*, 2017, under review.
- *Proceedings of the 10th International EAI International Wireless Internet Conference* (WiCON), Tianjin, China, December 2017.
- *Proceedings of the 10th International Conference on Mobile Ad-hoc and Sensor Networks* (IEEE MSN2014), Maui, Hawaii, USA, December 2014.
- *Proceedings of 22nd IEEE Newfoundland Electrical and Computer Engineering Conference* (NECEC'13), St. Johns, NL, Canada, November 2013.

expected data rate for downlink and uplink transmission can be at least 100Mb/s and 20Mb/s, respectively. Meanwhile, the end-to-end latency is maintained below 100ms [101]. The authors in [6] investigate the cooperative transmission in 5G small-cell networks to improve transmission capacity and reliability for vehicular communications. X. Ge *et al* have proposed a new vehicular network architecture where the 5G mobile communication technologies and software defined networks are integrated [7]. As one of the key technologies in 5G, massive MIMO has been considered in vehicular networks [7, 8, 33, 35, 36, 102]. At roadside unit (RSU), hundreds of antennas can be deployed (as illustrated in [33, 36, 102], by using millimeter-wave (Terahertz frequency band [8]), the elements at the RSU can reach upto 256), and the simultaneous vehicles in service can be tens. The massive MIMO architecture is quite suitable for the moving vehicles to access the roadside infrastructure [35]. Besides, massive MIMO has shown huge potential spectrum and energy efficiency [3, 29, 32, 34, 104]. Those benefits can be achieved by employing simple linear match-filter (MF), zero-forcing (ZF), and minimum mean-square error (MMSE) receivers [3,29,32,34,38]. However, the results in [3, 29, 32, 34] clearly show that in terms of both the achievable sum rate capacity and the energy efficiency, linear ZF or MMSE receivers show obvious performance gain over the MF receivers. Therefore, linear ZF or MMSE receivers are more preferable in real implementations [23, 81, 84–86, 105].

However, as the number of vehicular users under simultaneous service can be huge, the computational complexity for linear detection schemes is still considerable. Besides, the iterative methods, for example, Gauss-Seidel method, experience large processing delay caused by iterations and successive detection. In this chapter, we propose low complexity and fast processing algorithms to address those issues. The proposed schemes transform the large-scale matrix inverse problems in linear MMSE detection scheme into solving linear equations. We then introduce iterative methods to solve linear equations. To speed up

the updating process in the iterative method, we utilize the properties of a block matrix, and perform the updating process on a small size block independently. The independent processing can be done in parallel, which greatly reduces the overall processing time. We also evaluate the performance of the proposed schemes in terms of the probability that the convergence conditions are met, and the system bit error rate. The results show that the proposed schemes achieve good system performance and at low complexity and latency.

The main contributions are summarised as follows.

1) We propose to utilize the block diagonal matrices for the updating progress in iterative methods. By using small size block diagonal matrices, the matrix inversion issues degrade from a large scale to a small scale. Therefore, the overall computational complexity is greatly reduced;

2) We propose an alternative matrix inversion algorithm by using the block matrix inversion theorem. Compared to the well-known Cholesky decomposition based matrix inversion method, the proposed algorithm involves less memory storage and fewer multiplications;

3) We propose the block Gauss-Seidel method for the updating progress. The proposed Gauss-Seidel method is performed on small size block matrix independently, hence the computational complexity on each block is greatly reduced.

4) We propose using the parallel processing for the updating progress. By utilizing the independent structure, parallel processing can significantly reduce the overall processing time.

We also evaluate the performance of the proposed schemes in terms of the probability that the convergence conditions are met, and the system bit error rate (BER) in various system configurations. Numerical results have shown that the proposed schemes can achieve

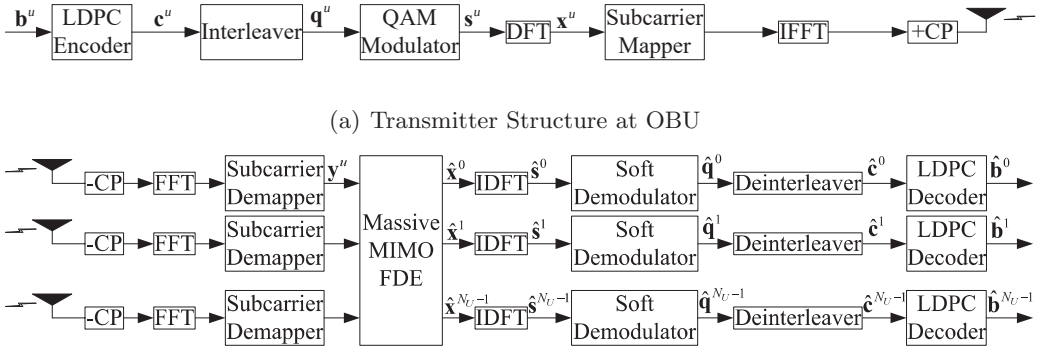
near linear MMSE detection performance but with low complexity and low processing delay. The superiority of the proposed algorithms is even more significant when the number of vehicles under simultaneous service becomes large. This superiority, i.e., low complexity and fast processing time, makes the proposed algorithms quite suitable for the applications of massive MIMO to modern vehicular networks.

The rest of this chapter is organized as follows. In Section 6.1, we introduce the system model and briefly review the linear detection scheme, Cholesky decomposition based matrix inversion method, Jacobi method, and the Gauss-Seidel method. The proposed iterative method that utilizes the block diagonal matrices is presented in Section 6.2, followed by the proposed block Gauss-Seidel method in Section 6.3. Numerical simulations and discussion are provided in Section 6.4 and conclusions are drawn in Section 6.5.

6.1 System Model

We consider an application scenario shown in Figure 6.1 where the roadside infrastructure and the moving vehicles on the road form a typical vehicular network. The moving vehicles are connected with neighbouring vehicles through vehicle-to-vehicle (V2V) links and they communicate with the RSU through vehicle-to-infrastructure (V2I) links. The RSU is deployed with massive/large-scale antenna array and is mounted on the top of the roadside buildings. To support ITS service, all vehicles collect traffic data and share with neighboring vehicles or report to the RSU. For example, when the vehicle accesses a school zone, the information includes location, speed limit, school bus, pedestrians, parking lot, and so on, will be collected, and shared with the vehicles behind. Those information can be also reported to the RSU, and RSU can broadcast to alert the vehicles in its communication area. Obviously, the number of vehicles, which are simultaneously communicating with the RSU, can be huge, and low complexity and fast processing data detection schemes are

Figure 6.1: RSU with massive antennas to serve multiple vehicles on the road



(a) Transmitter Structure at OBU

(b) Receiver Structure at RSU

Figure 6.2: The baseband signal processing diagram at OBU and RSU

required.

As pointed out in [10, 11], V2I communications play a significant role in ITS service, we mainly focus on the V2I links in this Chapter. For uplink transmission, at each vehicle, the

on-board unit (OBU) encodes the source information bitstream, and modulated to QAM symbols. To avoid high peak-to-average power ratio (PAPR), the single-carrier frequency division multiple access (SC-FDMA), also known as DFT-spread orthogonal frequency division multiplexing (DFT-S-OFDM), is adopted [23]. Compared to OFDM modulation, in SC-FDMA systems, DFT operations are performed before OFDM modulations.

The baseband signal processing diagram is presented in Figure 6.2. We assume the number of antenna elements at RSU is N_B, and the number of vehicles in simultaneous service is N_U. At the OBU side, for the u-th vehicular user, the LDPC encoder first encodes the source information bitstream (of length K), $\mathbf{b}^u = [b_0^u, b_1^u, \cdots, b_{K-1}^u]^{\mathrm{T}}$, to generate the corresponding coded bitstream (of length L), $\mathbf{c}^u = [c_0^u, c_1^u, \cdots, c_{L-1}^u]^{\mathrm{T}}$. The LDPC code rate is $R = K/L$. After interleaving, the output bitstream is $\mathbf{q}^u = [q_0^u, q_1^u, \cdots, q_{L-1}^u]^{\mathrm{T}}$, where $q_\ell^u = \Pi\left(c_l^u\right)$ and $\Pi(\cdot)$ is the interleaving function. $\mathbf{q}^u$ is fed into the QAM modulator to produce the transmitted symbol block $\mathbf{s}^u = [s_0^u, s_1^u, \cdots, s_{N-1}^u]^{\mathrm{T}}$, where $s_n^u = \mathscr{Q}\left(\left[q_{n,0}^u, q_{n,1}^u, \cdots, q_{n,\log_2 M-1}^u\right]\right)$. The QAM symbol s_n^u carries $\log_2 M$ bits information, and the m-th mapping bit is $q_{n,m}^u$. M is the cardinality of the QAM symbol set χ. Performing DFT operations, the frequency domain symbol vector $\mathbf{x}^u = [x_0^u, x_1^u, \cdots, x_{N-1}^u]^{\mathrm{T}}$ is obtained, with $x_k^u = \sum_{n=0}^{N-1} s_n^u \exp\left(-\frac{j2\pi kn}{N}\right)$. Redefine the transmitted signal vector over the k-th subcarrier as $\mathbf{x}_k = \left[x_k^0, x_k^1, \cdots, x_k^{N_U-1}\right]^{\mathrm{T}}$, we generalize the frequency domain signal vector as

$$\mathbf{x} = \left(\mathbf{F}_N \otimes \mathbf{I}_{N_U}\right)\mathbf{s}, \tag{6.1}$$

where $\mathbf{x} = \left[\mathbf{x}_0^{\mathrm{T}}, \mathbf{x}_1^{\mathrm{T}}, \cdots, \mathbf{x}_{N-1}^{\mathrm{T}}\right]^{\mathrm{T}}$, and $\mathbf{s} = \left[\mathbf{s}_0^{\mathrm{T}}, \mathbf{s}_1^{\mathrm{T}}, \cdots, s_{N-1}^{\mathrm{T}}\right]^{\mathrm{T}}$ with $\mathbf{s}_n = \left[s_n^0, s_n^1, \cdots, s_n^{N_U-1}\right]^{\mathrm{T}}$. After DFT, the frequency domain symbols are mapped to available subcarriers, and fed into an inverse-fast Fourier transform (IFFT) module. To overcome the inter-block-interference (IBI) in the multi-path frequency selective vehicular channel [5, 9], a cyclic prefix (CP) is put before sending to massive MIMO channel.

Assuming the length of the CP is no less than the maximum path delay, and after

removing CP, we perform the FFT operations on the received signal at each antenna. Through subcarrier demapping, we have

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{z}_k, \tag{6.2}$$

where $\mathbf{H}_k \in \mathbb{C}^{N_B \times N_U}$ is the frequency domain channel response. $\mathbf{z}_k = \left[z_{k,0}, z_{k,1}, \cdots, z_{k,N_B}\right]^{\mathrm{T}}$ is the noise vector over the k-th subcarrier, with each entry modeled as zero mean complex Gaussian circularly symmetric random variable [69]. $\mathrm{E}\left\{\mathbf{z}_k \mathbf{z}_k^{\mathrm{H}}\right\} = \sigma_z^2 \mathbf{I}_{N_B}$. $\mathbf{y}_k$ is used for later detection.

6.1.1 Linear MMSE Detection

From Equation (6.2), we denote the received signal over all subcarriers as $\mathbf{y} = \left[\mathbf{y}_0^{\mathrm{T}}, \mathbf{y}_1^{\mathrm{T}}, \cdots, \mathbf{y}_{N-1}^{\mathrm{T}}\right]^{\mathrm{T}}$, given by

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z} = \mathbf{H}\left(\mathbf{F}_N \otimes \mathbf{I}_{N_U}\right)\mathbf{s} + \mathbf{z}, \tag{6.3}$$

where $\mathbf{z} = \left[\mathbf{z}_0^{\mathrm{T}}, \mathbf{z}_1^{\mathrm{T}}, \cdots, \mathbf{z}_{N-1}^{\mathrm{T}}\right]^{\mathrm{T}}$, and the frequency domain channel matrix, $\mathbf{H}$, is given by

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_0 & & & \\ & \mathbf{H}_1 & & \\ & & \ddots & \\ & & & \mathbf{H}_N \end{bmatrix}.$$

With complete knowledge of the channel matrix, the linear MMSE detection can be expressed as

$$\hat{\mathbf{s}} = \left(\mathbf{F}_N^{\mathrm{H}} \otimes \mathbf{I}_{N_U}\right)\left(\mathbf{H}^{\mathrm{H}}\mathbf{H} + \sigma_z^2 \mathbf{I}_{N_U}\right)^{-1}\mathbf{H}^{\mathrm{H}}\mathbf{y}. \tag{6.4}$$

Equation (6.4) indicates that the linear MMSE estimation of $\hat{\mathbf{s}}$ consists of two stages, which are given by the following steps:

$$\hat{\mathbf{x}} = \left(\mathbf{H}^{\mathrm{H}}\mathbf{H} + \sigma_z^2 \mathbf{I}_{N_U}\right)^{-1}\mathbf{H}^{\mathrm{H}}\mathbf{y} = \mathbf{W}^{-1}\mathbf{y}^{\mathrm{MF}}, \tag{6.5}$$

$$\hat{\mathbf{s}} = \left(\mathbf{F}_N^{\mathrm{H}} \otimes \mathbf{I}_{N_U}\right) \hat{\mathbf{x}}, \tag{6.6}$$

where the equalization matrix $\mathbf{W} = \mathbf{G} + \sigma_z^2 \mathbf{I}_{N_U}$ and the Gram matrix $\mathbf{G} = \mathbf{H}^{\mathrm{H}}\mathbf{H}$. $\mathbf{y}^{\mathrm{MF}} = \mathbf{H}^{\mathrm{H}}\mathbf{y}$ is the match-filter output. $\mathbf{W}$ and $\mathbf{G}$ are both block diagonal matrices, with $\mathbf{W}_k \in \mathbb{C}^{N_U \times N_U}$ and $\mathbf{G}_k \in \mathbb{C}^{N_U \times N_U}$ respectively given by

$$\mathbf{W}_k = \mathbf{G}_k + \sigma_z^2 \mathbf{I}_{N_U}, \tag{6.7}$$

$$\mathbf{G}_k = \mathbf{H}_k^{\mathrm{H}} \mathbf{H}_k. \tag{6.8}$$

It is worth noting that if the noise component is not considered in the equalization process, the linear MMSE estimation degrades to the linear ZF detection.

We extract $\hat{s}_n^u$ from Equation (6.4), given by

$$\hat{s}_n^u = \rho^u s_n^u + \xi_n^u, \tag{6.9}$$

where ρ^u is the equivalent channel gain and ξ_n^u denotes the noise-plus-interference (NPI) component. ρ^u is given by

$$\rho^u = \frac{1}{N} \sum_{k=0}^{N-1} \mathbf{e}_u^{\mathrm{H}} \mathbf{W}_k^{-1} \mathbf{G}_k \mathbf{e}_u, \tag{6.10}$$

and the covariance of the NPI, $v^u = \mathrm{cov}\left\{\xi_n^u, \xi_n^u\right\}$, is given by

$$v^u = \rho^u \left(1 - \rho^u\right). \tag{6.11}$$

Therefore, the *posteriori* signal-to-noise-plus-interference ratio (SINR), γ_u, is given by

$$\gamma_u = \frac{\rho^u}{1 - \rho^u}. \tag{6.12}$$

In addition, the *max-log* likelihood ratio (LLR) of the bits associated with s_n^u is given by [23, 84]

$$L\left(q_{n,m}^u\right) = \gamma_u \left(\min_{s \in \chi_m^0} \left| \frac{\hat{s}_n^u}{\rho^u} - s \right|^2 - \min_{s' \in \chi_m^1} \left| \frac{\hat{s}_n^u}{\rho^u} - s' \right|^2 \right), \tag{6.13}$$

where $\chi_m^b \triangleq \left\{ s \,|s \in \chi, q_m = b \right\}$ denotes the subset of χ, where the m-th mapping bit associated with s is b. $L\left(q_{n,m}^u\right)$ is deinterleaved, and then fed into LDPC decoder for decoding process, as shown in Figure 6.2(b).

6.1.2 Matrix Inversion Based on Cholesky Decomposition

We note in Equation (6.5), the matrix inversion operations are required for linear MMSE detection. As $\mathbf{W}$ is a block diagonal matrix, the inversion can be implemented on each subcarrier. By noting that $\mathbf{W}_k$ is a Hermitian matrix, the matrix inversion based on Cholesky decomposition is usually adopted [23, 106]. To begin with, $\mathbf{W}_k$ can be factorised into

$$\mathbf{W}_k = \mathbf{R}_k^{\mathrm{H}} \mathbf{\Lambda}_k \mathbf{R}_k, \tag{6.14}$$

where $\mathbf{R}_k \in \mathbb{C}^{N_U \times N_U}$ is an upper triangular matrix where the diagonal elements are ones, and $\mathbf{\Lambda}_k = \mathrm{diag}\left\{\left[\Lambda_k^0, \Lambda_k^1, \cdots, \Lambda_k^{N_U-1}\right]^{\mathrm{T}}\right\}$ is a diagonal matrix. The diagonal elements, Λ_k^u, $u = 0, 1, \cdots, N_U - 1$, can be given as

$$\Lambda_k^u = \mathbf{W}_k^{(u,u)} - \sum_{l=0}^{u-1} \left|\mathbf{R}_k^{(l,u)}\right|^2 \Lambda_k^l. \tag{6.15}$$

The upper triangular elements, $\mathbf{R}_k^{(u,v)}$, $u < v$, are given by

$$\mathbf{R}_k^{(u,v)} = \frac{1}{\Lambda_k^u}\left(\mathbf{W}_k^{(u,v)} - \sum_{l=0}^{u-1} \left(\mathbf{W}_k^{(l,u)}\right)^* \mathbf{W}_k^{(l,v)} \Lambda_k^l\right). \tag{6.16}$$

Let $\mathbf{X}_k = \mathbf{W}_k^{-1}$ such that $\mathbf{W}_k \mathbf{X}_k = \mathbf{I}_{N_U}$, using the results in Equation (6.14), we then have

$$\mathbf{R}_k^{\mathrm{H}} \mathbf{\Lambda}_k \mathbf{B}_k = \mathbf{I}_{N_U}, \tag{6.17}$$

$$\mathbf{R}_k \mathbf{X}_k = \mathbf{B}_k. \tag{6.18}$$

Equation (6.17) indicates that $\mathbf{B}_k$ is the inverse of the matrix $\mathbf{R}_k^{\mathrm{H}} \mathbf{\Lambda}_k$; therefore, $\mathbf{B}_k$ is a lower triangular matrix. Meanwhile, as $\mathbf{X}_k$ is also a Hermitian matrix, only the diagonal and the above diagonal elements are required to obtain $\mathbf{X}_k$. By noting these properties [106], we use backward substitution to solve Equation (6.18), given by

$$\mathbf{X}_k^{(v,v)} = \frac{1}{\Lambda_k^v} - \sum_{l=v+1}^{N_U-1} \mathbf{R}_k^{(v,l)} \mathbf{X}_k^{(l,v)}, \tag{6.19}$$

$$\mathbf{X}_k^{(u,v)} = -\sum_{l=u+1}^{N_U-1} \mathbf{R}_k^{(u,l)} \mathbf{X}_k^{(l,v)}, \tag{6.20}$$

where only upper triangular elements, $\mathbf{X}_k^{(u,v)}$, $(u \le v)$, are solved, and $\mathbf{X}_k^{(v,u)} = \left(\mathbf{X}_k^{(u,v)}\right)^*$ $(u \le v)$ is used when needed [106].

The above solutions avoid computing $\mathbf{B}_k$ and fewer real number multiplications are required.

6.1.3 Jacobi Method

The above matrix inversion method based on Cholesky decomposition is a well-known matrix inversion method; however, the computational complexity is still in the order $O\left(N_U^3\right)$. Instead of computing the matrix inverse, we can transform the matrix inversion problem into solving linear equations. We then adopt the iterative methods to solve linear equations.

To be specific, we first rewrite Equation (6.5) into

$$\mathbf{W}\hat{\mathbf{x}} = \mathbf{y}^{\mathrm{MF}}. \tag{6.21}$$

Using the block matrix properties, we have N independent equations established on each subcarrier, given by

$$\mathbf{W}_k\hat{\mathbf{x}}_k = \mathbf{y}_k^{\mathrm{MF}}, \tag{6.22}$$

where $\mathbf{y}_k^{\mathrm{MF}} = \mathbf{H}_k^{\mathrm{H}}\mathbf{y}_k$ is the matched-filter output over the k-th subcarrier. We divide $\mathbf{W}_k$ into

$$\mathbf{W}_k = \mathbf{D}_k + \mathbf{E}_k, \tag{6.23}$$

where $\mathbf{D}_k = \mathrm{diag}\{\mathbf{d}_k\}$ with $\mathbf{d}_k = \left[d_k^0, d_k^1, \cdots, d_k^{N_U-1}\right]^{\mathrm{T}}$ where $d_k^u = \mathbf{W}_k^{(u,u)}$, and $\mathbf{E}_k$ consists of the rest of the elements in $\mathbf{W}_k$. With the Jacobi method [85], we have the updating process given by

$$\hat{\mathbf{x}}_k^{(i+1)} = \mathbf{D}_k^{-1}\left(\mathbf{y}_k^{\mathrm{MF}} - \mathbf{E}_k\hat{\mathbf{x}}_k^{(i)}\right), \tag{6.24}$$

where $\hat{\mathbf{x}}_k^{(i)}$ is the i-th estimation. Providing an initial estimate, we can follow the Jacobi iterative structure in Equation (6.24) to update the next estimation, eventually approaching the linear MMSE detection results.

It is worth noting that only matrix-vector products are involved in Equation (6.24), and the computational complexity of the Jacobi method is in the order $O\left(N_U^2\right)$. When the ratio $r = N_B/N_U$ is sufficiently large, only a few iterations are required for the Jacobi method, hence reducing the overall computational complexity relative to direct matrix inversion methods.

6.1.4 Gauss-Seidel Method

In the case where the ratio $r = N_B/N_U$ is not sufficiently large, the Jacobi method converges slowly [84]. The Gauss-Seidel method investigated in [84] has shown a faster convergence rate than Jacobi method. With the Gauss-Seidel method, $\mathbf{W}_k$ is first divided into

$$\mathbf{W}_k = \mathbf{D}_k + \mathbf{L}_k + \mathbf{L}_k^{\mathrm{H}}, \tag{6.25}$$

where $\mathbf{L}_k$ takes the lower triangular (below diagonal) elements in $\mathbf{W}_k$. The updating process can be expressed as

$$\hat{\mathbf{x}}_k^{(i+1)} = \left(\mathbf{D}_k + \mathbf{L}_k\right)^{-1}\left(\mathbf{y}_k^{\mathrm{MF}} - \mathbf{L}_k^{\mathrm{H}}\hat{\mathbf{x}}_k^{(i)}\right). \tag{6.26}$$

Although Equation (6.26) involves matrix inversion, successive detection can be employed as $\left(\mathbf{D}_k + \mathbf{L}_k\right)$ is a lower triangular matrix. Specifically, we have

$$\hat{x}_{k,u}^{(i+1)} = \frac{y_{k,u}^{\mathrm{MF}} - \sum_{v=0}^{u-1}\mathbf{W}_k^{(u,v)}\hat{x}_{k,v}^{(i+1)} - \sum_{v=u+1}^{N_U-1}\mathbf{W}_k^{(u,v)}\hat{x}_{k,v}^{(i)}}{d_k^u}, \tag{6.27}$$

where $\hat{x}_{k,u}^{(i)} = \mathbf{e}_n^{\mathrm{H}}\hat{\mathbf{x}}_k^{(i)}$ is the estimation of the u-th user's transmit symbol over the k-th sub-carrier in the i-th iteration, and $y_{k,u}^{\mathrm{MF}} = \mathbf{e}_u^{\mathrm{H}}\mathbf{y}_k^{\mathrm{MF}}$. Apparently, the computational complexity is in the order $O\left(N_U^2\right)$ in each iteration.

6.2 Block Diagonal Matrix Inversion Based Detection Method

The Gauss-Seidel method based detection scheme [84] has the following limitations. First of all, when the number of vehicles in service becomes very large, the successive detection introduces large processing delay since the last user needs to wait until all previous users have been detected. In addition to that, the successive signal processing structure is not efficient for hardware implementation [85]. In this section, we will introduce the use of the block diagonal matrix in massive MIMO data detection.

6.2.1 Development of the Proposed Scheme

It has been shown in [23] that $\mathbf{W}_k$ is a diagonal dominant matrix. Based on that, we define the block diagonal matrix, $\mathbf{P}_k$, which is given by

$$\mathbf{P}_k = \begin{bmatrix} \mathbf{P}_k(0) & & & \\ & \mathbf{P}_k(1) & & \\ & & \ddots & \\ & & & \mathbf{P}_k(B-1) \end{bmatrix}, \tag{6.28}$$

where the l-th block, $\mathbf{P}_k(l)$, consists of the elements at the rows and the columns both from $l \cdot N_L$ to $(l+1) \cdot N_L - 1$ in $\mathbf{W}_k$. B and N_L denote the number of blocks and the size of the block. For better illustration, we suppose $N_L \times B = N_U$[2]. According to the expressions in Equation (6.24) and (6.26), we have the iterative estimation given by

$$\hat{\mathbf{x}}_k^{(i+1)} = \mathbf{P}_k^{-1}\left(\mathbf{y}_k^{\mathrm{MF}} - \mathbf{Q}_k \hat{\mathbf{x}}_k^{(i)}\right), \tag{6.29}$$

where $\mathbf{Q}_k = \mathbf{W}_k - \mathbf{P}_k$. Although the matrix inversion, $\mathbf{P}_k^{-1}$, is involved in Equation (6.29), we can utilize the properties of the block diagonal matrix to reduce the computational complexity. Specifically, we define $\tilde{\mathbf{y}}_k^{(i)} = \mathbf{y}_k^{\mathrm{MF}} - \mathbf{Q}_k \hat{\mathbf{x}}_k^{(i)}$ which can be preprocessed when the

[2]If the equation does not hold, the last block consists of the rest elements in the last block.

estimate after i-th iteration, $\hat{\mathbf{x}}_k^{(i)}$, is available. We then divide $\hat{\mathbf{x}}_k^{(i)}$ and $\tilde{\mathbf{y}}_k^{(i)}$ into B blocks, respectively given by

$$\hat{\mathbf{x}}_k^{(i)} = \begin{bmatrix} \hat{\mathbf{x}}_k^{(i)}(0) \\ \hat{\mathbf{x}}_k^{(i)}(1) \\ \vdots \\ \hat{\mathbf{x}}_k^{(i)}(B-1) \end{bmatrix}, \tag{6.30}$$

$$\tilde{\mathbf{y}}_k^{(i)} = \begin{bmatrix} \tilde{\mathbf{y}}_k^{(i)}(0) \\ \tilde{\mathbf{y}}_k^{(i)}(1) \\ \vdots \\ \tilde{\mathbf{y}}_k^{(i)}(B-1) \end{bmatrix}, \tag{6.31}$$

where $\hat{\mathbf{x}}_k^{(i)}(l)$ is the l-th block in $\hat{\mathbf{x}}_k^{(i)}$, denoting the estimate of the frequency symbols over the k-th subcarrier for the users from $l \cdot N_L$ to $(l+1) \cdot N_L - 1$ in the i-th iteration; $\tilde{\mathbf{y}}_k^{(i)}(l)$ is the l-th block, consisting of the elements in rows from $l \cdot N_L$ to $(l+1) \cdot N_L - 1$ in $\tilde{\mathbf{y}}_k^{(i)}$. $\tilde{\mathbf{y}}_k^{(i)}(l)$ is given by

$$\tilde{\mathbf{y}}_k^{(i)}(l) = \mathbf{y}_k^{\mathrm{MF}}(l) - \mathbf{Q}_k(l)\,\hat{\mathbf{x}}_k^{(i)}, \tag{6.32}$$

where $\mathbf{Q}_k(l)$ consists of elements in rows from $l \cdot N_L$ to $(l+1) \cdot N_L - 1$ in $\mathbf{Q}_k$, and the elements in columns from $l \cdot N_L$ to $(l+1) \cdot N_L - 1$ are zeros; $\mathbf{y}_k^{\mathrm{MF}}(l)$ consists of the rows from $l \cdot N_L$ to $(l+1) \cdot N_L - 1$ in $\mathbf{y}_k^{\mathrm{MF}}$. In addition, we have

$$\hat{\mathbf{x}}_k^{(i+1)}(l) = \mathbf{P}_k^{-1}(l)\,\tilde{\mathbf{y}}_k^{(i)}(l). \tag{6.33}$$

That is to say, we only need to compute the matrix inverse, $\mathbf{P}_k^{-1}(l)$, with the matrix size $N_L = N_U/B$.

Therefore, we divide the updating progress in Equation (6.29) into two steps: the pre-processing expressed in Equation (6.32) to obtain $\tilde{\mathbf{y}}_k(l)$, and the updating progress given by Equation (6.33). Instead of solving a large size matrix inversion, we only need to compute a much small size matrix inversion, which requires fewer computations.

It is worth pointing out that if the number of blocks, $B = 1$, the block diagonal matrix inverse based detection scheme is equivalent to the direct matrix inversion in linear MMSE detection; if $B = N_U$, the proposed scheme becomes the Jacobi method.

6.2.2 A New Matrix Inversion Method

An matrix inversion based on Cholesky decomposition can be used for the computation of $\mathbf{P}_k^{-1}(l)$; however, as the size of the $\mathbf{P}_k^{-1}(l)$ is small, we have a better way to compute matrix inversion, which is based on block matrix inversion [107]. As we will show later, the new matrix inversion matrix is of low complexity and requires low storage space.

To begin with, we present the block matrix inversion given by

$$\begin{aligned}&\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} \\ =&\begin{bmatrix} \left(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right)^{-1} & -\mathbf{A}^{-1}\mathbf{B}\left(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right)^{-1} \\ -\left(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right)^{-1}\mathbf{C}\mathbf{A}^{-1} & \left(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right)^{-1} \end{bmatrix},\end{aligned} \tag{6.34}$$

where $\mathbf{A}$ and $\mathbf{D}$ are square matrix; $\mathbf{B}$ and $\mathbf{C}$ are arbitrary size. The Equation (6.34) holds when $\mathbf{A}$, $\mathbf{D}$, $\left(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right)$, and $\left(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right)$ are invertible.

Suppose we have a Hermitian matrix $\mathbf{A}_M \in \mathbb{C}^{M\times M}$, and its inversion $\mathbf{A}_M^{-1}$. The matrix $\mathbf{A}_{M+1}$ is given by

$$\mathbf{A}_{M+1} = \begin{bmatrix} \mathbf{A}_M & \mathbf{b} \\ \mathbf{b}^{\mathrm{H}} & a_{M+1} \end{bmatrix},$$

where $a_{M+1} > 0$, and $\left(a_{M+1} - \mathbf{b}^{\mathrm{H}}\mathbf{A}_M^{-1}\mathbf{b}\right) \neq 0$. Define $\mathbf{d} = \mathbf{A}_M^{-1}\mathbf{b}$, according to Equation (6.34), we have

$$\mathbf{A}_{M+1}^{-1} = \begin{bmatrix} \mathbf{B} & \mathbf{c} \\ \mathbf{c}^{\mathrm{H}} & d_{M+1} \end{bmatrix}, \tag{6.35}$$

where $\mathbf{B} \in \mathbb{C}^{M\times M}$, $\mathbf{c} \in \mathbb{C}^{M\times 1}$, and

$$d_{M+1} = \left(a_{M+1} - \mathbf{b}^{\mathrm{H}}\mathbf{d}\right)^{-1}, \tag{6.36}$$

$$\mathbf{c} = -d_{M+1}\mathbf{d}, \tag{6.37}$$

$$\mathbf{B} = \left(\mathbf{A}_M - a_{M+1}^{-1}\mathbf{b}\mathbf{b}^{\mathrm{H}}\right)^{-1}, \tag{6.38}$$

$\mathbf{B}$ is known as the Schur component of $\mathbf{A}_M$, which can be calculated as

$$\mathbf{B} = \mathbf{A}_M^{-1} + d_{M+1}\mathbf{d}\mathbf{d}^{\mathbf{H}}, \tag{6.39}$$

That is to say, if we have the vector $\mathbf{d}$ as $\mathbf{d} = \mathbf{A}_M^{-1}\mathbf{b}$, we then update d_{M+1}, $\mathbf{c}$, and $\mathbf{B}$ in Equations (6.36), (6.37), (6.39), respectively.

In summary, we present the new matrix inversion method in **Algorithm 6.1**. It is straightforward from **Algorithm 6.1** that the new matrix inversion method only needs $O(M)$ extra memory storage for $\mathbf{d}$. Compared to the Cholesky decomposition based matrix inversion method where the memory storage is required for both $\boldsymbol{\Lambda_k}$ and $\mathbf{R}_k$, the new method requires less storage space.

6.2.3 Initial Estimation

The initial estimate, $\hat{\mathbf{x}}_k^{(0)}$, is required for iterative methods. If we have an initial estimate that is close to the MMSE estimation, the number of iterations will be greatly reduced, hence reducing the overall processing delay. In priori work [23, 80, 84], the authors have demonstrated that $\mathbf{W}_k$ will be diagonal dominant when the ratio $r = N_B/N_U$ is sufficiently large. By using this property, we can use the initial estimation for the l-th block given by

$$\hat{\mathbf{x}}_k^{(0)}(l) = \mathbf{D}_k^{-1}(l)\,\mathbf{y}_k^{\mathrm{MF}}(l), \tag{6.40}$$

where the diagonal matrix $\mathbf{D}_k^{-1}$ is utilized to approximate $\mathbf{W}_k^{-1}$. Actually, this approximation is also used in [84] as an initial estimate. It is worth noting that the initial estimate provided in Equation (6.40) is better than zero vector used in [83], and the computational complexity is low.

Table 6.1: **Algorithm 6.1:** Compute the Inversion of a Hermitian Matrix

---

Input: The Hermitian Matrix $\mathbf{A} \in \mathbb{C}^{M \times M}$;

Output: The Matrix Inversion $\mathbf{B} = \mathbf{A}^{-1}$.

Initialization:

1. $\mathbf{B}(1,1) = 1/\mathbf{A}(1,1)$, $\mathbf{d} = \text{zeros}(1, M)$;

Loop:

2. for $m = 1:1:(M-1)$
3. $\quad \mathbf{d}(1:m) = \mathbf{B}(1:m, 1:m) \cdot \mathbf{A}(1:m, m+1)$;
4. $\quad \mathbf{B}(m+1, m+1) = \mathbf{A}(m+1, 1:m) \cdot \mathbf{d}(1:m)$;

   $\quad \mathbf{B}(m+1, m+1) = 1/(\mathbf{A}(m+1, m+1) - \mathbf{B}(m+1, m+1))$;
5. $\quad \mathbf{B}(1:m, m+1) = -\mathbf{B}(m+1, m+1) \cdot \mathbf{d}(1:m)$;
6. $\quad \mathbf{B}(1:m, 1:m) = \mathbf{B}(1:m, 1:m) - \mathbf{B}(1:m, m+1) \cdot \mathbf{d}^{\mathrm{H}}(1:m)$;
7. end

Return B.

6.2.4 LLR Computation

In order to provide LLR for the soft-input LDPC decoder, we need to compute the LLR of the bits associated with the symbol estimation. In the Neumann series expansion method [23], the truncated order is limited for computational complexity consideration as matrix multiplications are involved. In addition, in a system where N_B/N_U is low (for example, $N_B/N_U \leqslant 5$), the matrix inverse approximation with the truncation order of three is far away from $\mathbf{W}_k^{-1}$. Furthermore, we note that the matrix inverse is still involved in Equation (6.10) for the computation of the equivalent channel gain.

Considering all those factors above, the authors in [23] propose to use the first truncation order of the Neumann series expansion to replace $\mathbf{W}_k^{-1}$ in the computation of ρ^u and v^u. This indicates that the NPI after the first iteration is considered for all the rest of the iterations. Similar idea is presented in [84, 85]. However, in iterative method, the covariance of the NPI will decrease with iterations. Moreover, with a sufficiently large number of iterations, the final estimation $\hat{\mathbf{x}}_k^{(i)}$ will be quite close to the MMSE estimation given the convergence conditions are satisfied. Therefore, if ρ^u and v^u still adopt the method in [23, 84, 85], significant system performance loss will be shown compared to the true MMSE detection.

In order to address the issue, we propose to utilize the NPI in MMSE estimation. To be specific, we first write

$$\rho^u = \frac{1}{N}\sum_{k=0}^{N-1} \mathbf{e}_u^{\mathrm{H}} \mathbf{W}_k^{-1} \mathbf{G}_k \mathbf{e}_u = \frac{1}{N}\sum_{k=0}^{N-1} \left(1 - \sigma_z^2 \mathbf{e}_u^{\mathrm{H}} \mathbf{W}_k^{-1} \mathbf{e}_u\right), \tag{6.41}$$

where Equation (6.7) is applied. When N_B become sufficiently large, $\mathbf{W}_k$ will be diagonal dominant. Therefore, an approximation of ρ^u can be given by

$$\rho^u \approx 1 - \frac{\sigma_z^2}{N}\sum_{k=0}^{N-1} \frac{1}{d_k^u}, \tag{6.42}$$

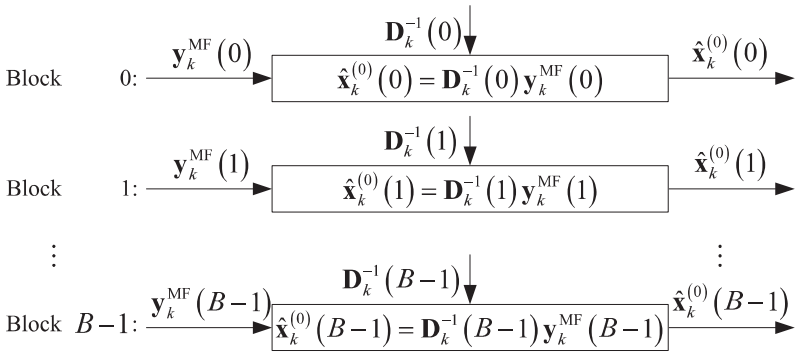
and the covariance of the NPI can be approximated using Equation (6.11). Using the approximation in (6.42) can greatly reduce the computational complexity to obtain the

equivalent channel gain ρ^u. Besides, with sufficient iterations (as we will show later, the computational complexity and processing delay for one iteration are low, and we can perform a large number of iterations), the relationship between v^u and ρ^u in Equation (6.11) can be held as the estimate is close to the MMSE estimate when convergence conditions are met. With v^u and ρ^u, the *posteriori* SINR is computed as Equation (6.12), and the LLRs are calculated as Equation (6.13).

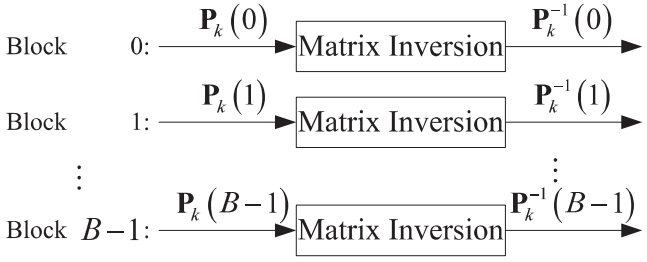
6.2.5 Parallel Processing

One of the most significant advantages to using the block diagonal matrix is to implement the detection scheme with parallel processing. As we illustrated previously, both the direct matrix inversion method and Gauss-Seidel method experience a long processing delay. The previous one requires huge computational load ($O\left(N_U^3\right)$), and the later one involves successive detection. By contrast, when we use the block diagonal matrix, the matrix inversion operations are performed on a small blocks, which requires less computational complexity for implementation. Besides, by noting the processing on each block is independent, the parallel processing structure can be utilized to speed up the detection progress.

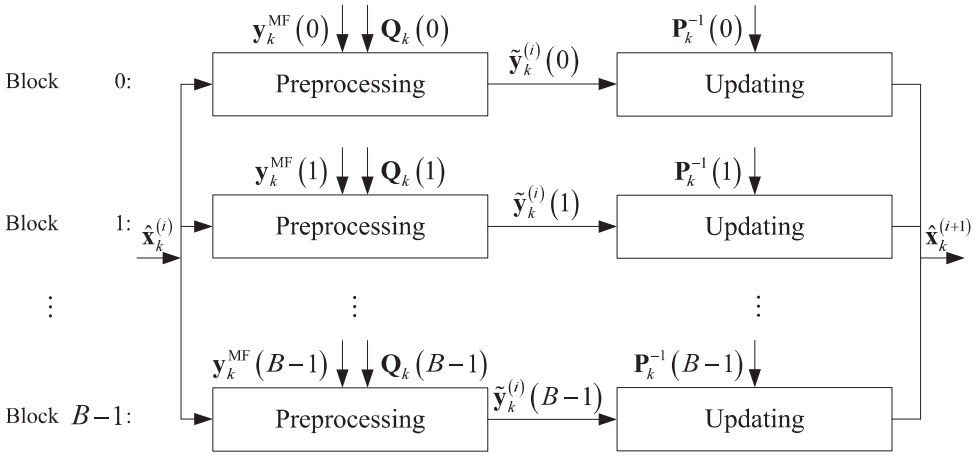
To be specific, the initial estimation provided in Equation (6.40) can be implemented on each block with the block diagonal matrix $\mathbf{D}_k^{-1}(l)$ available. We show the parallel processing for this progress in Figure 6.3(a). Apparently, as the size of $\mathbf{D}_k^{-1}(l)$ is much smaller than $\mathbf{D}_k^{-1}$, the processing delay is much reduced for computing the initial estimation. Besides, we can also apply the parallel processing on the block diagonal matrix inversion by using the property of the block diagonal matrix, which is shown in Figure 6.3(b). **Algorithm 6.1** is utilized on each block for matrix inversion. Compared to a large size matrix inversion, the matrix inversion on a small size matrix is of low complexity and processing time. Last, the iterative method can be implemented with the parallel processing structure as well, shown



(a) Initial Estimation



(b) Matrix Inversion



(c) Iterative Progress

Figure 6.3: Parallel processing on each block

in Figure 6.3(c). On each block, the progress for updating the next estimate is divided into two cascaded steps: preprocessing and updating. The preprocessing step is expressed in Equation (6.32) and the updating step is given by Equation (6.33). By using the parallel processing structure, the overall processing time for one iteration is greatly reduced relative to the processing time on each block.

Finally, we present the block diagonal matrix massive MIMO detection scheme in **Algorithm 6.2**.

6.2.6 Residual Estimation Error

The proposed block diagonal matrix based massive MIMO detection scheme is to approach linear MMSE estimation at each subcarrier by using the iterative method. To evaluate the error vector between the iterative estimation and the linear MMSE estimation, we define

$$\mathbf{e}_k^{(i)} = \hat{\mathbf{x}}_k^{(i)} - \hat{\mathbf{x}}_k, \tag{6.43}$$

where the iterative estimate $\hat{\mathbf{x}}_k^{(i)}$ is given by Equation (6.29), and the linear MMSE estimate $\hat{\mathbf{x}}_k = \mathbf{W}_k^{-1}\mathbf{y}_k^{\mathrm{MF}}$ is derived from Equation (6.22). According to Equation (6.29), we can derive

$$\begin{aligned}
\mathbf{e}_k^{(i)} =& \mathbf{P}_k^{-1}\left(\mathbf{y}_k^{\mathrm{MF}} - \mathbf{Q}_k\hat{\mathbf{x}}_k^{(i-1)}\right) - \hat{\mathbf{x}}_k \\
=& \mathbf{P}_k^{-1}\left(\mathbf{W}_k\hat{\mathbf{x}}_k - \mathbf{Q}_k\hat{\mathbf{x}}_k^{(i-1)}\right) - \hat{\mathbf{x}}_k \\
=& \mathbf{P}_k^{-1}\mathbf{Q}_k\left(\hat{\mathbf{x}}_k - \hat{\mathbf{x}}_k^{(i-1)}\right) \\
=& -\mathbf{P}_k^{-1}\mathbf{Q}_k\mathbf{e}_k^{(i-1)}.
\end{aligned} \tag{6.44}$$

Similar to [23], the ℓ_2-norm of the residual estimation error is defined as

$$J^{(i)} = \left\|\mathbf{e}_k^{(i)}\right\|_2 = \left\|-\mathbf{P}_k\mathbf{Q}_k\mathbf{e}_k^{(i-1)}\right\|_2 = \cdots = \left\|(-\mathbf{P}_k\mathbf{Q}_k)^i\mathbf{e}_k^{(0)}\right\|_2, \tag{6.45}$$

where $\mathbf{e}_k^{(0)} = \hat{\mathbf{x}}_k^{(0)} - \hat{\mathbf{x}}_k$ indicates the initial residual estimation error. Apparently, we have

$$J^{(i)} \leqslant \left\|(\mathbf{P}_k\mathbf{Q}_k)^i\right\|_{\mathrm{F}} \left\|\mathbf{e}_k^{(0)}\right\|_2, \tag{6.46}$$

Table 6.2: **Algorithm 6.2:** Block Diagonal Matrix Based Massive MIMO Detection

---

Input: $\mathbf{H}_k$, $\mathbf{y}_k^{\mathrm{MF}}$, σ_z^2, number of blocks B, and iterations I

Output: Estimation after I Iterations, $\hat{\mathbf{x}}_k^{(I)}$.

Initialization:

1. $\mathbf{G}_k = \mathbf{H}_k^{\mathrm{H}}\mathbf{H}_k$, $\mathbf{W}_k = \mathbf{G}_k + \sigma_z^2\mathbf{I}_{N_U}$, $\mathbf{y}_k^{\mathrm{MF}} = \mathbf{H}_k\mathbf{y}_k$;
2. Obtain $\mathbf{D}_k(l)$, $\mathbf{P}_k(l)$, $\mathbf{Q}_k(l)$, and $\mathbf{y}_k^{\mathrm{MF}}(l)$, $l = 0, 1, \cdots, B$;

Parallel Processing: Initial Estimation (3-5)

3. for $l = 0 : 1 : (B-1)$
4. $\quad \mathbf{x}_k^{(0)}(l) = \mathbf{D}_k^{-1}(l)\,\mathbf{y}_k^{\mathrm{MF}}(l)$;
5. end

Parallel Processing: Matrix Inversion (6-8)

6. for $l = 0 : 1 : (B-1)$
7. $\quad$ Compute $\mathbf{P}_k^{-1}(l)$ with **Algorithm 6.1**;
8. end

Iteration:

9. for $i = 0 : 1 : (I-1)$

Parallel Processing: Iterative Progress (10-13)

10. $\quad$ for $l = 0 : 1 : (B-1)$
11. $\quad\quad$ Preprocessing: $\tilde{\mathbf{y}}_k^{(i)}(l) = \mathbf{y}_k^{\mathrm{MF}}(l) - \mathbf{Q}_k(l)\,\hat{\mathbf{x}}_k^{(i)}$;
12. $\quad\quad$ Updating: $\hat{\mathbf{x}}_k^{(i+1)}(l) = \mathbf{P}_k^{-1}(l)\,\tilde{\mathbf{y}}_k^{(i)}(l)$;
13. $\quad$ end
14. end

Return $\hat{\mathbf{x}}_k^{(I)}$.

---

where $\|\mathbf{Ax}\|_2 \leqslant \|\mathbf{A}\|_{\mathrm{F}}\|\mathbf{x}\|_2$ is used for the above derivation. Given the convergence condition satisfied for the block diagonal matrix based massive MIMO detection scheme, i.e., $\lim_{i\to\infty}\left(\mathbf{P}_k^{-1}\mathbf{Q}_k\right)^i = \mathbf{0}$, the residual estimate will approach zero. We will show later that the probability that the convergence condition is satisfied will approach one with proper block size. That is to say, the iterative estimate will eventually approach the linear MMSE estimate.

6.2.7 Computational Complexity Analysis

We now evaluate the computational complexity of the block diagonal matrix based detection scheme. We adopt the number of real-valued multiplications for the analysis of the computational complexity [85]. It is clear from **Algorithm 6.2**, the overall computational complexity of the block diagonal matrix based detection scheme consists of two components: the initialization, and iteration steps.

In the initialization step, the computations of $\mathbf{G}_k$, $\mathbf{W}_k$, and the matched-filter output $\mathbf{y}_k^{\mathrm{MF}}$, are essential for most of the detection schemes [23,84–86]. Therefore, we only consider the computations for the initial estimation and matrix inversion. In **Algorithm 6.2**, we employ the parallel processing structure for the initial estimation (steps 3-5). For each block, as $\mathbf{D}_k(l)$ is a diagonal matrix, the real-valued multiplications required for each block is $2N_L$. We utilize the **Algorithm 6.1** for the matrix inversion. Suppose the complexity for a size M Hermitian matrix is $\mathrm{T}(M)$, according to the **Loop** in **Algorithm 6.1**, we have

$$\mathrm{T}(M+1) = \mathrm{T}(M) + (4M-2)M + (2M+1) + 2M + 2M^2, \tag{6.47}$$

where $(4M-2)M$, $(2M+1)$[3], $2M$, and $2M^2$, account for the number of real-valued multiplications in steps 3, 4, 5, 6 in **Algorithm 6.1**, respectively. It is worth noting that the

[3]We take the division operation in step 4 as one operation. This is to avoid $\mathrm{T}(1) = 0$ when counting the operations for size 1 matrix. However, as we can see, the overall number of division operations is M for size

property of a Hermitian matrix is applied; as a result, only the diagonal and above diagonal elements are considered for the computational complexity analysis. With Equation (6.47), and let $\mathrm{T}(1) = 1$, we can derive that

$$\mathrm{T}(M) = 2M^3 - 2M^2 + M. \tag{6.48}$$

That is to say, the proposed matrix inversion method requires $2N_L^3 - 2N_L^2 + N_L$ real-valued multiplications (and divisions) for each block matrix. Using the previous mentioned Cholesky decomposition based matrix inversion method, $\frac{7}{3}N_L^3 - \frac{1}{2}N_L^2 - \frac{5}{6}N_L$ real-valued multiplications (and divisions) are required. Both methods involves N_L divisions. Therefore, compared to the Cholesky decomposition based matrix inversion, our proposed matrix inversion method requires fewer computations.

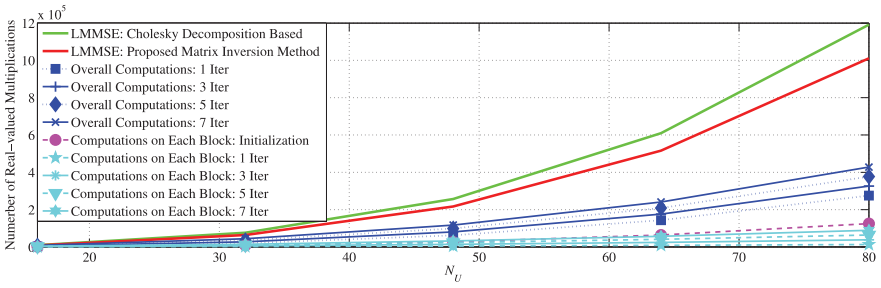
In the iteration step, the processing on each block involves two cascaded steps. The preprocessing in step 11 (**Algorithm 6.2**) requires $4(N_U - N_L)N_L$ multiplications. The updating process in step 12 involves $4N_L^2 - 2N_L$ real-valued multiplications. Therefore, the overall computational complexity required on each block is $4N_UN_L - 2N_L$.

To summarize, we present the computational complexity comparison of the proposed block diagonal matrix inversion based detection method with the linear MMSE detection scheme in Figure 6.4. For comparison, we also present the computations using the Cholesky decomposition based matrix inversion method as a benchmark. The following observations can be found.

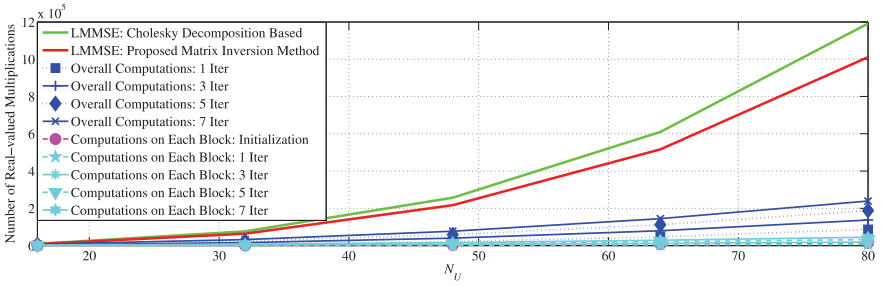
1) It is clear from Figure 6.4 that with the proposed matrix inversion method, the number of real-valued multiplications is less than that of using the Cholesky decomposition based matrix inversion method;

2) With the proposed block diagonal matrix based iterative method, the overall compu-

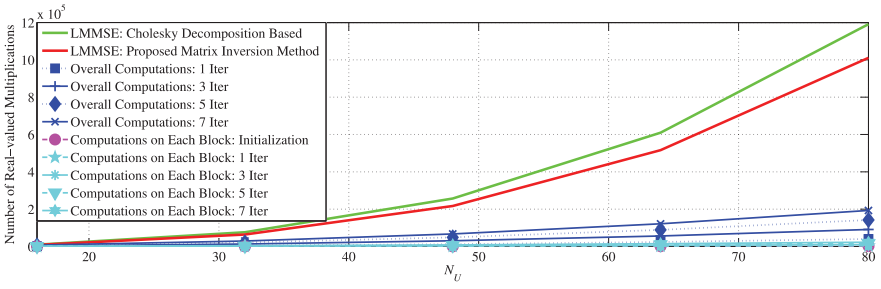
M matrix, which is the same as Cholesky decomposition method.



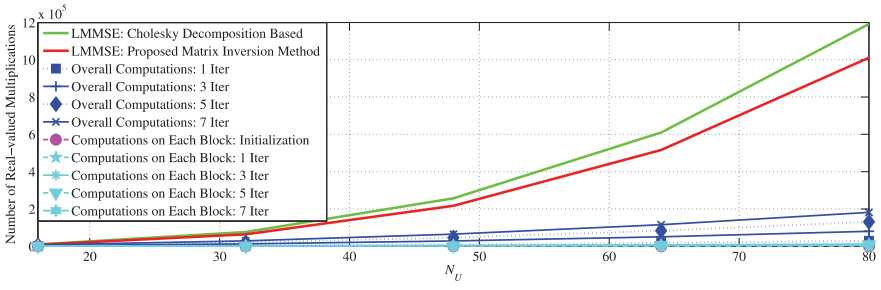
(a) $B = 2$



(b) $B = 4$



(c) $B = 8$



(d) $B = 16$

Figure 6.4: Computational complexity comparison for different block size

tations have been greatly reduced even with a large number iterations. In addition, from Figure 6.4(a) to Figure 6.4(d), it is clear that the overall computations decrease with the increase of block size;

3) With parallel processing, we can see that the computations on each block maintain very low level from various blocks. Furthermore, from Figure 6.4(a) to Figure 6.4(d), we can see that the increase of block size leads to the decrease of the computations on each block.

The results in Figure 6.4 show that proposed matrix inversion method is low complexity compared to the well-known Cholesky decomposition based matrix inversion method. Besides, the proposed iterative method based on block diagonal matrix inversion shows obvious superiority of low complexity in terms of overall computations and computations on each individual block.

6.3 Block Gauss-Seidel Method

As shown previously, the Gauss-Seidel method is of low complexity and has a fast convergence rate. However, two limitations with the original Gauss-Seidel method can be found: first, if the vehicular user number is large, the successive detection will introduce large processing delay for the last user; second, as the detection for later users needs to wait until all previous users have been detected, the successive signal processing structure is not efficient for hardware implementation. Inspired by the block diagonal matrix based detection scheme in previous section, we can employ the Gauss-Seidel method on small blocks.

6.3.1 Development of the Proposed Scheme

By utilizing the block diagonal matrix $\mathbf{P}_k$ in Equation (6.28), we have the updating progress given by Equation (6.29). By transforming the matrix inversion in Equation (6.29) into linear equations, we have

$$\mathbf{P}_k \hat{\mathbf{x}}_k^{(i+1)} = \tilde{\mathbf{y}}_k^{(i)}. \tag{6.49}$$

Since $\mathbf{P}_k$ is a block diagonal matrix, Equation (6.49) is equivalent to B independent linear equations, given by

$$\mathbf{P}_k(l)\,\hat{\mathbf{x}}_k^{(i+1)}(l) = \tilde{\mathbf{y}}_k^{(i)}(l), \tag{6.50}$$

where $l = 0, 1, B-1$. It is worth noting that $\mathbf{P}_k(l)$ is a Hermitian matrix as well; therefore, the small size linear equation (6.50) can be solved with Gauss-Seidel method. To be specific, we can divide $\mathbf{P}_k(l)$ as

$$\mathbf{P}_k(l) = \mathbf{D}_k(l) + \mathbf{T}_k(l) + \mathbf{T}_k^{\mathrm{H}}(l), \tag{6.51}$$

where $\mathbf{D}_k(l)$ is the diagonal matrix with the diagonal elements taken from the diagonal elements in $\mathbf{P}_k(l)$; $\mathbf{T}_k(l)$ is a lower triangular matrix with the diagonal elements all zeros and the below diagonal elements consisting of the below diagonal elements in $\mathbf{P}_k(l)$. Therefore, we have the updating progress in Equation (6.50) given by

$$\hat{\mathbf{x}}_k^{(i+1)}(l) = (\mathbf{D}_k(l) + \mathbf{T}_k(l))^{-1}\left(\tilde{\mathbf{y}}_k^{(i)}(l) - \mathbf{T}_k^{\mathrm{H}}(l)\,\hat{\mathbf{x}}_k^{(i)}(l)\right). \tag{6.52}$$

Similar to Equation (6.26), as $(\mathbf{D}_k(l) + \mathbf{T}_k(l))$ is a lower triangular matrix, successive detection can be introduced for the updating process.

It is noted that when the block size $B = 1$, the block Gauss-Seidel method reduces to the original Gauss-Seidel method; when $B = N_U$, the block Gauss-Seidel method becomes the Jacobi method. Although we introduce successive detection for Equation (6.52), the processing delay for the last users in the block is just $1/B$ of that in the original Gauss-Seidel method. This is quite obvious as the block matrix $\mathbf{P}_k(l)$ is of size $N_L = N_B/B$. Therefore,

instead of a long delay time, our proposed block Gauss-Seidel method has the advantage of much lower processing delay for one iteration. In addition, by using the Gauss-Seidel method on each block, we successfully avoid matrix inversion (even though the matrix size is small), which leads to further computations reduction. Moreover, the independent processing on each block can be implemented with a parallel processing structure, which is much efficient in hardware implementation.

6.3.2 Parallel Processing

With the availability of the previous estimate $\hat{\mathbf{x}}_k^{(i)}$, we update $\tilde{\mathbf{y}}_k^{(i)}(l)$ for each block according to Equation (6.32), and then perform the Gauss-Seidel method according to Equation (6.52). Therefore, we can utilize the parallel processing structure to complete the preprocessing and updating progress, which helps to greatly reduce the overall processing delay.

6.3.3 Residual Estimation Error

From Equation (6.52), we can derive that the matrix used for block Gauss-Seidel method, given by

$$\mathbf{\Gamma}_k = \begin{bmatrix} \mathbf{\Gamma}_k(0) & & & \\ & \mathbf{\Gamma}_k(1) & & \\ & & \ddots & \\ & & & \mathbf{\Gamma}_k(B-1) \end{bmatrix}, \tag{6.53}$$

where $\mathbf{\Gamma}_k(l) = \mathbf{D}_k(l) + \mathbf{T}_k(l)$. Similar to the previous Section, we have the ℓ_2-norm of the residual estimation error of the block Gauss-Seidel method given by

$$\begin{aligned} \tilde{J}^{(i)} &= \left\| \left(-\mathbf{\Gamma}_k(\mathbf{W}_k - \mathbf{\Gamma}_k)\right)^i \mathbf{e}_k^{(0)} \right\|_2 \\ &\leqslant \left\| \left(\mathbf{I}_{N_U} - \mathbf{\Gamma}_k \mathbf{W}_k\right)^i \right\|_{\mathrm{F}} \left\| \mathbf{e}_k^{(0)} \right\|_2. \end{aligned} \tag{6.54}$$

It is obvious that when the convergence condition, i.e., $\lim_{i\to\infty} \left(\mathbf{I}_{N_U} - \mathbf{\Gamma}_k \mathbf{W}_k\right)^i = \mathbf{0}$, is satisfied, the residual estimation error will approach 0. As we show later, with proper block size,

the probability that the convergence condition is satisfied approaches 1. Therefore, the proposed block Gauss-Seidel method can achieve near linear MMSE estimation.

6.3.4 Computational Complexity Analysis

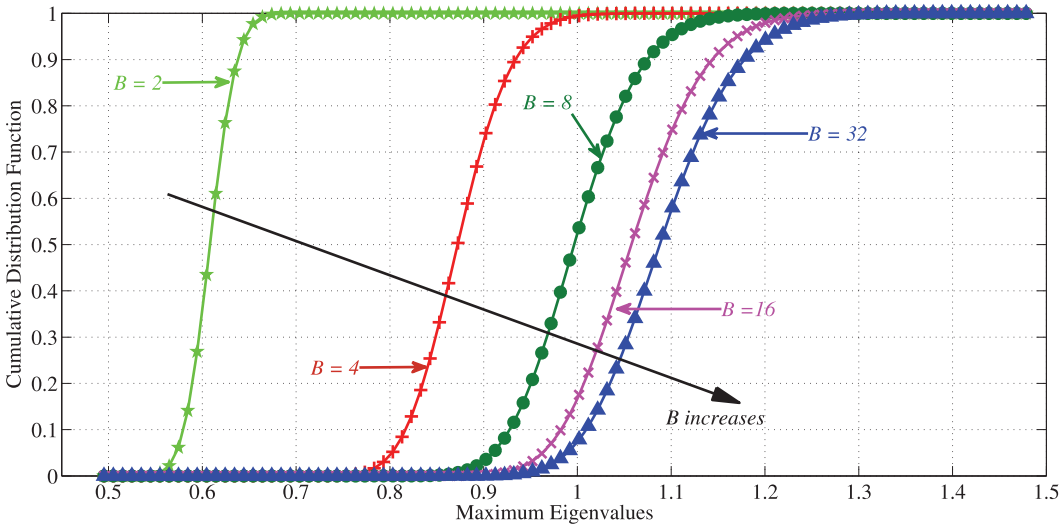
Compared to the block diagonal matrix based detection scheme, we successfully avoid matrix inversion. Instead, we introduce the Gauss-Seidel method for the updating progress. As the matrix size is N_L, the computations required for the updating on each block is $4N_L^2 - 2N_L$. The preprocessing is the same as the block diagonal matrix based detection scheme. Therefore, the overall amount of real-valued multiplications is $4N_U N_L - 2N_L$ on each block. Compared to the block diagonal matrix based detection scheme, block Gauss-Seidel method consumes less computations in initialization. Therefore, block Gauss-Seidel method is of low complexity and the computations on each block are greatly reduced with a parallel processing structure.

6.4 Numerical Simulations and Performance Evaluation

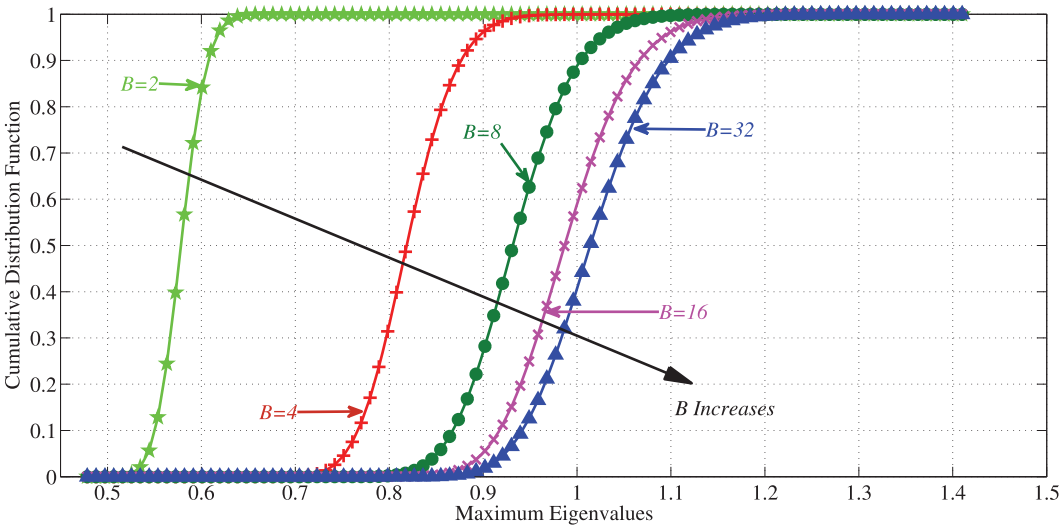
6.4.1 Convergence Conditions

We first investigate the convergence performance of the block diagonal matrix based detection method. Given the initial estimation $\hat{\mathbf{x}}_k^{(0)}$, according to Equation (6.29), we have

$$\begin{aligned}
\hat{\mathbf{x}}_k^{(i+1)} =& \mathbf{P}_k^{-1}\mathbf{y}_k^{\mathrm{MF}} - \mathbf{P}_k^{-1}\mathbf{Q}_k\hat{\mathbf{x}}_k^{(i)} \\
=& \mathbf{P}_k^{-1}\mathbf{y}_k^{\mathrm{MF}} - \mathbf{P}_k^{-1}\mathbf{Q}_k\left(\mathbf{P}_k^{-1}\mathbf{y}_k^{\mathrm{MF}} - \mathbf{P}_k^{-1}\mathbf{Q}_k\hat{\mathbf{x}}_k^{(i-1)}\right) \\
=& \left(\mathbf{I} - \mathbf{P}_k^{-1}\mathbf{Q}_k\right)\mathbf{P}_k^{-1}\mathbf{y}_k^{\mathrm{MF}} + \left(\mathbf{P}_k^{-1}\mathbf{Q}_k\right)^2\hat{\mathbf{x}}_k^{(i-1)} \\
=& \cdots \\
=& \left(\mathbf{I} - \mathbf{P}_k^{-1}\mathbf{Q}_k + \cdots + \left(-\mathbf{P}_k^{-1}\mathbf{Q}_k\right)^i\right)\mathbf{P}_k^{-1}\mathbf{y}_k^{\mathrm{MF}} \\
&+ \left(-\mathbf{P}_k^{-1}\mathbf{Q}_k\right)^{i+1}\hat{\mathbf{x}}_k^{(0)}.
\end{aligned} \tag{6.55}$$

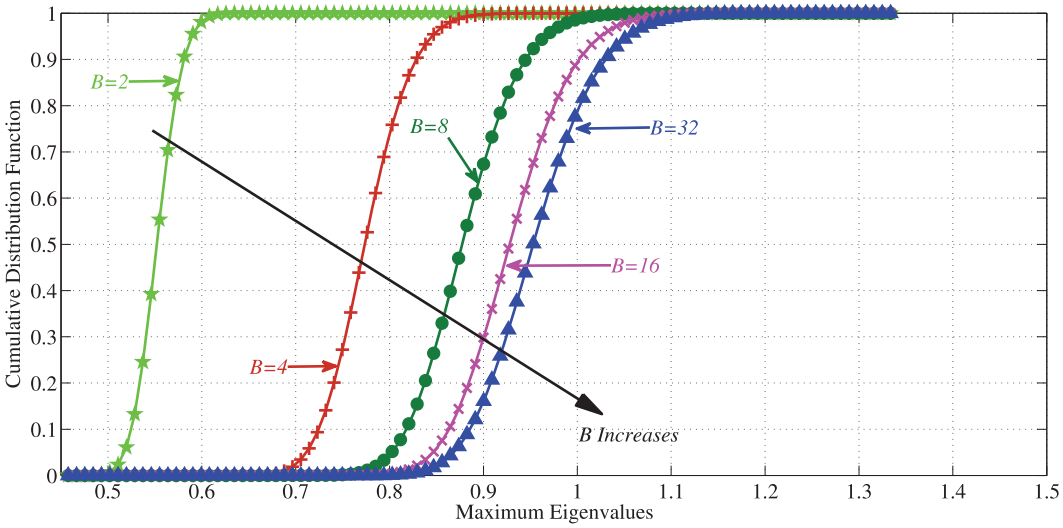


(a)

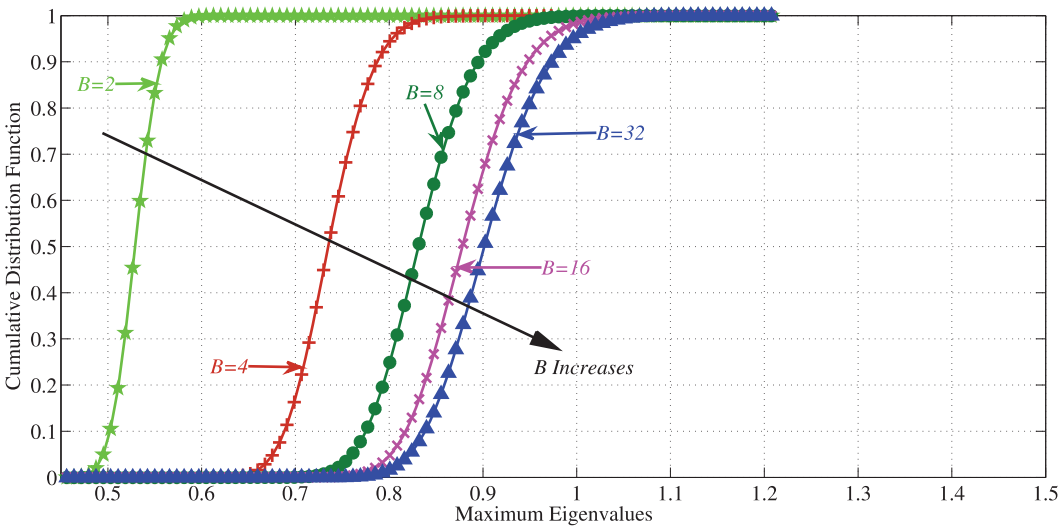


(b)

Figure 6.5: Cumulative distribution function of the maximum eigenvalues, block diagonal matrix based detection scheme, $N_U = 32$: (a) $N_B = 112$, (b) $N_B = 128$.



(a)



(b)

Figure 6.6: Cumulative distribution function of the maximum eigenvalues, block diagonal matrix based detection scheme, $N_U = 32$: (a) $N_B = 144$, (b) $N_B = 160$.

Therefore, the convergence condition for Equation (6.55) is

$$\lim_{i\to\infty}\left(\mathbf{P}_k^{-1}\mathbf{Q}_k\right)^i = \lim_{i\to\infty}\left(\mathbf{I}-\mathbf{P}_k^{-1}\mathbf{W}_k\right)^i = \mathbf{0}, \tag{6.56}$$

or equivalently,

$$\rho\left(\mathbf{I}-\mathbf{P}_k^{-1}\mathbf{W}_k\right) = |\lambda_0| < 1, \tag{6.57}$$

where $\rho(\mathbf{A})$ is the spectral radius of a matrix $\mathbf{A}$, and $|\lambda_0| \geqslant |\lambda_1| \geqslant \cdots \geqslant |\lambda_{N_U-1}|$ denote the N_U eigenvalues of $\left(\mathbf{I}-\mathbf{P}_k^{-1}\mathbf{W}_k\right)$.

We utilize Monte-Carlo simulations to obtain the cumulative distribution function (CDF) of λ_0 by randomly generating the channel matrix $\mathbf{H}_k$ [23]. And the results are shown in Figures 6.5 and 6.6, where $E_s/N_0 = 2$dB is used for the noise covariance computation. The following observations can be found:

1) It is noted that with the increase of B, the probability $\Pr(|\lambda_0| \leqslant a)$ decreases. For example, in Figure 6.5(a), $\Pr(|\lambda_0| \leqslant 1)$ for $B = 2$ is 1, and eventually decreases to 0.073 for $B = 32$. Similar results are shown in Figure 6.5(b), 6.6(a), and 6.6(b);

2) As $|\lambda_0| \leqslant 1$ is the convergence condition for the iterative methods, the results in Figure 6.5(a) indicates that $B \geqslant 8$ may be inapplicable for the system configuration. In particular, the Jacobi method (which corresponds $B = 32$) cannot converge for all cases. This is consistent with the results in [23, 80] where a large ratio of $r = N_B/N_U$ is required;

3) From Figure 6.5(a) to Figure 6.6(b), it is shown that with increasing N_B, the probability that the convergence conditions is satisfied (i.e., $\Pr(|\lambda_0| \leqslant 1)$) can be improved for various number of blocks. As a result, computational complexity for each iteration with larger B will be reduced (smaller block), and the processing delay can be further improved (high parallel efficiency);

4) The results in Figure 6.5 and 6.6 also reveal that $\Pr(|\lambda_0| \leqslant a)$, $a < 1$, is always greater for a smaller B than a larger B. As the maximum eigenvalue plays a dominant role in the convergence rate for iterative methods, we conclude that the convergence rate for smaller B is greater than that of a large B.

Therefore, on one hand, in order to maintain low complexity and processing delay, we require a large number of blocks; on the other hand, in order to achieve fast convergence, we require a small number of blocks. The results in Figure 6.5 and 6.6 can be utilized for a proper selection of B in practical implementation.

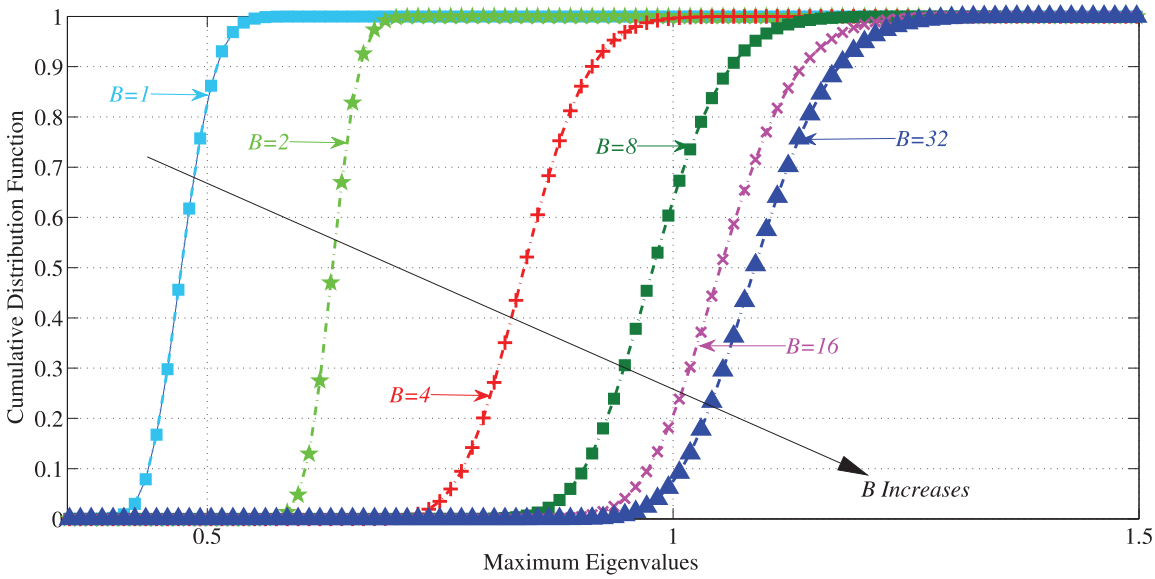
For the block Gauss-Seidel method, by replacing $\mathbf{P}_k$ with $\mathbf{\Gamma}_k$, we can also utilize the Monte-Carlo simulations to obtain the CDF of the maximum eigenvalue of the matrix $\left(\mathbf{I} - \mathbf{\Gamma}_k^{-1}\mathbf{W}_k\right)$, and the results are shown in Figure 6.7.

The CDF of the maximum eigenvalue for the original Gauss-Seidel method in [84] (corresponding to $B = 1$ in our case) and that for Jacobi method (corresponding to $B = N_U$) are presented in Figure 6.7 as benchmarks. Apparently, the original Gauss-Seidel method has a better convergence performance than the Jacobi method. This is validated from the following obvious observations:

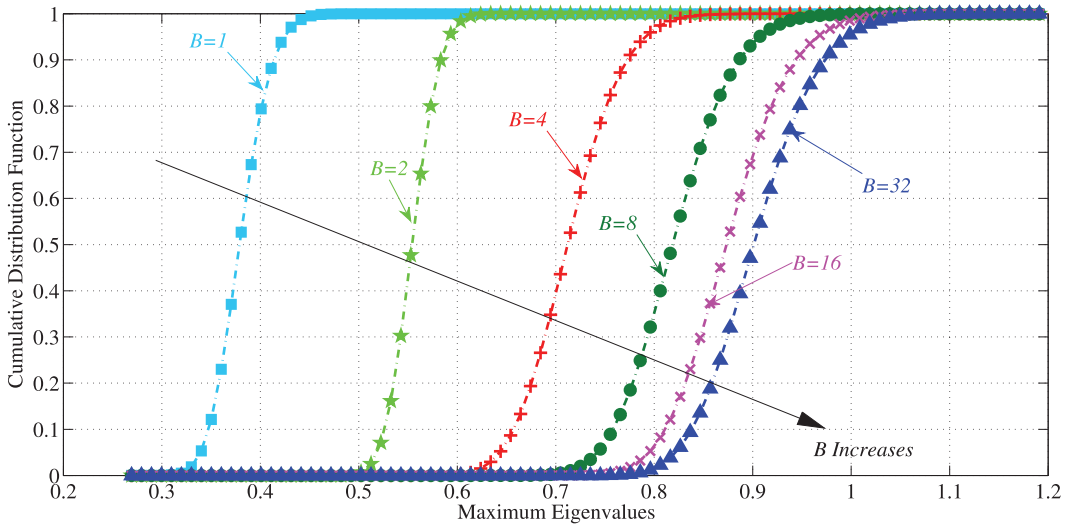
1) The probability that the convergence conditions are satisfied (i.e., $|\lambda_0| < 1$) is much greater for the original Gauss-Seidel method than that for the Jacobi method;

2) The probability $\Pr(|\lambda_0| \leqslant a)$, $a < 1$, for the original Gauss-Seidel method is much greater than that for the Jacobi method.

In particular, the latter observation indicates that the convergence rate is much faster than the Jacobi method. For $1 < B < N_U$, it is obvious that the convergence performance is between the original Gauss-Seidel method and the Jacobi method. However, when $B > 1$ is applicable, by utilizing the parallel processing, we can achieve near linear MMSE



(a) N_B = 112, N_U = 32



(b) N_B = 160, N_U = 32

Figure 6.7: Cumulative distribution function of the maximum eigenvalues, block Gauss-Seidel method.

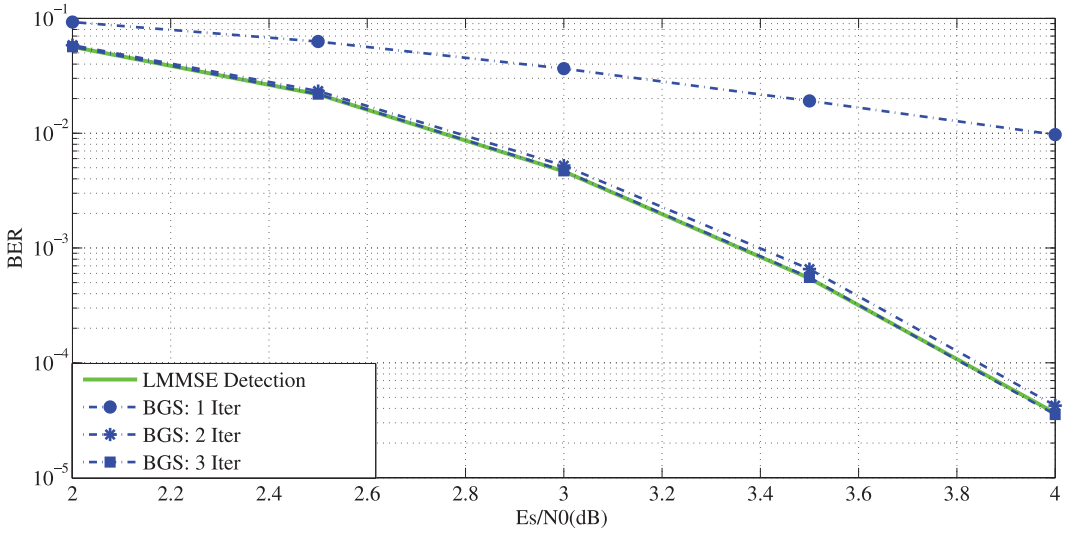
performance (when the performance converges) with fast processing time.

6.4.2 BER Performance

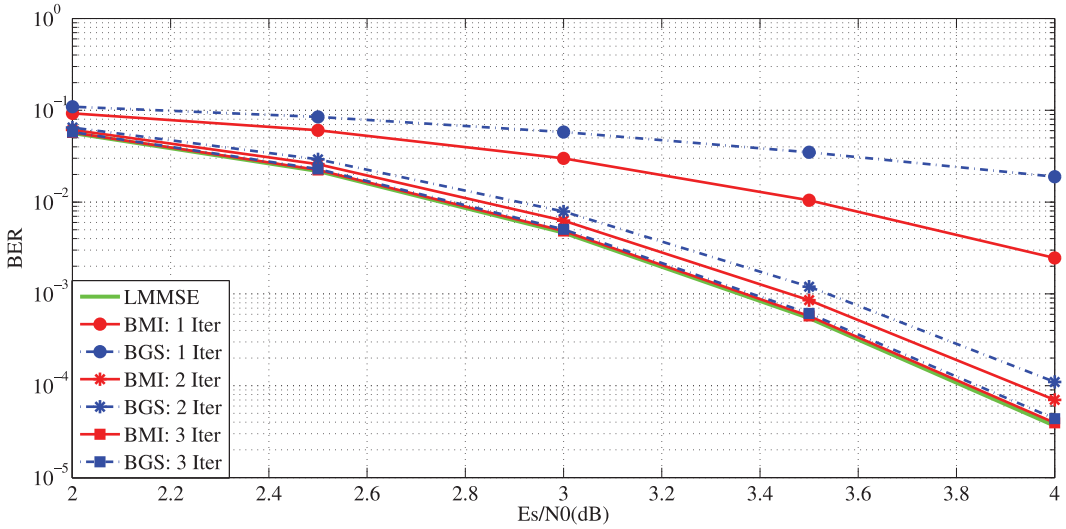
We now evaluate the system bit error rate performance. A LDPC code with code length 1152, and code rate 1/2 is adopted for the channel code scheme. The modulation scheme is 64QAM, which indicates each symbol carries 6 bits information. The extended vehicular A (EVA) model is used in the simulation to reflect the multi-path channel between the vehicles and the RSU.

In Figures 6.8, to 6.10, we present the BER performance of the proposed block diagonal matrix based detection scheme and the Gauss-Seidel method. The system configuration is 256×32, which indicates the base station is equipped with 256 receive antennas and 32 vehicular users are simultaneously served. For $B = 1$, the block diagonal matrix based detection scheme degrades to the linear MMSE detection scheme, and the block Gauss-Seidel method reduces to the original Gauss-Seidel method [84]. It is noted from Figure 6.8(a) that in this system configuration, the performance of the Gauss-Seidel method quickly converges to that of the linear MMSE detection scheme (only two iterations are required). For $B = N_U$, both the block diagonal matrix based detection scheme and the block Gauss-Seidel method become the well-known Jacobi method. From Figure 6.10(b), we can see that after 4-5 iterations, the BER performance of the Jacobi method converges to that of the linear MMSE detection scheme. For the number of blocks $1 < B < 32$, we present the BER performance in figures from Figure 6.8(b) to Figure 6.10(a). As we can see, with the increase of B, more iterations are required for the convergence of the system performance which is attributed to the slow convergence rate of larger B. This is consistent with the conclusions in the previous analysis.

The results in Figures 6.8, 6.9, and 6.10, also reveal the performance of the block diag-

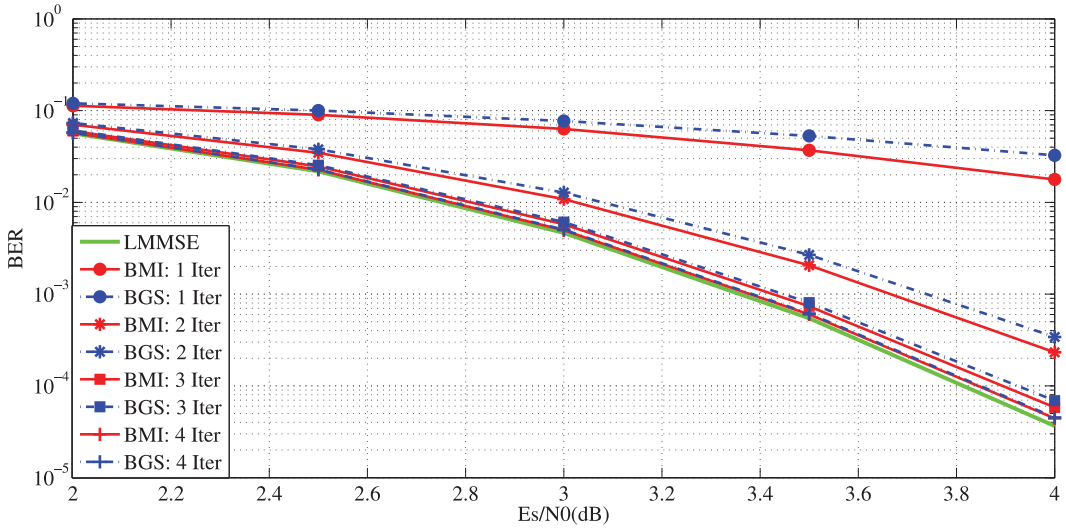


(a)

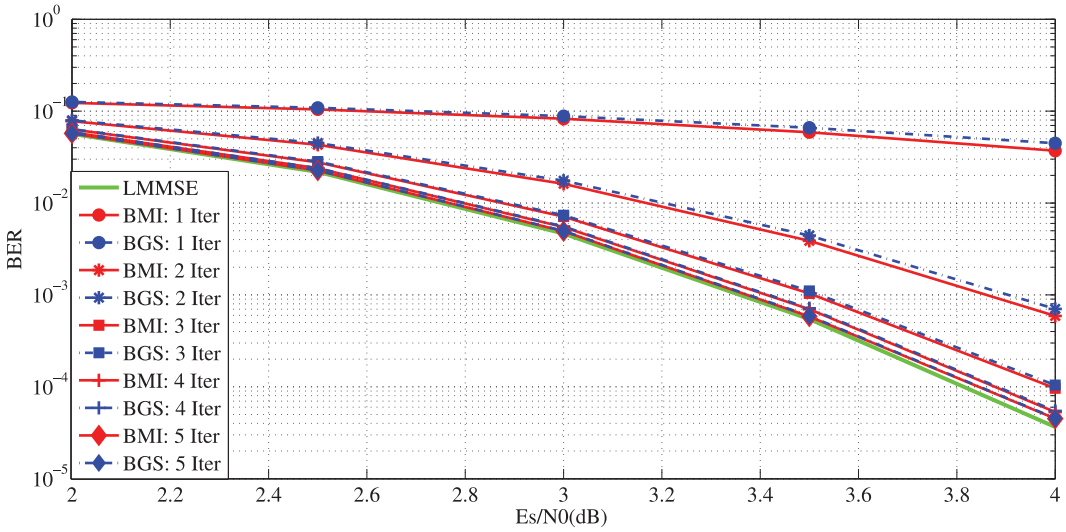


(b)

Figure 6.8: BER Performance, $N_B = 256$, $N_U = 32$. BMI: block diagonal matrix based detection scheme; BGS: block Gauss-Seidel method. (a) $B = 1$; (b) $B = 2$.

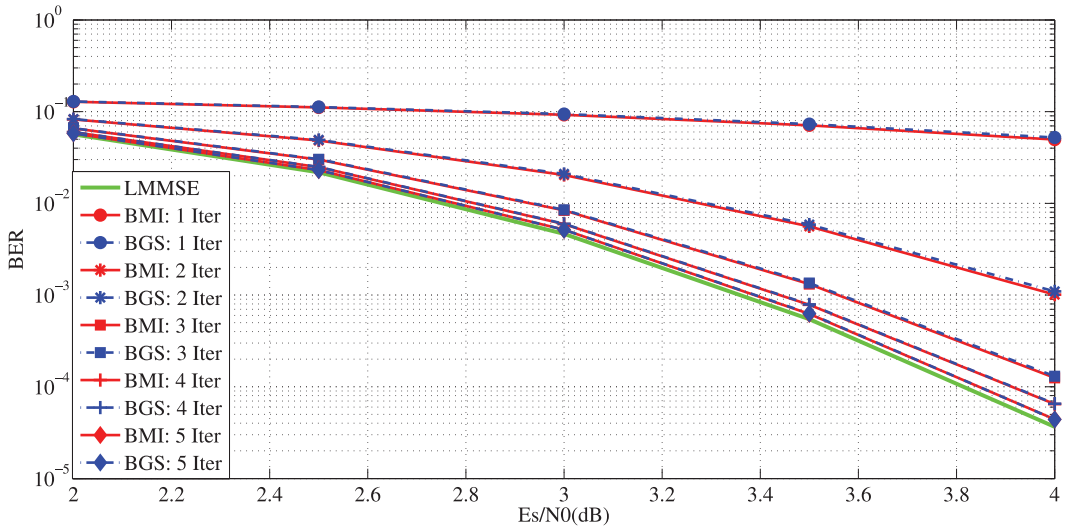


(a) $B = 4$

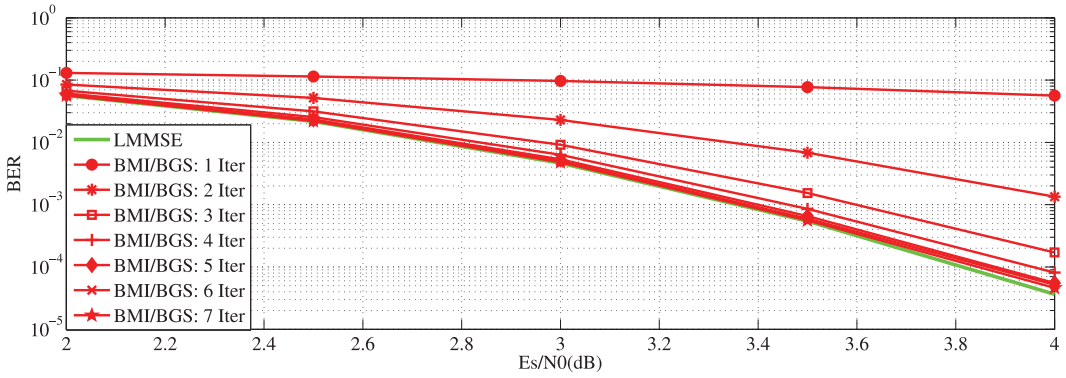


(b) $B = 8$

Figure 6.9: BER Performance, $N_B = 256$, $N_U = 32$. BMI: block diagonal matrix based detection scheme; BGS: block Gauss-Seidel method. (a) $B = 4$; (b) $B = 8$.



(a)



(b)

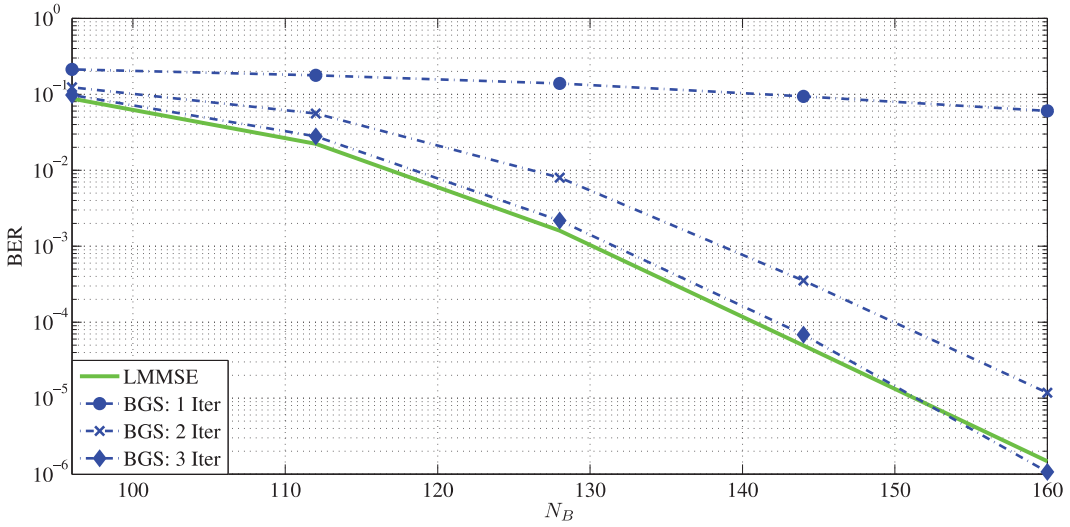
Figure 6.10: BER Performance, $N_B = 256$, $N_U = 32$. BMI: block diagonal matrix based detection scheme; BGS: block Gauss-Seidel method. (a) $B = 16$, (b) $B = 32$.

onal matrix based detection scheme is litter better than the Gauss-Seidel method when the same iterations are processed. This is understandable as the block Gauss-Seidel method utilizes the lower triangular matrix of the block diagonal matrix to update the next iteration. However, the results show that the block Gauss-Seidel method achieves very close performance compared to the block diagonal matrix based detection scheme, especially when the number of blocks increases.

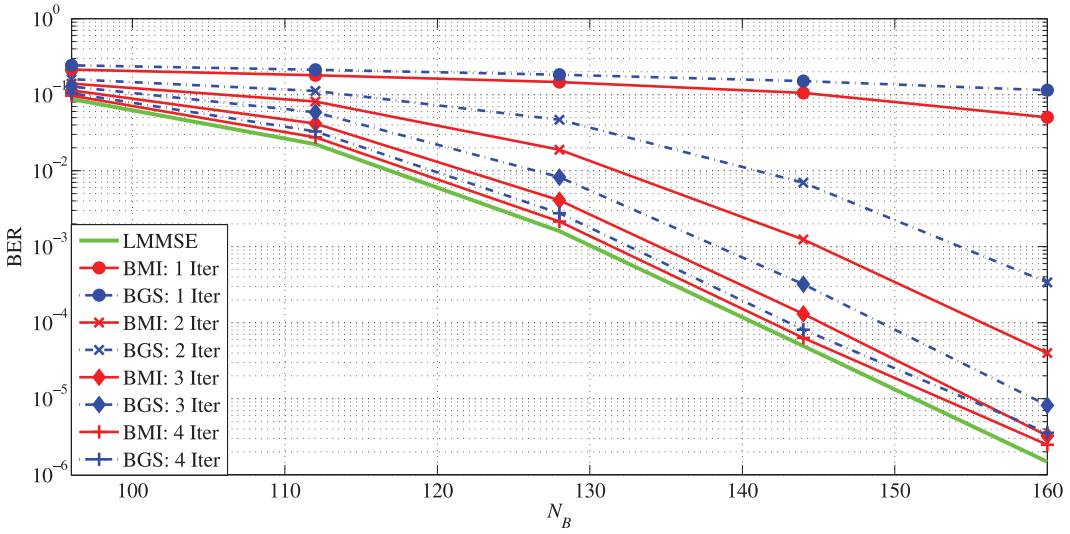
However, as shown previously, when the ratio $r = N_B/N_U$ is not sufficient large, the convergence conditions may not be satisfied for a large B. For better illustration, we present the BER performance with the increase of N_B, with fixed $N_U = 32$, shown in Figures 6.11 and 6.12. The average received signal-to-noise ratio ($E_s/N_0 = 7dB$). It is noted that from Figure 6.11(a) that the original Gauss-Seidel method (corresponds to $B = 1$) works well and it takes about three iterations to converge to the performance of the linear MMSE detection scheme. Figure 6.11(b) shows that about 4 iterations are required for both the block diagonal matrix based detection and block Gauss-Seidel method schemes to converge to the performance of the linear MMSE schemes. However, with the increase of B, we note from Figure 6.12(a) and Figure 6.12(b) that the performance of the block diagonal matrix based detection scheme and block Gauss-Seidel method cannot converge to that of the linear MMSE detection scheme, especially when N_B is not sufficient large. The results are also consistent with the previous analysis where we noted that when we have a large B, the convergence conditions may not be satisfied with a high probability, which causes performance gap compared to the linear MMSE detection scheme.

6.4.3 Processing Time

One of the most significant contributions of our work is to develop fast processing detection methods, and the parallel processing structure is proposed to speedup the processing time.

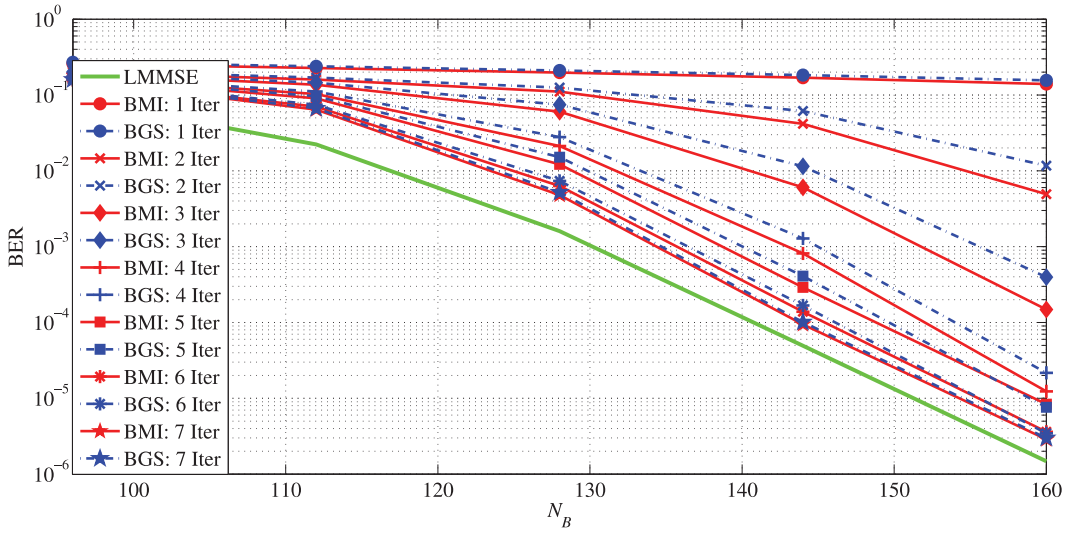


(a) $B = 1$

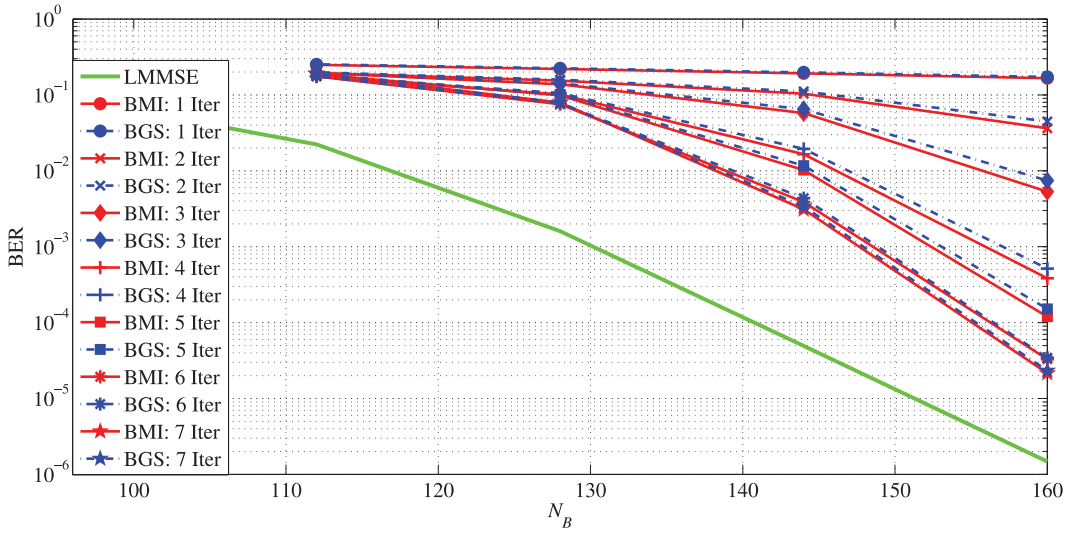


(b)

Figure 6.11: BER Performance versus N_B, $N_U = 32$, BMI: block diagonal matrix based detection scheme; BGS: block Gauss-Seidel method. (a) $B = 1$; (b) $B = 2$.



(a) $B = 4$



(b) $B = 8$

Figure 6.12: BER Performance versus N_B, $N_U = 32$, BMI: block diagonal matrix based detection scheme; BGS: block Gauss-Seidel method. (a) $B = 4$; (b) $B = 8$.

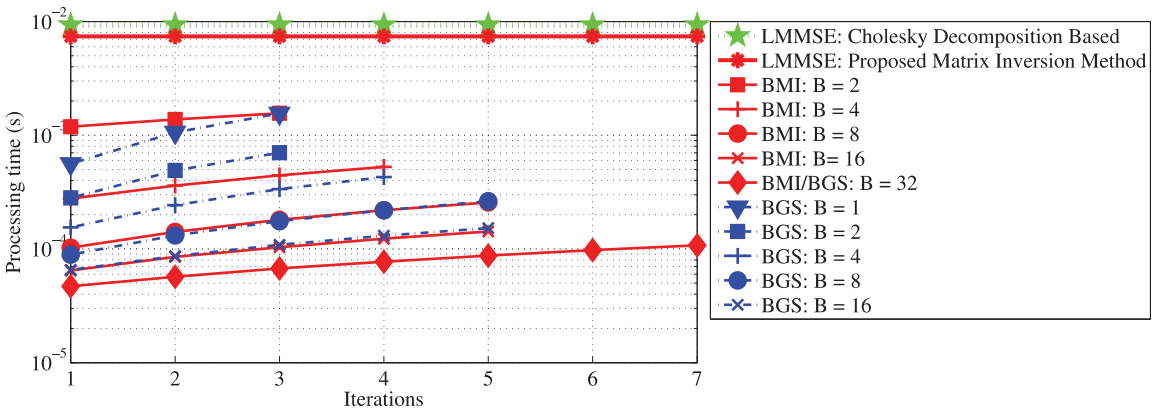


Figure 6.13: Processing time of the proposed detection schemes using parallel processing, BMI: block diagonal matrix based detection scheme; BGS: block Gauss-Seidel method.

In this subsection, we use the running time of the program to evaluate the processing time of the proposed schemes. In the proposed schemes, the processing on each individual block is independent. That is to say, the overall processing delay is reduced to the processing time on each block if the parallel processing structure is utilized. Therefore, we use the running time of the process on each block to reflect the overall processing time of the proposed schemes.

We use MATLAB to obtain the running time of the programs which are running on a 3.5-GHz Intel core i7-4770K CPU with a 16G RAM and MATLAB R2012a platform. The results are shown in Figure 6.13 with average of 400000 random simulations and the system configuration is the same as the setup in Figure 6.8. For comparison, we present the running time of the linear MMSE detection scheme as a benchmark where the Cholesky decomposition algorithm is used for matrix inversion.

From Figure 6.13, we can see the overall processing time is about 10ms (millisecond) if the Cholesky decomposition algorithm is used. The large processing delay comes from the required sufficient computations and back-forward solutions involved in Cholesky decomposition algorithm. With the proposed new matrix inversion method, we can successfully

reduce the overall processing time to 7ms (about 30% improvement). The improvement can be attributed to the computational complexity reduction which is shown in Figure 6.4. The significant improvement of the proposed detection schemes using parallel processing structure is clearly shown in Figure 6.13. Specifically, with the block diagonal matrix based detection scheme, the overall processing time is reduced to 1-2ms for $B = 2$ and that processing delay is dramatically reduced to less than 0.1ms when the block size increases to 32. For the block Gauss-Seidel method, we can see that in the original Gauss-Seidel method (corresponding to $B = 1$), the processing time is maintained at the 1ms level. With the increase of the block size, the overall processing delay is eventually reduced to less than 0.1ms. The significant reduction in processing time of the proposed schemes can be attributed to the significant computation reduction on each small block. Although the proposed schemes require sufficient iterations to converge, the results in Figure 6.13 indicate the increase of the processing delay with iterations is not so sensitive. This is attributed to the slightly increased computation on each block when more iterations are required (This is validated with the results in Figure 6.4). Figure 6.13 also indicates that the processing time of the block Gauss-Seidel method is less than the block diagonal matrix based detection scheme with the same block size and iterations. This is because the computation in the initialization steps for the block Gauss-Seidel method are fewer as no block matrix inversion is required. However, with the increase of block size, this gap eventually decreases since the computation of the matrix inversion on small blocks will be less significant.

6.4.4 Discussion

Numerical results have demonstrated that using the parallel processing structure, we can achieve near linear MMSE detection performance with low complexity and fast processing time. The results in Figure 6.4 and 6.13 indicate that the larger the block size is, the higher

the level of parallelism that can be achieved, resulting in much lower computation on each individual block and lower processing delay of the detection procedure. However, the results in Figures 6.11 and 6.12 indicate that if we increase block size, the detection performance may not converge to that of the linear MMSE detection scheme. This is revealed by the convergence performance evaluation in Figures 6.5, 6.6, and 6.7. Therefore, a proper block size should be selected to meet the convergence conditions.

6.5 Summary

In this chapter, we have studied the practical massive MIMO detection schemes in modern vehicular networks. We consider an application scenario where the roadside unit is equipped with a large antenna array to simultaneously serve multiple vehicles on the road. As the number of vehicles is huge, low complexity and fast processing detection schemes are highly demanded. We propose a block diagonal matrix based detection scheme and block Gauss-Seidel method to greatly reduce the overall computational complexity in linear MMSE detection. Using the properties of the block matrix, we successfully degrade a large-scale array issue to a small-scale one, and propose parallel processing structures for real implementation. The parallel processing structures for the proposed schemes can greatly reduce the overall processing delay in iterative methods and is efficient for hardware implementation. We also investigate the convergence performance of the proposed scheme by studying the probability that the convergence conditions are satisfied. Finally, we validate the system BER performance in a practical system configuration. The results show that with proper block size, the performance of the proposed schemes can eventually approach the performance of the linear MMSE detection scheme, but requires low computational complexity and low processing delay.

Chapter 7

Efficient and Fast Processing Large Array Signal Detection in Underwater Acoustic Communications

[1] Underwater wireless communications have been essential for humans to investigate, and exploit the oceans. Various applications, such as off-shore oil field monitoring, remote control in oil and gas industry, disaster detection and warning, underwater surveillance, as well as military applications, demand the deployment of underwater wireless networks [13]. Among those wireless techniques, underwater acoustic (UWA) communications have the advantages

[1]The related work has been published in or submitted to

- *IEEE Global Communications Conference* (GLOBECOM'18), Abu Dhabi, UAE, 2018, under review.
- *Proceedings of the IEEE Global Communications Conference* (GLOBECOM'17), Singapore, December 2017.

of long distance transmission, which have appealed extensive attention [14–18, 108]. However, to provide reliable and efficient communications through UWA channels is a challenging task [14, 15, 109]. First of all, the UWA channel is known to be highly environmentally dependent [14]. Various factors, including water depth, wind, salinity, geometry, frequency band, as well as other environmental conditions, can affect the UWA channel impulse response [14, 15]. Besides, UWA channels are characterised by huge numbers of multi-path, long propagation delay, and attenuation [14]. Last but not least, the available frequency band for UWA communications is limited, which prohibits the applications of UWA communications to some of the underwater services where high throughput is requested [110].

Deploying a large-scale array of hydrophones for underwater acoustic (UWA) communications has numerous benefits, including high spectrum and energy efficiency, high rate communication. However, along with the benefits, large array signal processing is computational costly and requires long processing delay. Using the Gauss-Seidel method, we can achieve low complexity and good system performance. However, the Gauss-Seidel method introduces successive data detection, causing large processing delay and low efficiency for hardware implementation. In this Chapter, we apply the block Gauss-Seidel method for large array signal detection in UWA communications. In the proposed scheme, the Gauss-Seidel method is performed on a small size block matrix, and the processing on each block can be paralleled. Therefore, the total processing delay is greatly reduced. Moreover, the parallel processing structure is quite efficient for hardware implementation. We also utilize the UWA channel model developed in recent work to investigate the performance of the proposed scheme, and the results are promising.

The rest of this chapter is organized as follows. In Section 7.1, we briefly introduce the system model. We present the proposed low-complexity and fast processing detection scheme in Section 7.2. The performance evaluation and discussion are presented in Section

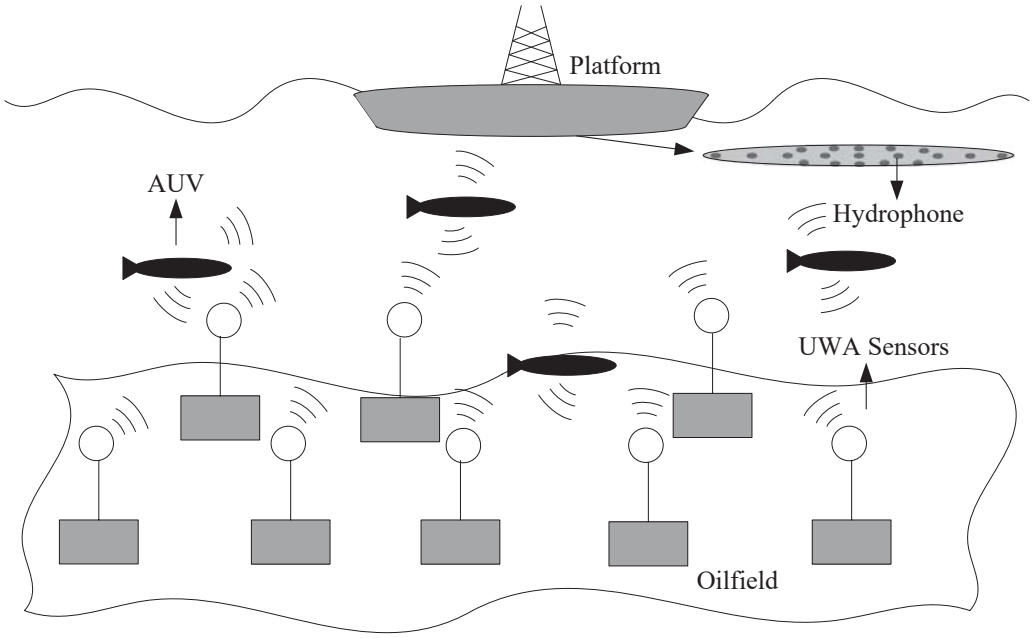


Figure 7.1: An application scenario of UWA communications with large-scale hydrophones at platform.

7.3. Finally, we draw the conclusions in Section 7.4.

7.1 System Model

We consider an UWA communication system for offshore oilfield monitoring, shown in Figure 7.1. Around the oilfield, a distributed UWA sensor network is deployed on the seafloor. The sensors collect real-time data, including conditions of the construction, environmental data, and status of the pipeline, and report to the platform through UWA channels. The autonomous underwater vehicles (AUVs) are also periodically launched for emergent data collection. At the platform where data is collected, a large array of hydrophones are deployed. The platform can simultaneously communicate with a large number of AUVs and UWA sensors. When the number of AUVs and sensors is large, the signal processing at

the platform will be computational costly. In this application scenario, a fast and efficient signal detection scheme is significant.

With OFDM modulation, the received signal vector over the k-th subcarrier, $\mathbf{y}_k \in \mathbb{C}^{N_R \times 1}$, is given by [15]

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{z}_k, \tag{7.1}$$

where $\mathbf{x}_k \in \mathbb{C}^{N_V \times 1}$ with $x_{k,u}$ denoting the transmitted symbol at the k-th subcarrier from the u-th AUV (or UWA sensor). $\mathbf{z}_k$ denotes the noise vector satisfying $\mathrm{E}\left\{\mathbf{z}_k \mathbf{z}_k^{\mathrm{H}}\right\} = \sigma_z^2 \mathbf{I}_{N_R}$ and $\mathbf{H}_k \in \mathbb{C}^{N_R \times N_V}$ represents the channel matrix. N_R is the number of hydrophones, and N_V is the number of AUVs or UWA sensors in communication ($N_R \gg N_V$).

7.1.1 UWA Channel Model

We utilize the statistic channel model developed in [14], where the time-varying channel transfer function is given by

$$H(f,t) = \bar{H}_0 \sum_{p=0}^{P-1} h_p \tilde{\gamma}_p(f,t)\, e^{-j2\pi f \tau_p}, \tag{7.2}$$

where $\tilde{\gamma}_p(f,t) = \gamma_p(f,t)\, e^{j2\pi a_p f t}$ is the overall small-scale coefficient for path p. h_p and τ_p describes the large-scale coefficient and the path delay. a_p denotes the Doppler scaling factor.

It has been experimentally validated that the channel modelling in Equation (7.2) is able to account for both large scale and small scale variations in the UWA channel. Therefore, we will utilize this model as our channel model.

7.1.2 Gauss-Seidel Method

As demonstrated in [3,111], when $N_R \gg N_V$, by employing a linear minimum mean-square error (MMSE) detection scheme, we can achieve near-optimal system performance. For

Equation (7.1), the linear MMSE estimate is given by

$$\hat{\mathbf{x}}_k = \left(\mathbf{H}_k^{\mathrm{H}}\mathbf{H}_k + \sigma_z^2\mathbf{I}_{N_V}\right)^{-1}\mathbf{H}_k^{\mathrm{H}}\mathbf{y}_k = \mathbf{W}_k^{-1}\mathbf{y}_k^{\mathrm{MF}}, \tag{7.3}$$

where $\mathbf{W}_k = \mathbf{H}_k^{\mathrm{H}}\mathbf{H}_k + \sigma_z^2\mathbf{I}_{N_V}$ is the equalization matrix, and $\mathbf{y}_k^{\mathrm{MF}} = \mathbf{H}_k^{\mathrm{H}}\mathbf{y}_k$ is the matched filter output over the k-th subcarrier.

The estimate in (7.3) involves matrix inverse; therefore, the computational complexity is high when the matrix size of $\mathbf{W}_k$ is large. To maintain low complexity in detection, iterative methods are usually adopted. Among those iterative methods, Gauss-Seidel method [84] has the advantages of fast convergence rate and better system performance compared to Neumann series expansion based detection schemes [23].

In the Gauss-Seidel method, Equation (7.3) is first rewritten to

$$\mathbf{W}_k\hat{\mathbf{x}}_k = \mathbf{y}_k^{\mathrm{MF}}. \tag{7.4}$$

That is, the matrix inversion problem is transformed into solving linear equations. Gauss-Seidel method is to utilize iterative estimation to approach the solutions to linear equation. This is implemented by providing an initial estimate, then following the Gauss-Seidel iterative estimation structure, and finally, providing the final estimate when it converges.

Specifically, the iterative estimate is given by

$$\hat{\mathbf{x}}_k^{(i+1)} = \left(\mathbf{D}_k + \mathbf{L}_k\right)^{-1}\left(\mathbf{y}_k^{\mathrm{MF}} - \mathbf{L}_k^{\mathrm{H}}\hat{\mathbf{x}}_k^{(i)}\right). \tag{7.5}$$

where $\mathbf{D}_k$ is a diagonal matrix with the diagonal elements taken from $\mathbf{W}_k$, and $\mathbf{L}_k$ consists of the elements below the diagonal elements in $\mathbf{W}_k$. $(\mathbf{D}_k + \mathbf{L}_k)$ is a lower triangular matrix; therefore, successive detection can be adopted in Equation (7.5) to update the next iterative estimate.

7.2 Block Gauss-Seidel Method

Although theGauss-Seidel method has the advantages of low complexity and fast convergence rate, it suffers from large processing delay since successive detection is introduced in each iteration. To speed up the processing time in each iteration, we propose the block Gauss-Seidel method in this section.

7.2.1 The Development

To begin with, we extract a block diagonal matrix, $\mathbf{P}_k$, from the equalization matrix $\mathbf{W}_k$, given by

$$\mathbf{P}_k = \begin{bmatrix} \mathbf{P}_k(0) & & & \\ & \mathbf{P}_k(1) & & \\ & & \ddots & \\ & & & \mathbf{P}_k(B-1) \end{bmatrix}, \tag{7.6}$$

where B is the number of blocks. The l-th block $\mathbf{P}_k(l)$ consists of the elements at the rows and the columns both from $l \cdot N_L$ to $(l+1) \cdot N_L - 1$ in $\mathbf{W}_k$, where N_L is the size of $\mathbf{P}_k(l)$ (suppose $N_V = BN_L$. If not, the last block takes the rest of the elements with a smaller size). We then rewrite the iterative estimation of $\hat{\mathbf{x}}_k$ as

$$\mathbf{P}_k \hat{\mathbf{x}}_k^{(i+1)} = \left(\mathbf{y}_k^{\mathrm{MF}} - \mathbf{Q}_k \hat{\mathbf{x}}_k^{(i)} \right), \tag{7.7}$$

where $\mathbf{Q}_k = \mathbf{W}_k - \mathbf{P}_k$, which takes the remaining elements in $\mathbf{W}_k$.

Define $\tilde{\mathbf{y}}_k^{(i)} = \mathbf{y}_k^{\mathrm{MF}} - \mathbf{Q}_k \hat{\mathbf{x}}_k^{(i)}$ which is the preprocessed vector after the i-th iteration. Using the properties of block diagonal matrix, we divide $\hat{\mathbf{x}}_k^{(i+1)}$ and $\tilde{\mathbf{y}}_k^{(i)}$ into B blocks, with each block consisting of N_L elements. We then have

$$\mathbf{P}_k(l) \hat{\mathbf{x}}_k^{(i+1)}(l) = \tilde{\mathbf{y}}_k^{(i)}(l). \tag{7.8}$$

It is worth noting that in Equation (7.8), the size of $\mathbf{P}_k(l)$ is N_L, which is only $1/B$ of the original matrix $\mathbf{P}_k$. This indicates the updating progress in Equation (7.7) can be implemented in a small scale matrix. It is easy to derive that $\mathbf{P}_k(l)$ is a Hermitian matrix; therefore, we can divide $\mathbf{P}_k(l)$ into

$$\mathbf{P}_k(l) = \mathbf{D}_k(l) + \mathbf{E}_k(l) + \mathbf{E}_k^{\mathrm{H}}(l), \tag{7.9}$$

where $\mathbf{D}_k(l)$ is the diagonal matrix of $\mathbf{P}_k(l)$, and $\mathbf{E}_k(l)$ consists of the elements below the diagonal elements in $\mathbf{P}_k(l)$. Performing the Gauss-Seidel method on the small block, we have

$$(\mathbf{D}_k(l) + \mathbf{E}_k(l))\,\hat{\mathbf{x}}_k^{(i+1)}(l) = \left(\tilde{\mathbf{y}}_k^{(i)}(l) - \mathbf{E}_k^{\mathrm{H}}(l)\,\hat{\mathbf{x}}_k^{(i)}(l)\right). \tag{7.10}$$

The same with Equation (7.5), successive detection can be adopted to solve (7.10).

With the previous estimation, $\tilde{\mathbf{y}}_k^{(i)}$ can be preprocessed; therefore, the updating progress on the small block matrix can be speeded up, without waiting for a long processing delay in the original Gauss-Seidel method. The independent updating progress can be implemented with parallel processing structure, which is detailed in the next subsection.

7.2.2 Parallel Processing

The preprocessing of $\tilde{\mathbf{y}}_k^{(i)}$ can be implemented on small size block, given by

$$\tilde{\mathbf{y}}_k^{(i)}(l) = \mathbf{y}_k^{\mathrm{MF}}(l) - \mathbf{Q}_k(l)\,\hat{\mathbf{x}}_k^{(i)}, \tag{7.11}$$

where $\mathbf{Q}_k(l)$ consists of rows from $l \cdot N_L$ to $(l+1) \cdot N_L - 1$ in $\mathbf{Q}_k$. $\mathbf{y}_k^{\mathrm{MF}}(l)$ is given by

$$\mathbf{y}_k^{\mathrm{MF}}(l) = \mathbf{H}_k^{\mathrm{H}}(l)\,\mathbf{y}_k, \tag{7.12}$$

where $\mathbf{H}_k^{\mathrm{H}}(l)$ consists of the rows from $l \cdot N_L$ to $(l+1) \cdot N_L - 1$ in $\mathbf{H}_k^{\mathrm{H}}$.

After the preprocessing on each block, the block Gauss-Seidel method is performed on each block. Therefore, we summarize the parallel processing structure in Figure 7.2. It is

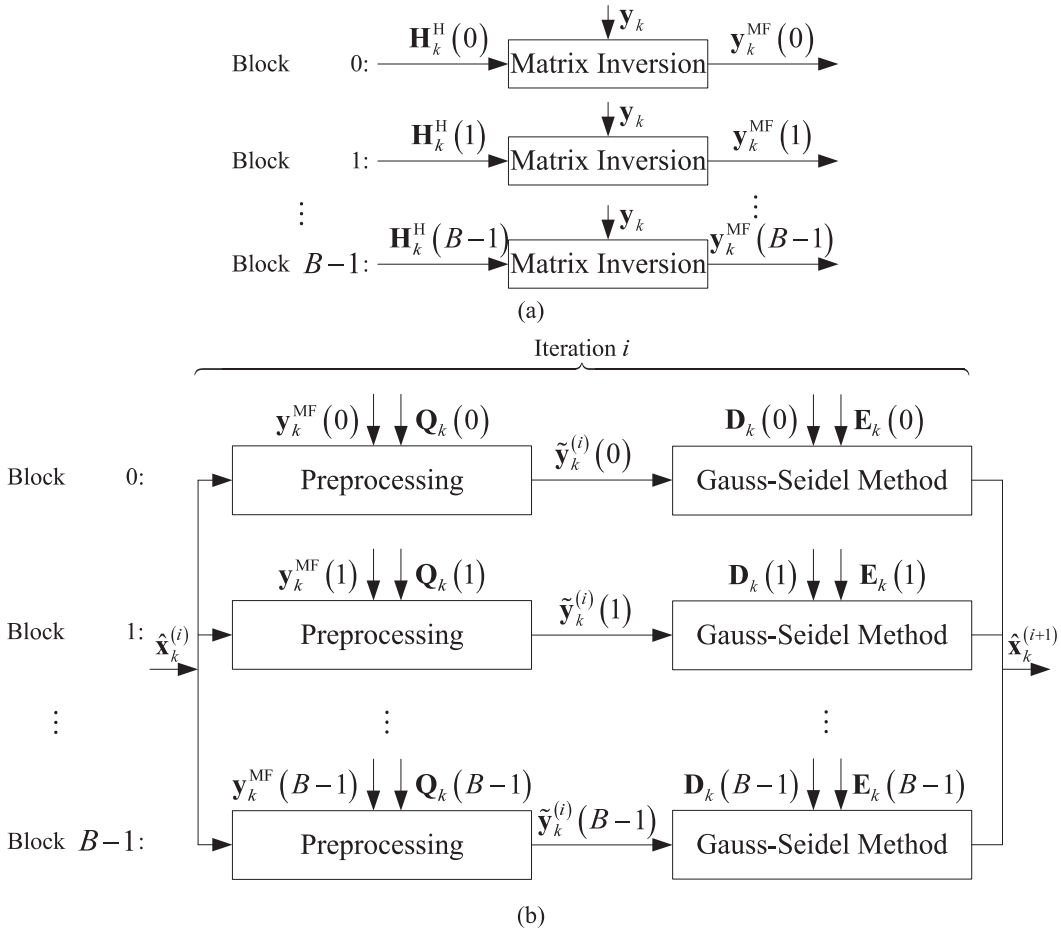


Figure 7.2: Block diagram of the parallel processing

clear from Figure 7.2 that after obtaining $\mathbf{y}_k^{\mathrm{MF}}(l)$, the matched filter output is fed into the corresponding block for preprocessing. The updating processing using Gauss-Seidel method is performed on each block independently.

By using the parallel processing structure, for each iteration, the overall processing time will be only $1/B$ of the original Gauss-Seidel method due to a small size of block matrix. Besides, the parallel processing structure is efficient for hardware implementation, which is much more appealing in practical applications.

7.2.3 Initial Estimation

Intuitively, if the initial estimation is very close to the linear MMSE estimation, the iterations required for convergence for the proposed iterative method will be few [85]. By noting the matrix $\mathbf{W}_k$ is diagonal dominant, we use the initial estimate for the iterative method, given by

$$\mathbf{x}_k^{(0)}(l) = \mathbf{D}_k^{-1}(l)\,\mathbf{y}_k^{\mathrm{MF}}(l). \tag{7.13}$$

It is worth noting that the initial estimate provided by Equation (7.13) is of low complexity. This is because $\mathbf{D}_k$ is a diagonal matrix, and only $O(N_L)$ multiplications are required for each block.

7.2.4 LLR Computation

As the soft-input channel code schemes are usually adopted in the UWA communication system, we need to derive the soft-output of the detector. For linear MMSE detection in Equation (7.3), we separate the estimation of the symbol from the u-th AUV or UWA sensor, given by

$$\hat{x}_{k,u} = \mathbf{e}_u^{\mathrm{H}}\hat{\mathbf{x}}_k = \rho_{k,u}x_{k,u} + \zeta_{k,u}, \tag{7.14}$$

where $\rho_{k,u}$ is the equivalent channel gain, and $\zeta_{k,u}$ denotes the noise plus interference (NPI). $\rho_{k,u}$ and the covariance of NPI are respectively given by

$$\rho_{k,u} = \mathbf{e}_u^{\mathrm{H}} \mathbf{W}_k^{-1} \mathbf{H}_k^{\mathrm{H}} \mathbf{H}_k \mathbf{e}_u, \tag{7.15}$$

$$v_{k,u} = \mathrm{cov}\left\{\zeta_{k,u}, \zeta_{k,u}\right\} = \rho_{u,k}\left(1 - \rho_{u,k}\right). \tag{7.16}$$

Therefore, the max log likelihood ratio (LLR) is given by

$$L\left(b_{k,u}^m\right) = \frac{\rho_{u,k}}{1-\rho_{u,k}} \left(\min_{s \in \chi_m^0} \left| \frac{\hat{x}_{k,u}}{\rho_{u,k}} - s \right|^2 - \min_{s' \in \chi_m^1} \left| \frac{\hat{x}_{k,u}}{\rho_{u,k}} - s' \right|^2 \right), \tag{7.17}$$

where $b_{k,u}^m$ is the m-th mapping bit of $x_{k,u}$. χ_m^0 and χ_m^1 denote the subset of the constellation symbols where the m-th mapping bit is 0 or 1, respectively.

If the convergence conditions for the block Gauss-Seidel method are satisfied[2], with sufficiently large iterations, the proposed Gauss-Seidel Method will approach the linear MMSE estimation vector. Therefore, Equation (7.17) can be applied to derive the LLRs. However, we note that the computation of $\rho_{u,k}$ in Equation (7.15) involves the matrix inversion, which is not expected for low-complexity consideration. To address this issue, we rewrite Equation (7.15) into

$$\rho_{k,u} = \mathbf{e}_u^{\mathrm{H}} \mathbf{W}_k^{-1} \left(\mathbf{W}_k - \sigma_z^2 \mathbf{I}_{N_V} \right) \mathbf{e}_u = 1 - \sigma_z^2 \mathbf{e}_u^{\mathrm{H}} \mathbf{W}_k^{-1} \mathbf{e}_u. \tag{7.18}$$

That is to say, we only need to know the diagonal elements of $\mathbf{W}_k^{-1}$ for the computation of $\rho_{k,u}$. In addition, as we know that $\mathbf{W}_k$ is diagonally dominant, the approximation of $\rho_{k,u}$ can be given as

$$\rho_{k,u} \approx 1 - \sigma_z^2 \mathbf{e}_u^{\mathrm{H}} \mathbf{D}_k^{-1} \mathbf{e}_u. \tag{7.19}$$

The approximation in Equation (7.19) has significant performance improvement than the LLR computation method in [84, 85].

[2]The iterative method will converge if $\lim_{l \to \infty} \left(\mathbf{I}_{N_V} - \mathbf{P}_k^{-1} \mathbf{W}_k \right)^l = \mathbf{0}$, or equivalently, $\left| \lambda \left(\mathbf{I}_{N_V} - \mathbf{P}_k^{-1} \mathbf{W}_k \right) \right|_{\max} < 1$.

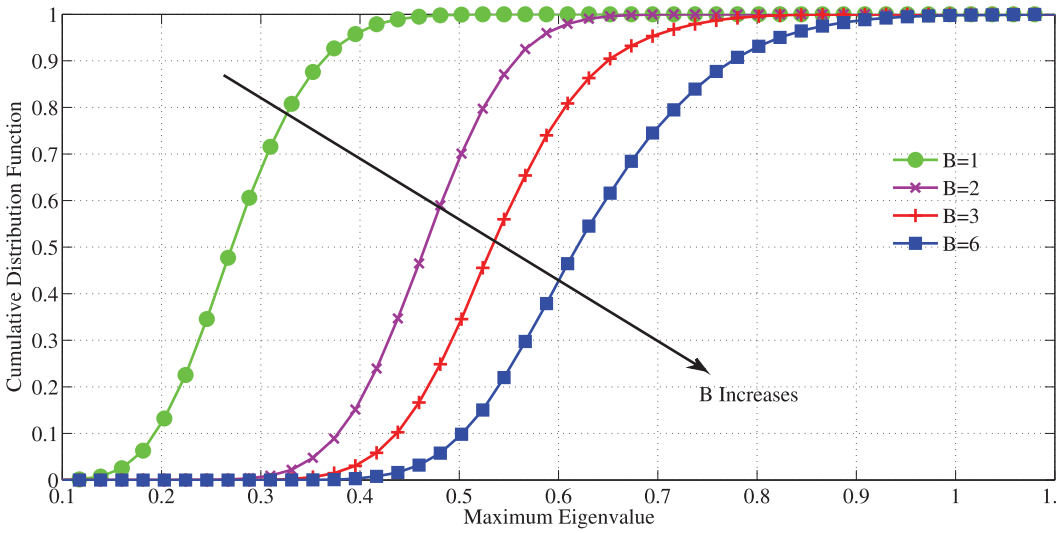


Figure 7.3: Cumulative distribution function of the maximum eigenvalue

To summarize, we present the soft-output block Gauss-Seidel method in **Algorithm 7.1**.

7.3 Performance Evaluation and Discussion

7.3.1 Computational Complexity

From **Algorithm** 1, we can see the main processing delay is determined by the iterative progress. By using the block diagonal matrix in our proposed scheme, the iterative progress can be implemented on each block independently. For each iteration, the computational complexity in each block is in the order of $O\left(N_V N_L\right)$. Compared to the processing in (7.5), where the overall complexity is $O\left(N_V^2\right)$, our proposed scheme is of low complexity in each block. Besides, we propose to use the parallel processing structure in each iterative progress, the overall processing delay in one iteration is just $1/B$ of the original Gauss-Seidel method.

Table 7.1: **Algorithm 7.1:** Proposed Block Gauss-Seidel Method on Each Subcarrier

Input: $\mathbf{H}_k$, $\mathbf{y}_k$, σ_z^2, number of blocks B, and iterations I

Output: LLRs of the bits associated with transmitted symbols.

Initialization:

1. $\mathbf{W}_k = \mathbf{H}_k^{\mathrm{H}}\mathbf{H}_k + \sigma_z^2\mathbf{I}_{N_U}$, $\mathbf{y}_k^{\mathrm{MF}} = \mathbf{H}_k\mathbf{y}_k$;
2. Obtain $\mathbf{D}_k(l)$, $\mathbf{Q}_k(l)$, $\mathbf{E}_k(l)$, and $\mathbf{y}_k^{\mathrm{MF}}(l)$, $l = 0, 1, \cdots, B$;

Parallel Processing: Initial Estimation (3-5)

3. for $l = 0 : 1 : (B-1)$
4. $\quad \mathbf{x}_k^{(0)}(l) = \mathbf{D}_k^{-1}(l)\,\mathbf{y}_k^{\mathrm{MF}}(l)$;
5. end

Iteration:

6. for $i = 0 : 1 : (I-1)$

Parallel Processing: Iterative Progress (7-10)

7. $\quad$ for $l = 0 : 1 : (B-1)$
8. $\quad\quad$ Preprocessing: $\tilde{\mathbf{y}}_k^{(i)}(l) = \mathbf{y}_k^{\mathrm{MF}}(l) - \mathbf{Q}_k(l)\,\hat{\mathbf{x}}_k^{(i)}$;
9. $\quad\quad$ Updating: $\hat{\mathbf{x}}_k^{(i+1)}(l) = (\mathbf{D}_k(l) + \mathbf{E}_k(l))^{-1}$
$$\left(\tilde{\mathbf{y}}_k^{(i)}(l) - \mathbf{E}_k^{\mathrm{H}}(l)\,\hat{\mathbf{x}}_k^{(i)}(l)\right);$$
10. $\quad$ end
11. end

LLR Computation

12. Compute $\rho_{u,k}$;
13. Compute $L\left(b_{k,u}^m\right)$ with Equation (7.17).

Return $L\left(b_{k,u}^m\right)$.

7.3.2 Convergence Performance

We now evaluate the convergence performance of the proposed block Gauss-Seidel method. We randomly generate the channel matrix with $N_R = 30$, $N_V = 6$, and the average signal to noise ratio (SNR) at receiver is set to 6.5dB. We then use Monte-Carlo simulations to obtain the cumulative distribution function of the maximum eigenvalue for the convergence matrix $\mathbf{I}_{N_V} - \mathbf{P}_k^{-1}\mathbf{W}_k$. The results are shown in Figure 7.3. Note in Figure 7.3, $B = 1$ corresponds to the original Gauss-Seidel method, while $B = 6$ represents the Jacobi method (or equivalently the Neumann series expansion based detection method in [23]). Apparently, with the increase of the block size, the probability that $\Pr\left(\left|\lambda\left(\mathbf{I}_{N_V} - \mathbf{P}_k^{-1}\mathbf{W}_k\right)\right|_{\max} < 1\right)$ will decrease, which indicates that Jacobi method (or Neumann series expansion based detection scheme) may be inapplicable as the convergence condition is not satisfied. Besides, the results in Figure 7.3 also reveal reveal that $\Pr\left(\left|\lambda\left(\mathbf{I}_{N_V} - \mathbf{P}_k^{-1}\mathbf{W}_k\right)\right|_{\max} < a\right)$, $0 < a < 1$, for small size B is always greater than that for large size B. This indicates that with small B, the convergence rate will be faster than that with a large B. However, small size B indicates large processing delay in processing. Therefore, a proper B is required to achieve a good trade-off between the convergence rate and efficiency in parallel processing.

7.3.3 BER Performance

The acoustic channel simulator in [14] is adopted to generate channel data for BER performance evaluation. The OFDM parameters are given in [93]. Each AUV and UWA sensor adopts an LDPC code with code length 1152, code rate 0.5 and 16QAM modulation. 30 hydrophones are deployed at the platform and 6 AUVs are served simultaneously[3]. As we can see from Fig. 7.4, the original Gauss-Seidel method requires 2-3 iterations to converge.

[3]With large array of hydrophones, the number of AUVs in service can be much more, and the performance improvement of the proposed scheme can be much more obvious.

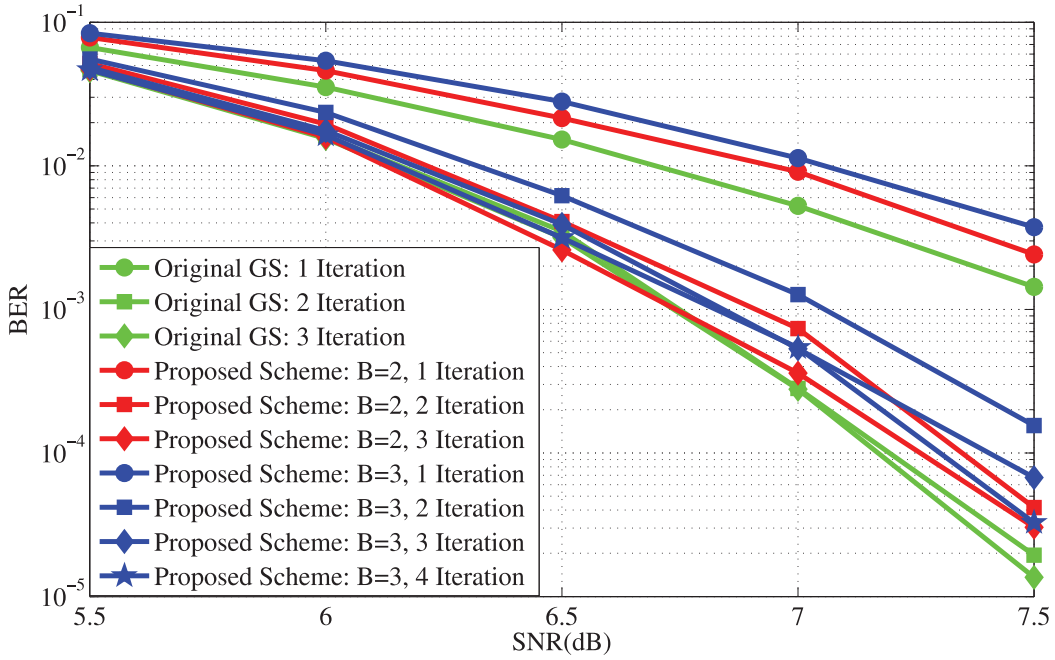


Figure 7.4: BER Performance, LDPC code rate 0.5, code length 1152, 16QAM

However, the iterations are about 3 and 4 for $B = 2$ and $B = 3$, respectively. Considering that the processing time for each block is just $1/B$ of the original Gauss-Seidel method, the overall processing delay for both cases is much less than the original Gauss-Seidel method.

7.4 Summary

We propose the block Gauss-Seidel method for signal detection in UWA communications system. With a large array of hydrophones, the signal process is computational costly and requires long processing time; however, our proposed scheme can efficiently address these issues. By performing the Gauss-Seidel method on a small scale block matrix, and using a parallel processing structure, the proposed scheme achieves almost the same system performance compared to the original Gauss-Seidel method, but with much less processing delay.

Chapter 8

Conclusions and Future Work

8.1 Conclusions

We have proposed various low-complexity and efficient data detection schemes for conventional small-scale and large-scale MIMO systems. Generally, we propose turbo equalization schemes for conventional small-scale MIMO system, and near linear MMSE detection schemes for massive MIMO systems. The main conclusions are listed as follows:

- The proposed soft input soft output MMSE-SQRD based turbo equalization for MIMO-OFDM systems under imperfect channel estimation has shown enhanced system BER performance compared to conventional turbo equalization schemes;
- The proposed turbo equalizer conditioned on estimated channel for MIMO MMSE receiver can efficiently utilize the *a priori* information from channel decoder to account for a generalized expression of existing MMSE and MMSE-SQRD based equalizers;
- The proposed SQRD-based soft interference cancelation scheme performs interference cancelation for both detected and undetected users, and the system BER performance is further improved;

- The proposed data detection scheme using the stair matrix has shown much improved system performance over the use of the diagonal matrix, in terms of the the convergence performance, mean-square error to approach the matrix inversion, residual estimation error to approach the linear MMSE estimation vector, and the system BER performance;

- The proposed block diagonal matrix based massive MIMO uplink detection scheme and the block Gauss-Seidel method for V2I communications have shown much reduced processing delay and achieved nearly the performance of the linear MMSE detection scheme;

- The block Gauss-Seidel method have shown reduced computational complexity and processing delay to achieve BER performance close to original Gauss-Seidel method in the applications to large-scale array signal detection in UWA communications.

8.2 Future Work

Based on the research work and promising results from this project, some of the future work can be conducted as follows.

1. The analysis of data detection schemes for massive MIMO under imperfect channel estimation

   Our results have shown that the system performance degradation is inevitable when channel estimation error is presented. However, by taking into account the channel estimation error in data detection schemes, we can achieve significant performance gain. These results are good references for the design of data detection schemes for massive MIMO with channel estimation error.

2. High parallelism and fast convergence detection schemes for massive MIMO

   We have shown that the detection scheme using the stair matrix in the development can improve the system convergence performance, while the block diagonal matrix based data detection and block Gauss-Seidel method can greatly reduce the overall processing delay. Inspired by these results, we can further investigate the high parallelism and fast convergence data detection scheme for massive MIMO.

3. Incorporating compressed sensing based detection schemes in massive MIMO

   We have shown that when the ratio $r = N_B/N_U$ is not sufficiently large, the linear MMSE detection scheme cannot achieve near optimal system performance. On the other hand, using compressed sensing based detection schemes, we can greatly improve the system performance, much better than that of the linear MMSE detection scheme. Therefore, incorporating compressed sensing based data detection in massive MIMO can be a promising research topic in future.

Appendix

A Preliminaries

We first present the preliminary lemmas.

Lemma A.1. *Let $a_k \sim CN(0,1)$, we then have*

$$E\left\{|a_k|^2\right\} = 1, \tag{A.1}$$

$$E\left\{|a_k|^4\right\} = 2, \tag{A.2}$$

$$E\left\{|a_k|^6\right\} = 6, \tag{A.3}$$

$$E\left\{|a_k|^8\right\} = 24, \tag{A.4}$$

Proof. We first obtain the joint probability density function (PDF) $f(a_k) = \frac{1}{\pi}\exp\left(-|a_k|^2\right)$, and then apply $\mathrm{E}\{g(a_k)\} = \int_{\mathbb{C}} g(a_k)\cdot f(a_k)\,da_k$ to obtain the results in Equations (A.1) - (A.4). □

Lemma A.2. *Let $\mathbf{a} = [a_1, a_2, \cdots, a_{N_B}]^T$ with each entry $a_k \sim CN(0,1)$, independent and identically distributed (i.i.d.). We then have*

$$E\left\{\mathbf{a}^H\mathbf{a}\right\} = N_B, \tag{A.5}$$

$$E\left\{\left|\mathbf{a}^H\mathbf{a}\right|^4\right\} = A_3, \tag{A.6}$$

$$E\left\{\left|\mathbf{a}^{H}\mathbf{a}\right|^{-4}\right\} = \frac{1}{B_1}, \tag{A.7}$$

where

$$\begin{aligned} A_3 = {} & 24N_B + N_B\left(N_B-1\right)\left(N_B-2\right)\left(N_B-3\right) \\ & + 36N_B\left(N_B-1\right) + 12N_B\left(N_B-1\right)\left(N_B-2\right), \end{aligned} \tag{A.8}$$

$$B_1 = \left(N_B-1\right)\left(N_B-2\right)\left(N_B-3\right)\left(N_B-4\right). \tag{A.9}$$

Proof. Through $\mathrm{E}\left\{\mathbf{a}^{\mathrm{H}}\mathbf{a}\right\} = \mathrm{E}\left\{\sum\limits_{k=1}^{N_B}\left|a_k\right|^2\right\} = \sum\limits_{k=1}^{N_B}\mathrm{E}\left\{\left|a_k\right|^2\right\}$ and the results from Equation (A.1), we can derive the results in Equation (A.5).

We first write $\left|\mathbf{a}^{\mathrm{H}}\mathbf{a}\right|^4 = \sum\limits_{k=1}^{N_B}\sum\limits_{l=1}^{N_B}\sum\limits_{m=1}^{N_B}\sum\limits_{n=1}^{N_B}\left|a_k\right|^2\left|a_l\right|^2\left|a_m\right|^2\left|a_n\right|^2$. Therefore, we have $\mathrm{E}\left\{\left|\mathbf{a}^{\mathrm{H}}\mathbf{a}\right|^4\right\}$ given by

$$\begin{aligned} \mathrm{E}\left\{\left|\mathbf{a}^{\mathrm{H}}\mathbf{a}\right|^4\right\} = & \mathrm{E}\left\{\sum_{k=1}^{N_B}\sum_{l=1}^{N_B}\sum_{m=1}^{N_B}\sum_{n=1}^{N_B}\left|a_k\right|^2\left|a_l\right|^2\left|a_m\right|^2\left|a_n\right|^2\right\} \\ = & N_B\mathrm{E}\left\{\left|a_k\right|^8\right\} + 4N_B\left(N_B-1\right)\mathrm{E}\left\{\left|a_k\right|^6\right\}\mathrm{E}\left\{\left|a_l\right|^2\right\} \\ & + 6N_B\left(N_B-1\right)\left(N_B-2\right)\mathrm{E}\left\{\left|a_k\right|^4\right\}\mathrm{E}\left\{\left|a_l\right|^2\right\}\mathrm{E}\left\{\left|a_m\right|^2\right\} \\ & + 3N_B\left(N_B-1\right)\mathrm{E}\left\{\left|a_k\right|^4\right\}\mathrm{E}\left\{\left|a_l\right|^4\right\} \\ & + N_B\left(N_B-1\right)\left(N_B-2\right)\left(N_B-3\right)\mathrm{E}\left\{\left|a_k\right|^2\right\}\mathrm{E}\left\{\left|a_l\right|^2\right\}\mathrm{E}\left\{\left|a_m\right|^2\right\}\mathrm{E}\left\{\left|a_n\right|^2\right\} \end{aligned} \tag{A.10}$$

The results can be derived as follows. Since the elements in $\mathbf{a}$ are i.i.d., the non-zero terms come from five cases:

Case1: $\mathrm{E}\left\{\left|a_k\right|^8\right\}$, which corresponds to $k = l = m = n$;

Case2: $\mathrm{E}\left\{\left|a_k\right|^6\right\}\mathrm{E}\left\{\left|a_l\right|^2\right\}$, which corresponds to $k = m = n$ and $k \neq l$;

Case3: $\mathrm{E}\left\{\left|a_k\right|^4\right\}\mathrm{E}\left\{\left|a_l\right|^2\right\}\mathrm{E}\left\{\left|a_m\right|^2\right\}$, which corresponds to $k = n, k \neq l \neq m$;

Case4: $\mathrm{E}\left\{\left|a_k\right|^4\right\}\mathrm{E}\left\{\left|a_l\right|^4\right\}$, which corresponds to $k = m, l = n, k \neq l$;

Case5: $\mathrm{E}\left\{\left|a_k\right|^2\right\}\mathrm{E}\left\{\left|a_l\right|^2\right\}\mathrm{E}\left\{\left|a_m\right|^2\right\}\mathrm{E}\left\{\left|a_n\right|^2\right\}$, which corresponds to $k \neq l \neq m \neq n$.

After expansion, we can see there are:

- N_B terms in **Case1**;
- $\begin{pmatrix} 4 \\ 3 \end{pmatrix} N_B \left(N_B - 1\right)$ terms in **Case2**;
- $\begin{pmatrix} 4 \\ 2 \end{pmatrix} N_B \left(N_B - 1\right)\left(N_B - 2\right)$ terms in **Case3**;
- $\frac{1}{2}\begin{pmatrix} 4 \\ 2 \end{pmatrix} N_B \left(N_B - 1\right)$ terms in **Case4**;
- $N_B \left(N_B - 1\right)\left(N_B - 2\right)\left(N_B - 3\right)$ terms in **Case5**.

Using the results in Equations (A.1) - (A.4), we derive the result in Equation (A.6).

Equation (A.7) is obtained by noting that $\left|\mathbf{a}^{\mathrm{H}}\mathbf{a}\right|^{-1}$ follows an inverse-Gamma distribution, and the 4-th moment is obtained in [100]. □

Lemma A.3. *Let* $\mathbf{a} = \left[a_1, a_2, \cdots, a_{N_B}\right]^T$, $\mathbf{b} = \left[b_1, b_2, \cdots, b_{N_B}\right]^T$, *with each entry* $a_k \sim CN\left(0,1\right)$, $b_k \sim CN\left(0,1\right)$, *and i.i.d., we then have*

$$E\left\{\left|\mathbf{a}^H\mathbf{b}\right|^4\right\} = A_1, \tag{A.11}$$

$$E\left\{\left|\mathbf{a}^H\mathbf{b}\right|^8\right\} = A_5. \tag{A.12}$$

where

$$A_1 = 2N_B\left(N_B + 1\right), \tag{A.13}$$

$$\begin{aligned} A_5 = {} & 576N_B + 24N_B\left(N_B - 1\right)\left(N_B - 2\right)\left(N_B - 3\right) \\ & + 864N_B\left(N_B - 1\right) + 288N_B\left(N_B - 1\right)\left(N_B - 2\right). \end{aligned} \tag{A.14}$$

Proof. We first write

$$\left|\mathbf{a}^{\mathrm{H}}\mathbf{b}\right|^2 = \left(\sum_{k=1}^{N_B} a_k^* b_k\right)\left(\sum_{l=1}^{N_B} a_l b_l^*\right) = \sum_{k=1}^{N_B}\sum_{l=1}^{N_B} a_k^* a_l b_l^* b_k.$$

In addition, we have

$$\begin{aligned}
\mathrm{E}\left\{\left|\mathbf{a}^{\mathrm{H}}\mathbf{b}\right|^4\right\} &= \mathrm{E}\left\{\sum_{k=1}^{N_B}\sum_{l=1}^{N_B}\sum_{m=1}^{N_B}\sum_{n=1}^{N_B} a_k^* a_l a_m^* a_n b_l^* b_k b_n^* b_m\right\} \\
&= \sum_{k=1}^{N_B}\sum_{l=1}^{N_B}\sum_{m=1}^{N_B}\sum_{n=1}^{N_B} \mathrm{E}\left\{a_k^* a_l a_m^* a_n b_l^* b_k b_n^* b_m\right\} \qquad \text{(A.15)} \\
&= 2N_B\left(N_B-1\right)\mathrm{E}\left\{|a_k|^2\right\}^2\mathrm{E}\left\{|b_k|^2\right\}^2 + N_B\mathrm{E}\left\{|a_k|^4\right\}\mathrm{E}\left\{|b_k|^4\right\}.
\end{aligned}$$

In (A.15), the i.i.d. assumption is used for the derivation and only non-zero terms are considered. With the results in Equations (A.1) and (A.2), we derive the result in Equation (A.11). Moreover, we write $\mathrm{E}\left\{\left|\mathbf{a}^{\mathrm{H}}\mathbf{b}\right|^8\right\}$ as

$$\begin{aligned}
\mathrm{E}\left\{\left|\mathbf{a}^{\mathrm{H}}\mathbf{b}\right|^8\right\} &= \sum_{k=1}^{N_B}\sum_{k_1=1}^{N_B}\sum_{l=1}^{N_B}\sum_{l_1=1}^{N_B}\sum_{m=1}^{N_B}\sum_{m_1=1}^{N_B}\sum_{n=1}^{N_B}\sum_{n_1=1}^{N_B} \mathrm{E}\left\{a_k^* a_{k_1}^* a_l a_{l_1} a_m^* a_{m_1}^* a_n a_{n_1}\right\}\mathrm{E}\left\{b_k b_{k_1} b_l^* b_{l_1}^* b_m b_{m_1} b_n^* b_{n_1}^*\right\} \\
&= \alpha_1\left(\mathrm{E}\left\{|a_k|^2\right\}\right)^8 + \alpha_2\left(\mathrm{E}\left\{|a_k|^2\right\}\right)^4\left(\mathrm{E}\left\{|a_l|^4\right\}\right)^2 + \alpha_3\left(\mathrm{E}\left\{|a_k|^2\right\}\right)^2\left(\mathrm{E}\left\{|a_l|^6\right\}\right)^2 \\
&\quad + \alpha_4\left(\mathrm{E}\left\{|a_k|^4\right\}\right)^4 + \alpha_5\mathrm{E}\left\{|a_k|^8\right\}\mathrm{E}\left\{|a_l|^8\right\},
\end{aligned} \qquad \text{(A.16)}$$

where only the non-zero terms are considered, and the i.i.d. assumption is used for the derivation. The coefficients α_i, $i = 1, 2, \cdots, 5$, account for the number of the non-zero terms, and are given by

$$\begin{aligned}
\alpha_1 &= \begin{pmatrix} 4 \\ 1 \end{pmatrix}\cdot\begin{pmatrix} 3 \\ 1 \end{pmatrix}\cdot\begin{pmatrix} 2 \\ 1 \end{pmatrix} N_B\left(N_B-1\right)\left(N_B-2\right)\left(N_B-3\right) \\
&= 24N_B\left(N_B-1\right)\left(N_B-2\right), \\
\alpha_2 &= \begin{pmatrix} 4 \\ 2 \end{pmatrix}\cdot\begin{pmatrix} 4 \\ 2 \end{pmatrix}\cdot 2N_B\left(N_B-1\right)\left(N_B-2\right) \\
&= 72N_B\left(N_B-1\right)\left(N_B-2\right),
\end{aligned}$$

$$\alpha_3 = \begin{pmatrix} 4 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 4 \\ 1 \end{pmatrix} N_B \left(N_B - 1\right) = 16 N_B \left(N_B - 1\right),$$

$$\alpha_4 = \frac{1}{2} \begin{pmatrix} 4 \\ 2 \end{pmatrix} \cdot \begin{pmatrix} 4 \\ 2 \end{pmatrix} N_B \left(N_B - 1\right) = 18 N_B \left(N_B - 1\right),$$

$$\alpha_5 = N_B.$$

With the results in Equations (A.1) - (A.4), we have the result in Equation (A.12). □

Lemma A.4. *Let* $A = \mathbf{a}^H \mathbf{b}\mathbf{b}^H \mathbf{c}$, *where* $\mathbf{a} = \left[a_1, a_2, \cdots, a_{N_B}\right]^T$, $\mathbf{b} = \left[b_1, b_2, \cdots, b_{N_B}\right]^T$, *and* $\mathbf{c} = \left[c_1, c_2, \cdots, c_{N_B}\right]^T$, *with each entry* $a_k \sim CN\left(0,1\right)$, $b_k \sim CN\left(0,1\right)$, *and* $c_k \sim CN\left(0,1\right)$, *and i.i.d., we then have*

$$E\left\{|A|^4\right\} = A_2, \tag{A.17}$$

where

$$\begin{aligned} A_2 = {} & 96 N_B + 4 N_B \left(N_B - 1\right)\left(N_B - 2\right)\left(N_B - 3\right) \\ & + 144 N_B \left(N_B - 1\right) + 48 N_B \left(N_B - 1\right)\left(N_B - 2\right), \end{aligned} \tag{A.18}$$

Proof. We first write

$$A = \mathbf{a}^{\mathrm{H}} \mathbf{b}\mathbf{b}^{\mathrm{H}} \mathbf{c} = \sum_{k=1}^{N_B} a_k^* b_k \sum_{l=1}^{N_B} b_l^* c_l = \sum_{k=1}^{N_B} \sum_{l=1}^{N_B} a_k^* b_k b_l^* c_l,$$

and then we have

$$|A|^2 = \sum_{k=1}^{N_B} \sum_{l=1}^{N_B} \sum_{m=1}^{N_B} \sum_{n=1}^{N_B} a_k^* a_m b_k b_l^* b_m^* b_n c_l c_n^*.$$

Therefore, we have $\mathrm{E}\left\{|A|^4\right\} = \mathrm{E}\left\{|A|^2 \cdot |A|^2\right\}$ given by

$$\begin{aligned} \mathrm{E}\left\{|A|^4\right\} &= \sum_{k=1}^{N_B} \sum_{l=1}^{N_B} \sum_{m=1}^{N_B} \sum_{n=1}^{N_B} \sum_{o=1}^{N_B} \sum_{p=1}^{N_B} \sum_{q=1}^{N_B} \sum_{r=1}^{N_B} \mathrm{E}\left\{a_k^* a_m a_o^* a_q b_k b_l^* b_m^* b_n b_o b_p^* b_q^* b_r c_l c_n^* c_p c_r^*\right\} \\ &= \alpha_1 \left(\mathrm{E}\left\{|a_k|^4\right\}\right)^2 \mathrm{E}\left\{|a_l|^8\right\} + \alpha_2 \left(\mathrm{E}\left\{|a_k|^4\right\}\right)^4 + \alpha_3 \left(\mathrm{E}\left\{|a_k|^2\right\}\right)^4 \left(\mathrm{E}\left\{|a_l|^4\right\}\right)^2 \\ &\quad + \alpha_4 \left(\mathrm{E}\left\{|a_k|^2\right\}\right)^3 \mathrm{E}\left\{|a_l|^4\right\} \mathrm{E}\left\{|a_m|^6\right\} + \alpha_5 \left(\mathrm{E}\left\{|a_k|^2\right\}\right)^8 + \alpha_6 \left(\mathrm{E}\left\{|a_k|^2\right\}\right)^6 \mathrm{E}\left\{|a_l|^4\right\}, \end{aligned} \tag{A.19}$$

where

$$\alpha_1 = N_B,$$

$$\alpha_2 = N_B\left(N_B - 1\right),$$

$$\alpha_3 = 4N_B\left(N_B - 1\right)\left(N_B - 2\right) + 8N_B\left(N_B - 1\right),$$

$$\alpha_4 = 8N_B\left(N_B - 1\right),$$

$$\alpha_5 = 4N_B\left(N_B - 1\right)\left(N_B - 2\right)\left(N_B - 3\right),$$

$$\alpha_6 = 16N_B\left(N_B - 1\right)\left(N_B - 2\right).$$

Table A.1: Non-zero Terms in Equation (A.19)

Non-zero Terms	**Cases**
$\left(\mathrm{E}\left\{\left\|a_k\right\|^4\right\}\right)^2\mathrm{E}\left\{\left\|a_l\right\|^8\right\}$	$k = l = m = n = o = p = q = r$: N_B
$\left(\mathrm{E}\left\{\left\|a_k\right\|^4\right\}\right)^4$	$k = m = o = q, l = n = p = r, k \neq l$: $N_B\left(N_B - 1\right)$
$\left(\mathrm{E}\left\{\left\|a_k\right\|^2\right\}\right)^4\left(\mathrm{E}\left\{\left\|a_l\right\|^4\right\}\right)^2$	$k = m = o = q, l = n, p = r, k \neq l \neq p$: $2 \cdot 2N_B\left(N_B - 1\right)\left(N_B - 2\right)$ or $k = m = l = n, o = q = p = r, k \neq o$: $4 \cdot 2N_B\left(N_B - 1\right)$
$\left(\mathrm{E}\left\{\left\|a_k\right\|^2\right\}\right)^3\mathrm{E}\left\{\left\|a_l\right\|^4\right\}\mathrm{E}\left\{\left\|a_m\right\|^6\right\}$	$k = m = o = q = l = n, p = r, k \neq p$: $4 \cdot 2N_B\left(N_B - 1\right)$
$\left(\mathrm{E}\left\{\left\|a_k\right\|^2\right\}\right)^8$	$k = m, o = q, l = n, p = r, k \neq o \neq l \neq p$: $4N_B\left(N_B - 1\right)\left(N_B - 2\right)\left(N_B - 3\right)$
$\left(\mathrm{E}\left\{\left\|a_k\right\|^2\right\}\right)^6\mathrm{E}\left\{\left\|a_l\right\|^4\right\}$	$k = m, o = q = l = n, p = r, k \neq o \neq p$: $16N_B\left(N_B - 1\right)\left(N_B - 2\right)$

In Equation (A.19), only the non-zero terms are considered in the expectation and the i.i.d. assumption is used for the derivation. We summarize the non-zero terms in Table A.1. It worth noting that we only provide an example to show how the non-zero term is generated, and the coefficients correspond to similar cases that a non-zero term is generated. For example, for the non-zero term $\left(\mathrm{E}\left\{|a_k|^2\right\}\right)^8$, except for the cases in Table A.1, we also have the following cases to generate the non-zero term.

$$k = q, m = o, l = n, p = r, k \neq o \neq l \neq p;$$

$$k = m, o = q, l = r, p = n, k \neq o \neq l \neq p;$$

$$k = q, m = o, l = r, p = n, k \neq o \neq l \neq p.$$

That is why the coefficient 4 is present. The same for other non-zero terms in TABLE A.1.

With the results in Equations (A.1) - (A.4), we derive the result in Equation (A.17). □

Lemma A.5. *Let* $A = \mathbf{a}^H\mathbf{a}\mathbf{b}^H\mathbf{b}\mathbf{c}^H\mathbf{b}\mathbf{b}^H\mathbf{d}\mathbf{a}^H\mathbf{c}\mathbf{d}^H\mathbf{a}$, *where* $\mathbf{a} = [a_1, a_2, \cdots, a_{N_B}]^T$, $\mathbf{b} = [b_1, b_2, \cdots, b_{N_B}]^T$, $\mathbf{c} = [c_1, c_2, \cdots, c_{N_B}]^T$, *and* $\mathbf{d} = [d_1, d_2, \cdots, d_{N_B}]^T$, *with each entry* $a_k \sim CN(0,1)$, $b_k \sim CN(0,1)$, $c_k \sim CN(0,1)$, *and* $d_k \sim CN(0,1)$, *and i.i.d., we then have*

$$E\left\{A^2\right\} = A_4, \tag{A.20}$$

where

$$
\begin{aligned}
A_4 =& N_B\left(N_B-1\right)\left(N_B-2\right)^3\left(N_B-3\right)^3+26N_B\left(N_B-1\right)\left(N_B-2\right)^3\left(N_B-3\right)^2\\
&+46N_B\left(N_B-1\right)\left(N_B-2\right)^2\left(N_B-3\right)^2+4N_B\left(N_B-1\right)^2\left(N_B-2\right)^3\left(N_B-3\right)\\
&+220N_B\left(N_B-1\right)\left(N_B-2\right)^3\left(N_B-3\right)+48N_B\left(N_B-1\right)^2\left(N_B-2\right)^2\left(N_B-3\right)\\
&+808N_B\left(N_B-1\right)\left(N_B-2\right)^2\left(N_B-3\right)+128N_B\left(N-1\right)^2\left(N_B-2\right)\left(N_B-3\right)\\
&+832N_B\left(N_B-1\right)\left(N_B-2\right)\left(N_B-3\right)+40N_B\left(N_B-1\right)^2\left(N_B-2\right)^3\\
&+600N_B\left(N_B-1\right)\left(N_B-2\right)^3+4N_B\left(N_B-1\right)^3\left(N_B-2\right)^2+576N_B\left(N_B-1\right)^2\left(N_B-2\right)^2\\
&+3480N_B\left(N_B-1\right)\left(N_B-2\right)^2+64N_B\left(N_B-1\right)^3\left(N_B-2\right)\\
&+2592N_B\left(N_B-1\right)^2\left(N_B-2\right)+8064N_B\left(N_B-1\right)\left(N_B-2\right)\\
&+256N_B\left(N_B-1\right)^3+4352N_B\left(N_B-1\right)^2+9888N_B\left(N_B-1\right)+2304N_B.
\end{aligned}
\tag{A.21}
$$

Table A.2: Non-zero Terms in Equation (A.22)

Non-zero Terms	**Number of Terms**
$\left(\mathrm{E}\left\{\left\|a_k\right\|^2\right\}\right)^{12}$	$2N_B\left(N_B-1\right)\left(N_B-2\right)^2\left(N_B-3\right)^2$
	$+N_B\left(N_B-1\right)\left(N_B-2\right)^3\left(N_B-3\right)^3$
	$+2N_B\left(N_B-1\right)\left(N_B-2\right)^3\left(N_B-3\right)^2$
$\left(\mathrm{E}\left\{\left\|a_k\right\|^2\right\}\right)^{10}\mathrm{E}\left\{\left\|a_l\right\|^4\right\}$	$12N_B\left(N_B-1\right)\left(N_B-2\right)^3\left(N_B-3\right)^2$
	$+20N_B\left(N_B-1\right)\left(N_B-2\right)^2\left(N_B-3\right)$
	$+2N_B\left(N_B-1\right)^2\left(N_B-2\right)^3\left(N_B-3\right)$
	$+20N_B\left(N_B-1\right)\left(N_B-2\right)^3\left(N_B-3\right)$

$\left(\mathrm{E}\left\{\lvert a_k\rvert^2\right\}\right)^9 \mathrm{E}\left\{\lvert a_l\rvert^6\right\}$	$8N_B\left(N_B-1\right)\left(N_B-2\right)\left(N_B-3\right)$
	$+4N_B\left(N_B-1\right)\left(N_B-2\right)^2\left(N_B-3\right)^2$
	$+4N_B\left(N_B-1\right)^2\left(N_B-2\right)^2\left(N_B-3\right)$
	$+8N_B\left(N_B-1\right)\left(N_B-2\right)^2\left(N_B-3\right)$
$\left(\mathrm{E}\left\{\lvert a_k\rvert^2\right\}\right)^8 \left(\mathrm{E}\left\{\lvert a_l\rvert^4\right\}\right)^2$	$5N_B\left(N_B-1\right)\left(N_B-2\right)^2\left(N_B-3\right)^2$
	$+6N_B\left(N_B-1\right)^2\left(N_B-2\right)^2\left(N_B-3\right)$
	$+45N_B\left(N_B-1\right)\left(N_B-2\right)^3\left(N_B-3\right)$
	$+8N_B\left(N_B-1\right)\left(N_B-2\right)\left(N_B-3\right)$
	$+50N_B\left(N_B-1\right)\left(N_B-2\right)^2$
	$+N_B\left(N_B-1\right)^3\left(N_B-2\right)^2$
	$+10N_B\left(N_B-1\right)^2\left(N_B-2\right)^3$
	$+8N_B\left(N_B-1\right)\left(N_B-2\right)^2\left(N_B-3\right)$
	$+50N_B\left(N_B-1\right)\left(N_B-2\right)^3$
$\left(\mathrm{E}\left\{\lvert a_k\rvert^2\right\}\right)^8 \mathrm{E}\left\{\lvert a_l\rvert^8\right\}$	$2N_B\left(N_B-1\right)\left(N_B-2\right)^2\left(N_B-3\right)$
$\left(\mathrm{E}\left\{\lvert a_k\rvert^2\right\}\right)^7 \mathrm{E}\left\{\lvert a_l\rvert^4\right\} \mathrm{E}\left\{\lvert a_m\rvert^6\right\}$	$8N_B\left(N_B-1\right)^2\left(N_B-2\right)\left(N_B-3\right)$
	$+28N_B\left(N_B-1\right)\left(N_B-2\right)^2\left(N_B-3\right)$
	$+40N_B\left(N_B-1\right)\left(N_B-2\right)$
	$+4N_B\left(N_B-1\right)^3\left(N_B-2\right)$
	$+24N_B\left(N_B-1\right)^2\left(N_B-2\right)^2$
	$+40N_B\left(N_B-1\right)\left(N_B-2\right)^2$

$\left(\mathrm{E}\left\{\lvert a_k\rvert^2\right\}\right)^6\left(\mathrm{E}\left\{\lvert a_l\rvert^4\right\}\right)^3$	$38N_B\left(N_B-1\right)\left(N_B-2\right)^2\left(N_B-3\right)$
	$+34N_B(N_B-1)^2(N_B-2)^2$
	$+4N_B(N_B-1)^2\left(N_B-2\right)\left(N_B-3\right)$
	$+50N_B\left(N_B-1\right)\left(N_B-2\right)^3$
	$+40N_B\left(N_B-1\right)\left(N_B-2\right)$
	$+2N_B(N_B-1)^3\left(N_B-2\right)$
	$+40N_B\left(N_B-1\right)\left(N_B-2\right)^2$
$\left(\mathrm{E}\left\{\lvert a_k\rvert^2\right\}\right)^6\mathrm{E}\left\{\lvert a_l\rvert^4\right\}\mathrm{E}\left\{\lvert a_m\rvert^8\right\}$	$4N_B\left(N_B-1\right)\left(N_B-2\right)\left(N_B-3\right)$
	$+2N_B(N_B-1)^2\left(N_B-2\right)$
	$+10N_B\left(N_B-1\right)\left(N_B-2\right)^2$
$\left(\mathrm{E}\left\{\lvert a_k\rvert^2\right\}\right)^6\left(\mathrm{E}\left\{\lvert a_l\rvert^6\right\}\right)^2$	$8N_B\left(N_B-1\right)$
	$+4N_B\left(N_B-1\right)\left(N_B-2\right)\left(N_B-3\right)$
	$+4N_B(N_B-1)^3$
	$+8N_B(N_B-1)^2\left(N_B-2\right)$
	$+8N_B\left(N_B-1\right)\left(N_B-2\right)$
$\left(\mathrm{E}\left\{\lvert a_k\rvert^2\right\}\right)^5\left(\mathrm{E}\left\{\lvert a_l\rvert^4\right\}\right)^2\mathrm{E}\left\{\lvert a_m\rvert^6\right\}$	$12N_B\left(N_B-1\right)\left(N_B-2\right)\left(N_B-3\right)$
	$+60N_B(N_B-1)^2\left(N_B-2\right)$
	$+40N_B\left(N_B-1\right)\left(N_B-2\right)^2$
	$+16N_B\left(N_B-1\right)+4N_B(N_B-1)^3$
	$+16N_B\left(N_B-1\right)\left(N_B-2\right)$
$\left(\mathrm{E}\left\{\lvert a_k\rvert^2\right\}\right)^5\mathrm{E}\left\{\lvert a_l\rvert^6\right\}\mathrm{E}\left\{\lvert a_m\rvert^8\right\}$	$4N_B(N_B-1)^2+4N_B\left(N_B-1\right)\left(N_B-2\right)$

$\left(\mathrm{E}\left\{\lvert a_k\rvert^2\right\}\right)^4\left(\mathrm{E}\left\{\lvert a_l\rvert^4\right\}\right)^4$	$N_B(N_B-1)^2(N_B-2)^2$ $+8N_B\left(N_B-1\right)\left(N_B-2\right)\left(N_B-3\right)$ $+65N_B\left(N_B-1\right)\left(N_B-2\right)^2$ $+32N_B(N_B-1)^2\left(N_B-2\right)$ $+8N_B\left(N_B-1\right)+N_B(N_B-1)^3$ $+8N_B\left(N_B-1\right)\left(N_B-2\right)$
$\left(\mathrm{E}\left\{\lvert a_k\rvert^2\right\}\right)^4\left(\mathrm{E}\left\{\lvert a_l\rvert^4\right\}\right)^2\mathrm{E}\left\{\lvert a_m\rvert^8\right\}$	$24N_B\left(N_B-1\right)\left(N_B-2\right)+2N_B(N_B-1)^2$
$\left(\mathrm{E}\left\{\lvert a_k\rvert^2\right\}\right)^4\mathrm{E}\left\{\lvert a_l\rvert^4\right\}\left(\mathrm{E}\left\{\lvert a_m\rvert^6\right\}\right)^2$	$16N_B(N_B-1)^2+8N_B\left(N_B-1\right)\left(N_B-2\right)$
$\left(\mathrm{E}\left\{\lvert a_k\rvert^2\right\}\right)^4\left(\mathrm{E}\left\{\lvert a_l\rvert^8\right\}\right)^2$	$N_B\left(N_B-1\right)$
$\left(\mathrm{E}\left\{\lvert a_k\rvert^2\right\}\right)^3\left(\mathrm{E}\left\{\lvert a_l\rvert^4\right\}\right)^3\mathrm{E}\left\{\lvert a_m\rvert^6\right\}$	$4N_B(N_B-1)^2\left(N_B-2\right)+36N_B\left(N_B-1\right)\left(N_B-2\right)$ $+24N_B(N_B-1)^2$
$\left(\mathrm{E}\left\{\lvert a_k\rvert^2\right\}\right)^3\mathrm{E}\left\{\lvert a_l\rvert^4\right\}\mathrm{E}\left\{\lvert a_m\rvert^6\right\}\mathrm{E}\left\{\lvert a_n\rvert^8\right\}$	$8N_B\left(N_B-1\right)$
$\left(\mathrm{E}\left\{\lvert a_k\rvert^2\right\}\right)^2\left(\mathrm{E}\left\{\lvert a_l\rvert^4\right\}\right)^5$	$2N_B(N_B-1)^2\left(N_B-2\right)+28N_B\left(N_B-1\right)\left(N_B-2\right)$ $+8N_B(N_B-1)^2$
$\left(\mathrm{E}\left\{\lvert a_k\rvert^2\right\}\right)^2\left(\mathrm{E}\left\{\lvert a_l\rvert^4\right\}\right)^3\mathrm{E}\left\{\lvert a_m\rvert^8\right\}$	$2N_B\left(N_B-1\right)\left(N_B-2\right)+8N_B\left(N_B-1\right)$
$\left(\mathrm{E}\left\{\lvert a_k\rvert^2\right\}\right)^2\left(\mathrm{E}\left\{\lvert a_l\rvert^4\right\}\right)^2\left(\mathrm{E}\left\{\lvert a_m\rvert^6\right\}\right)^2$	$4N_B(N_B-1)^2+4N_B\left(N_B-1\right)$
$\mathrm{E}\left\{\lvert a_k\rvert^2\right\}\left(\mathrm{E}\left\{\lvert a_l\rvert^4\right\}\right)^4\mathrm{E}\left\{\lvert a_m\rvert^6\right\}$	$4N_B(N_B-1)^2+8N_B\left(N_B-1\right)$
$\mathrm{E}\left\{\lvert a_k\rvert^2\right\}\left(\mathrm{E}\left\{\lvert a_l\rvert^4\right\}\right)^2\mathrm{E}\left\{\lvert a_m\rvert^6\right\}\mathrm{E}\left\{\lvert a_n\rvert^8\right\}$	$4N_B\left(N_B-1\right)$
$\left(\mathrm{E}\left\{\lvert a_k\rvert^4\right\}\right)^6$	$N_B\left(N_B-1\right)^2+4N_B\left(N_B-1\right)$
$\left(\mathrm{E}\left\{\lvert a_k\rvert^4\right\}\right)^4\mathrm{E}\left\{\lvert a_l\rvert^8\right\}$	$2N_B\left(N_B-1\right)$
$\left(\mathrm{E}\left\{\lvert a_k\rvert^4\right\}\right)^2\left(\mathrm{E}\left\{\lvert a_l\rvert^8\right\}\right)^2$	N_B

Proof. We first write

$$A=\sum_{k=1}^{N_B}\sum_{l=1}^{N_B}\sum_{m=1}^{N_B}\sum_{n=1}^{N_B}\sum_{o=1}^{N_B}\sum_{p=1}^{N_B}\lvert a_k\rvert^2\lvert b_l\rvert^2c_m^*b_mb_n^*d_na_o^*c_od_p^*a_p,$$

we then have $\mathrm{E}\left\{|A|^2\right\}$ given by

$$\begin{aligned}\mathrm{E}\left(|A|^2\right) = \sum_{k=1}^{N}\sum_{l=1}^{N}\sum_{m=1}^{N}\sum_{n=1}^{N}\sum_{o=1}^{N}\sum_{p=1}^{N}\sum_{k_1=1}^{N}\sum_{l_1=1}^{N}\sum_{m_1=1}^{N}\sum_{n_1=1}^{N}\sum_{o_1=1}^{N}\sum_{p_1=1}^{N}\mathrm{E}\left(|a_k|^2|a_{k_1}|^2 a_o^* a_p a_{o_1}^* a_{p_1}\right) \\ \cdot \mathrm{E}\left(|b_l|^2|b_{l_1}|^2 b_m b_n^* b_{m_1} b_{n_1}^*\right) \\ \cdot \mathrm{E}\left(c_m^* c_o c_{m_1}^* c_{o_1}\right)\mathrm{E}\left(d_n d_p^* d_{n_1} d_{p_1}^*\right)\end{aligned} \tag{A.22}$$

where the i.i.d. assumption is used for the derivation. Similar to the process in ***Lemma A.4***, we only consider the non-zero terms in the expectation, and summarize the non-zero terms in Table 5.4. We take the non-zero term $\left(\mathrm{E}\left(|a_k|^2\right)\right)^{12}$ as an example to show the number of terms as follows. In Equation (A.22), the following cases contribute to the expected non-zero term:

$$o = p = m = n, o_1 = p_1 = m_1 = n_1,$$
$$k \neq k_1 \neq o \neq o_1, l \neq l_1 \neq o \neq o_1;$$

$$o = o_1 = n = n_1, p = p_1 = m = m_1,$$
$$k \neq k_1 \neq o \neq p, l \neq l_1 \neq o \neq p;$$

$$o = o_1, p = p_1, m = m_1, n = n_1,$$
$$k \neq k_1 \neq o \neq p, l \neq l_1 \neq m \neq n;$$

$$o = o_1 = n = n_1, p = p_1, m = m_1,$$
$$k \neq k_1 \neq o \neq p, l \neq l_1 \neq o \neq p;$$

$$o = o_1, n = n_1, p = p_1 = m = m_1,$$
$$k \neq k_1 \neq o \neq p, l \neq l_1 \neq o \neq p.$$

for each of the first two cases $N_B\left(N_B-1\right)\left(N_B-2\right)^2\left(N_B-3\right)^2$ terms can be derived; for the third case, $N_B\left(N_B-1\right)\left(N_B-2\right)^3\left(N_B-3\right)^3$ terms are counted; and for the lase two cases,

we obtain $N_B\left(N_B-1\right)\left(N_B-2\right)^3\left(N_B-3\right)^2$ terms. Therefore, the final number of terms in Table 5.4 for the non-zero term $\left(\mathrm{E}\left(|a_k|^2\right)\right)^{12}$ is obtained. Following a similar analysis, we can derive the rest of the non-zero terms in Table 5.4.

Using the results in Equations (A.1) - (A.4), we have the result in (A.20). □

B Proof of Theorem 3

To begin with, we have the following Lemma:

Lemma A.6. $\mathbf{B}_{(u,v)}$ *is given by Equation* (5.22)*. When* $N_B > 4$*, we have*

$$E\left\{\left|\mathbf{B}_{(u,v)}\right|^2\right\} \leqslant \sqrt{\frac{A_1}{B_1}}, \tag{A.23}$$

if $u \in \mathbb{U}_1, v \neq u$*;*

$$E\left\{\left|\mathbf{B}_{(u,v)}\right|^2\right\} \leqslant \frac{\sqrt{A_2}}{B_1}, \tag{A.24}$$

if $v \in \{u-1, u+1\}$*;*

$$E\left\{\left|\mathbf{B}_{(u,v)}\right|^2\right\} \leqslant \sqrt{\frac{12A_2A_3 + 6A_1A_3^2 + 24A_4 + 48\sqrt{A_1A_2A_3^3}}{B_1^3}} \tag{A.25}$$

if $u \in \mathbb{U}_2$*,* $v \notin \{u-1, u, u+1\}$*;*

$$E\left\{\left|\mathbf{B}_{(u,u)}\right|^2\right\} \leqslant \sqrt{\frac{16A_3A_5}{B_1^3}}, \tag{A.26}$$

if $u \in \mathbb{U}_2$*,* $v = u$*.* $A_1 - A_5$ *and* B_1 *are respectively given by Equations* (A.13)*,* (A.18)*,* (A.8)*,* (A.21)*,* (A.14)*,* (A.9)*.*

Proof. The derivation of inequations (A.23), (A.24), (A.25), and (A.26), are respectively detailed in Appendix C, D, E, F. □

With the results in ***Lemma A.6***, we have

$$\begin{aligned}
&\mathrm{E}\left\{\|\mathbf{B}\|_{\mathrm{F}}^2\right\} = \sum_{u=1}^{N_U}\sum_{v=1}^{N_U}\mathrm{E}\left\{\left|\mathbf{B}_{(u,v)}\right|^2\right\} \\
&\leqslant \frac{N_U^2-1}{2}\sqrt{\frac{A_1}{B_1}} + (N_U-1)\frac{\sqrt{A_2}}{B_1} + \frac{(N_U-1)}{2}\sqrt{\frac{16A_3A_5}{B_1^3}} \\
&+\frac{N_U^2-4N_U+3}{2}\sqrt{\frac{12A_2A_3+6A_1A_3^2+24A_4+48\sqrt{A_1A_2A_3^3}}{B_1^3}}
\end{aligned} \tag{A.27}$$

Apparently, at the right hand side of the inequality (A.27), as the power in numerator is much less than that in denominator, we can derive

$$\lim_{N_B\to\infty}\mathrm{E}\left\{\|\mathbf{B}\|_{\mathrm{F}}^2\right\} = 0. \tag{A.28}$$

Applying the Markov's inequality, we have

$$\Pr\left\{\|\mathbf{B}\|_{\mathrm{F}}^2 < 1\right\} = 1 - \Pr\left\{\|\mathbf{B}\|_{\mathrm{F}}^2 \geqslant 1\right\} \geqslant 1 - \mathrm{E}\left\{\|\mathbf{B}\|_{\mathrm{F}}^2\right\}. \tag{A.29}$$

As $\|\mathbf{B}\|_{\mathrm{F}}^2 = \sum_{i=0}^{N_U-1}|\lambda_i|^2$, we have

$$\Pr\left\{|\lambda_0| < 1\right\} > \Pr\left\{\|\mathbf{B}\|_{\mathrm{F}}^2 < 1\right\} \geqslant 1 - \mathrm{E}\left\{\|\mathbf{B}\|_{\mathrm{F}}^2\right\}. \tag{A.30}$$

Therefore, we have

$$\lim_{N_B\to\infty}\Pr\left\{|\lambda_0| < 1\right\} = 1. \tag{A.31}$$

This complete the proof to ***Theorem 3***.

C Derivation of inequation (A.23)

For $u \in \mathbb{U}_1, v \neq u$, from Equation (5.22), we have

$$\mathrm{E}\left\{\left|\mathbf{B}_{(u,v)}\right|^2\right\} = \mathrm{E}\left\{\frac{\left|\mathbf{W}_{(u,v)}\right|^2}{\left|\mathbf{W}_{(u,u)}\right|^2}\right\} \leqslant \sqrt{\mathrm{E}\left\{\left|\mathbf{W}_{(u,v)}\right|^4\right\}\cdot\mathrm{E}\left\{\left|\mathbf{W}_{(u,u)}\right|^{-4}\right\}}, \tag{A.32}$$

where the Cauchy-Schwarz inequality is applied [23]. From ***Lemma A.3*** and ***Lemma A.2***, we have

$$\mathrm{E}\left\{|\mathbf{W}_{u,v}|^4\right\} = A_1, \tag{A.33}$$

$$\mathrm{E}\left\{\left|\mathbf{W}_{(u,u)}\right|^{-4}\right\} = \frac{1}{B_1}. \tag{A.34}$$

Therefore, inequation (A.23) is established.

D Derivation of inequation (A.24)

For $u \in \mathbb{U}_2$, $v = u - 1$, from Equation (5.22), we have

$$\mathbf{B}_{(u,u-1)} = \frac{\mathbf{G}_{(u,u+1)}\mathbf{G}_{(u+1,u-1)}}{\mathbf{G}_{(u,u)}\mathbf{G}_{(u+1,u+1)}}.$$

Applying the Cauchy-Schwarz inequality, we have

$$\mathrm{E}\left\{\left|\mathbf{B}_{(u,u-1)}\right|^2\right\} \leqslant \sqrt{\mathrm{E}\left\{\left|\mathbf{G}_{(u,u+1)}\mathbf{G}_{(u+1,u-1)}\right|^4\right\}} \cdot \sqrt{\mathrm{E}\left\{\left|\left(\mathbf{G}_{(u,u)}\mathbf{G}_{(u+1,u+1)}\right)^{-1}\right|^4\right\}} \tag{A.35}$$

According to ***Lemma A.4*** and ***Lemma A.2***, we have

$$\mathrm{E}\left\{\left|\mathbf{G}_{(u,u+1)}\mathbf{G}_{(u+1,u-1)}\right|^4\right\} = A_2, \tag{A.36}$$

$$\mathrm{E}\left\{\left|\left(\mathbf{G}_{(u,u)}\mathbf{G}_{(u+1,u+1)}\right)^{-1}\right|^4\right\} = \mathrm{E}\left\{\left|\left(\mathbf{G}_{(u,u)}\right)^{-1}\right|^4\right\} \cdot \mathrm{E}\left\{\left|\left(\mathbf{G}_{(u,u)}\right)^{-1}\right|^4\right\} = \frac{1}{B_1^2}. \tag{A.37}$$

For $u \in \mathbb{U}_2$, $v = u + 1$, following a similar process, we have the same result above. Therefore, we complete the derivation of the inequation (A.24).

E Derivation of inequation (A.25)

For $u \in \mathbb{U}_2$, $v \notin \{u-1, u, u+1\}$, from Equation (5.22), we have $\mathrm{E}\left\{\left|\mathbf{B}_{(u,v)}\right|^2\right\}$ given by

$$\begin{aligned}
&\mathrm{E}\left\{\left|\mathbf{B}_{(u,v)}\right|^2\right\} \\
=&\mathrm{E}\left\{\frac{\left|\mathbf{G}_{(u+1,u+1)}\mathbf{G}_{(u,u-1)}\mathbf{G}_{(u-1,v)} + \mathbf{G}_{(u-1,u-1)}\mathbf{G}_{(u,u+1)}\mathbf{G}_{(u+1,v)} - \mathbf{G}_{(u-1,u-1)}\mathbf{G}_{(u+1,u+1)}\mathbf{G}_{(u,v)}\right|^2}{\left|\mathbf{G}_{(u-1,u-1)}\mathbf{G}_{(u,u)}\mathbf{G}_{(u+1,u+1)}\right|^2}\right\} \\
\leqslant&\sqrt{\mathrm{E}\left\{\left|\mathbf{G}_{(u+1,u+1)}\mathbf{G}_{(u,u-1)}\mathbf{G}_{(u-1,v)} + \mathbf{G}_{(u-1,u-1)}\mathbf{G}_{(u,u+1)}\mathbf{G}_{(u+1,v)} - \mathbf{G}_{(u-1,u-1)}\mathbf{G}_{(u+1,u+1)}\mathbf{G}_{(u,v)}\right|^4\right\}} \\
&\cdot \sqrt{\mathrm{E}\left\{\left|\mathbf{G}_{(u-1,u-1)}\mathbf{G}_{(u,u)}\mathbf{G}_{(u+1,u+1)}\right|^{-4}\right\}}
\end{aligned} \tag{A.38}$$

where the Cauchy-Schwarz inequality is applied. Next, we have the first expectation in the right hand side of the inequality (A.38) given by

$$\begin{aligned}&\mathrm{E}\left\{\left|\mathbf{G}_{(u+1,u+1)}\mathbf{G}_{(u,u-1)}\mathbf{G}_{(u-1,v)}+\mathbf{G}_{(u-1,u-1)}\mathbf{G}_{(u,u+1)}\mathbf{G}_{(u+1,v)}-\mathbf{G}_{(u-1,u-1)}\mathbf{G}_{(u+1,u+1)}\mathbf{G}_{(u,v)}\right|^4\right\}\\&=\mathrm{E}\left\{(A+B+C+D+E+F)^2\right\}\leqslant 6\mathrm{E}\left\{A^2+B^2+C^2+D^2+E^2+F^2\right\}\end{aligned}\tag{A.39}$$

where

$$\begin{aligned}A&=\left|\mathbf{G}_{(u+1,u+1)}\right|^2\left|\mathbf{G}_{(u,u-1)}\mathbf{G}_{(u-1,v)}\right|^2,\\B&=\left|\mathbf{G}_{(u-1,u-1)}\right|^2\left|\mathbf{G}_{(u,u+1)}\mathbf{G}_{(u+1,v)}\right|^2,\\C&=\left|\mathbf{G}_{(u-1,u-1)}\right|^2\left|\mathbf{G}_{(u+1,u+1)}\right|^2\left|\mathbf{G}_{(u,v)}\right|^2,\\D&=2\,\mathrm{Re}\left(\mathbf{G}_{(u,u-1)}\mathbf{G}_{(u-1,v)}\mathbf{G}^*_{(u,u+1)}\mathbf{G}^*_{(u+1,v)}\right)\\&\quad\cdot\mathbf{G}_{(u+1,u+1)}\mathbf{G}_{(u-1,u-1)},\\E&=-2\,\mathrm{Re}\left(\mathbf{G}_{(u,u-1)}\mathbf{G}_{(u-1,v)}\mathbf{G}^*_{(u,v)}\right)\\&\quad\cdot\left|\mathbf{G}_{(u+1,u+1)}\right|^2\mathbf{G}_{(u-1,u-1)},\\F&=-2\,\mathrm{Re}\left(\mathbf{G}_{(u,u+1)}\mathbf{G}_{(u+1,v)}\mathbf{G}^*_{(u,v)}\right)\\&\quad\cdot\left|\mathbf{G}_{(u-1,u-1)}\right|^2\mathbf{G}_{(u+1,u+1)}.\end{aligned}$$

The inequality (A.39) holds by noting that

$$\begin{aligned}&(A+B+C+D+E+F)^2\\&\leqslant 6\left(A^2+B^2+C^2+D^2+E^2+F^2\right),\end{aligned}$$

where A, B, C, D, E, F are both real numbers. Next, we derive the expectations as follows individually.

With the results in ***Lemma A.2*** and ***Lemma A.4***, we have $\mathrm{E}\left(A^2\right)=\mathrm{E}\left(B^2\right)$ given by

$$\mathrm{E}\left(A^2\right)=\mathrm{E}\left(B^2\right)=A_2A_3.\tag{A.40}$$

$\mathrm{E}\left(C^2\right)$ is given by

$$\mathrm{E}\left\{C^2\right\} = A_1 {A_3}^2. \tag{A.41}$$

where the results in ***Lemma A.2*** and ***Lemma A.3*** are applied.

By using $\left(\mathrm{Re}\left(a\right)\right)^2 \leqslant |a|^2$, we derive the result of $\mathrm{E}\left\{D^2\right\}$, given by

$$\mathrm{E}\left\{D^2\right\} \leqslant 4\mathrm{E}\left\{\left|\mathbf{G}_{(u+1,u+1)}\mathbf{G}_{(u-1,u-1)}\mathbf{G}_{(u,u-1)}\mathbf{G}_{(u-1,v)}\mathbf{G}^*_{(u,u+1)}\mathbf{G}^*_{(u+1,v)}\right|^2\right\} = 4A_4, \tag{A.42}$$

where A_4 is obtained through ***Lemma A.5***.

Applying the Cauchy-Schwarz inequality, we have

$$\begin{aligned}\mathrm{E}\left\{E^2\right\} &\leqslant 4\mathrm{E}\left\{\left|\mathbf{G}_{(u-1,u-1)}\mathbf{G}_{(u,u-1)}\mathbf{G}_{(u-1,v)}\mathbf{G}^*_{(u,v)}\right|^2\right\} \cdot \mathrm{E}\left\{\left|\mathbf{G}_{(u+1,u+1)}\right|^4\right\} \\ &\leqslant 4\mathrm{E}\left\{\left|\mathbf{G}_{(u+1,u+1)}\right|^4\right\} \sqrt{\mathrm{E}\left\{\left|\mathbf{G}_{(u,u-1)}\mathbf{G}_{(u-1,v)}\right|^4\right\}} \cdot \sqrt{\mathrm{E}\left\{\left|\mathbf{G}_{(u-1,u-1)}\mathbf{G}_{(u,v)}\right|^4\right\}}\end{aligned} \tag{A.43}$$

With the results in ***Lemma A.2***, ***Lemma A.3***, and ***Lemma A.4***, we derive the result of $\mathrm{E}\left\{E^2\right\} = \mathrm{E}\left\{F^2\right\}$, given by

$$\mathrm{E}\left\{E^2\right\} = \mathrm{E}\left\{F^2\right\} \leqslant 4A_3\sqrt{A_1A_2A_3}. \tag{A.44}$$

Therefore, we derive

$$\mathrm{E}\left\{\left|\mathbf{B}_{(u,v)}\right|^2\right\} \leqslant \sqrt{\frac{12A_2A_3 + 6A_1A_3^2 + 24A_4 + 48\sqrt{A_1A_2A_3^3}}{B_1^3}} \tag{A.45}$$

F Derivation of inequation (A.26)

For $u \in \mathbb{U}_2$, $v = u$, from Equation (5.22), we have $\mathrm{E}\left\{\left|\mathbf{B}_{(u,v)}\right|^2\right\}$ given by

$$\begin{aligned}\mathrm{E}\left\{\left|\mathbf{B}_{(u,u)}\right|^2\right\} =& \mathrm{E}\left\{\frac{\left|\mathbf{G}_{(u+1,u+1)}\left|\mathbf{G}_{(u,u-1)}\right|^2 + \mathbf{G}_{(u-1,u-1)}\left|\mathbf{G}_{(u,u+1)}\right|^2\right|^2}{\left|\mathbf{G}_{(u,u)}\mathbf{G}_{(u-1,u-1)}\mathbf{G}_{(u+1,u+1)}\right|^2}\right\} \\ \leqslant& \sqrt{\mathrm{E}\left\{\left|\mathbf{G}_{(u+1,u+1)}\left|\mathbf{G}_{(u,u-1)}\right|^2 + \mathbf{G}_{(u-1,u-1)}\left|\mathbf{G}_{(u,u+1)}\right|^2\right|^4\right\}} \\ &\cdot \sqrt{\mathrm{E}\left\{\left|\mathbf{G}_{(u,u)}\mathbf{G}_{(u-1,u-1)}\mathbf{G}_{(u+1,u+1)}\right|^{-4}\right\}},\end{aligned} \tag{A.46}$$

where the Cauchy-Schwarz inequality is applied. By using $|a+b|^2 \leqslant 2\left(|a|^2+|b|^2\right)$, we have

$$\begin{aligned}
&\left|\mathbf{G}_{(u+1,u+1)}\left|\mathbf{G}_{(u,u-1)}\right|^2+\mathbf{G}_{(u-1,u-1)}\left|\mathbf{G}_{(u,u+1)}\right|^2\right|^2 \\
\leqslant\ & 2\left(\left|\mathbf{G}_{(u+1,u+1)}\right|^2\left|\mathbf{G}_{(u,u-1)}\right|^4+\left|\mathbf{G}_{(u-1,u-1)}\right|^2\left|\mathbf{G}_{(u,u+1)}\right|^4\right), \\
&\left|\mathbf{G}_{(u+1,u+1)}\left|\mathbf{G}_{(u,u-1)}\right|^2+\mathbf{G}_{(u-1,u-1)}\left|\mathbf{G}_{(u,u+1)}\right|^2\right|^4 \\
\leqslant\ & 8\left(\left|\mathbf{G}_{(u+1,u+1)}\right|^4\left|\mathbf{G}_{(u,u-1)}\right|^8+\left|\mathbf{G}_{(u-1,u-1)}\right|^4\left|\mathbf{G}_{(u,u+1)}\right|^8\right).
\end{aligned} \tag{A.47}$$

Therefore, we derive

$$\begin{aligned}
&\mathrm{E}\left\{\left|\mathbf{G}_{(u+1,u+1)}\left|\mathbf{G}_{(u,u-1)}\right|^2+\mathbf{G}_{(u-1,u-1)}\left|\mathbf{G}_{(u,u+1)}\right|^2\right|^4\right\} \\
\leqslant & 8\mathrm{E}\left(\left|\mathbf{G}_{(u+1,u+1)}\right|^4\right)\mathrm{E}\left(\left|\mathbf{G}_{(u,u-1)}\right|^8\right)+8\mathrm{E}\left(\left|\mathbf{G}_{(u-1,u-1)}\right|^4\right)\mathrm{E}\left(\left|\mathbf{G}_{(u,u+1)}\right|^8\right).
\end{aligned} \tag{A.48}$$

With the results in ***Lemma A.2*** and ***A.3***, we have

$$\mathrm{E}\left\{\left|\mathbf{B}_{(u,u)}\right|^2\right\} \leqslant \sqrt{\frac{16A_3A_5}{B_1^3}}. \tag{A.49}$$

Bibliography

[1] M. Shafi, A. Molisch, P. Smith, T. Haustein, P. Zhu, P. Silva, F. Tufvesson, A. Benjebbour, and G. Wunder. 5g: A tutorial overview of standards, trials, challenges, deployment, and practice. *IEEE Journal on Selected Areas in Communications*, 35(6):1201–1221, 2017.

[2] T. Marzetta. Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas. *IEEE Transactions on Wireless Communications*, 9(11):3590–3600, Nov. 2010.

[3] E. Larsson, O. Edfors, F. Tufvesson, and T. Marzetta. Massive MIMO for Next Generation Wireless Systems. *IEEE Communications Magzine*, 52(2):186–195, Feb. 2014.

[4] T. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. W. J. Schulz, M. Samimi, and F. Gutierrez. Millimeter Wave Mobile Communications for 5G Cellular: It Will Work! *IEEE Access*, 1:335–349, May 2013.

[5] C. Mecklenbrauker, A. Molisch, J. Karedal, F. Tufvesson, A. Paier, L. Bernado, T. Zemen, O. Klemp, and N. Czink. Vehicular Channel Characterization and Its Implications for Wireless System Design and Performance. *IEEE Proceedings*, 99(7):1189–1212, July 2011.

[6] X. Ge, H. Cheng, G. Mao, Y. Yang, and S. Tu. Vehicular Communications for 5G Cooperative Small-Cell Networks. *IEEE Transactions on Vehicular Technology*, 65(10):7882–7894, Oct. 2016.

[7] X. Ge, Z. Li, and S. Li. 5G Software Defined Vehicular Networks. *IEEE Communications Magzine*, 55(7):87–93, 2017.

[8] S. Mumtaz, J. Jornet, J. Aulin, W. Gerstacker, X. Dong, and B. Ai. Terahertz Communication for Vehicular Networks. *IEEE Transactions on Vehicular Technology*, 66(7):5617–5625, Jul. 2017.

[9] W. Viriyasitavat, M. Boban, H. Tsai, and A. Vasilakos. Vehicular Communications: Survey and Challenges of Channel and Propagation Models. *IEEE Vehicular Technology Magzine*, 10(2):55–66, 2015.

[10] Z. Su, Y. Hui, and Q. Yang. The Next Generation Vehicular Networks: A Content-Centrics Framework. *IEEE Wireless Communications Magzine*, 24(1):60–66, Feb. 2017.

[11] P. Li, T. Zhang, C. Huang, X. Chen, and B. Fu. RSU-Assisted Geocast in Vehicular Ad Hoc Networks. *IEEE Wireless Communications Magzine*, 24(1):53–59, Feb. 2017.

[12] K. Zheng, Q. Zheng, P. Chatzimisios, W. Xiang, and Y. Zhou. Heterogeneous Vehicular Networking: A Survey on Architecture, Challenges, and Solutions. *IEEE Communications Survey & Tutorials*, 17(4):2377–2396, 2015.

[13] X. Zhang, J. Cui, S. Das, M. Gerla, and M. Chitre. Underwater Wireless Communications and Networks: Theory and Application: Part 1 [Guest Editorial]. *IEEE Communications Magzine*, 53(11):40–41, Nov. 2015.

[14] P. Qarabaqi and M. Stojanovic. Statistical Characterization and Computationally Efficient Modeling of a Class of Underwater Acoustic Communication Channels. *IEEE Journal of Oceanic Engineering*, 38(4):701–717, Oct. 2013.

[15] S. Zhou and Z. Wang. *OFDM for Underwater Acoustic Communications*. John Wiley & Sons., 2014.

[16] A. Rafati, H. Lou, and C. Xiao. Soft-Decision Feedback Turbo Equalization for LDPC-Coded MIMO Underwater Acoustic Communications. *IEEE Journal of Oceanic Engi-*

neering, 39(1):90–99, Jan. 2014.

[17] J. Tao. Turbo Equalization for MIMO SC-FDMA Underwater Acoustic Communications, *Proceedings of the MTS/IEEE OCEANS*, Monterey, CA, USA, 2016.

[18] X. Zhao and D. Pompili. AMMCA: Acoustic Massive MIMO with Carrier Aggregation to Boost the Underwater Communication Data Rate. in *Proceedings of the 10th International Conference on Underwater Networks & Systems*, Arlington, VA, USA, Oct. 2015.

[19] A. Aminjavaheri and B. Farhang-Boroujeny. UWA Massive MIMO Communications, in *Proceedings of the MTS/IEEE OCEANS*, Washington, DC, USA, Oct. 2015.

[20] S. Moshavi. Multi-user Detection for DS-CDMA Communications. *IEEE Communications Magzine*, 34(10):124–136, Oct. 1996.

[21] T. Li, W. Wang, and X. Gao. Turbo Equalization for LTE Uplink under Imperfect Channel Estimation. in *Proceedings of the IEEE 20th International Symposium on Personal, Indoor and Mobile Radio Communications*, pages: 330–334, Tokyo, Japan, Sept. 2009.

[22] S. Chen, W. Wang, and X. Gao. Sorted QR Decomposition Based Detection for MU-MIMO LTE Uplink. in *Proceedings of the IEEE 71st Vehicular Technology Conference*, pages: 1–5, Taipei, May 2010.

[23] M. Wu, B. Yin, G. Wang, C. Dick, J. Cavallaro, and C. Studer. Large-Scale MIMO Detection for 3GPP LTE: Algorithms and FPGA Implementations, *IEEE Journal on Selected Topics in Signal Processing*, 8(5):916–929, Oct. 2014.

[24] A. Ghazal, C. Wang, B. Ai, D. Yuan, and H. Haas. A Nonstationary Wideband MIMO Channel Model for High-Mobility Intelligent Transportation Systems, *IEEE Transactions on Vehicular Technology*, 16(2):885–897, 2015.

[25] D. Tian, J. Zhou, Z. Sheng, and V. Leung. Robust Energy-Efficient MIMO Trans-

mission for Cognitive Vehicular Networks, *IEEE Transactions on Vehicular Technology*, 65(6):3845–3859, 2016.

[26] J. Mietzner, R. Schober, L. Lampe, W. Gerstacker, P. Hoeher. Multiple-antenna Techniques for Wireless Communications - A Comprehensive Literature Survey. *IEEE Communications Survey & Tutorials*, 11(2):87–105, Second Quarter 2009.

[27] G. Foschini. Layered Space-time Architecture for Wireless Communication in a Fading Environment When Using Multi-element Antennas. *Bell Labs Technical Journal*, pp. 41–59, Autumn 1996.

[28] S. Alamouti. A Simple Transmit Diversity Technique for Wireless Communications. *IEEE Journal on Selected Areas in Communications*, 16(8):1451–1458, Oct. 1998.

[29] H. Ngo, E. Larsson, T. Marzetta. Energy and Spectral Efficiency of Very Large Multiuser MIMO Systems. *IEEE Transactions on Communications*, 61(4):1436–1449, Apr. 2013.

[30] J. Hoydis, S. Brink, and M. Debbah. Massive MIMO in the UL/DL of Cellular Networks: How Many Antennas Do We Need?. *IEEE Journal on Selected Areas in Communications*, 31(2):160–171, Feb. 2013.

[31] E. Bjornson, L. Sanguinetti, J. Hoydis, and M. Debbah. Optimal Design of Energy-Efficient Multi-User MIMO Systems: Is Massive MIMO the Answer?. *IEEE Transactions on Wireless Communications*, 14(6):3059–3075, Jun. 2015.

[32] W. Liu, S. Han, and C. Yang. Energy Efficiency Scaling Law of Massive MIMO Systems. *IEEE Transactions on Communications*, 65(1):107–121, Jan. 2017.

[33] J. Choi, J. Mo, and R. Heath. Near Maximum-Likelihood Detector and Channel Estimator for Uplink Multiuser Massive MIMO Systems With One-Bit ADCs, *IEEE Transactions on Vehicular Technology*, 64(5):2005–2018, May 2016.

[34] L. Lu, G. Li, A. Swindlehurst, A. Ashikhmin, and R. Zhang. An Overview of Massive

MIMO: Benefits and Challenges, *IEEE Journal on Selected Topics in Signal Processing*, 8(5):742–758, Oct. 2014.

[35] R. Zhang, Z. Zhong, J. Zhao, B. Li, K. Wang. Channel Measurement and Packet-Level Modeling for V2I Spatial Multiplexing Uplinks Using Massive MIMO, *IEEE Transactions on Vehicular Technology*, 65(10):7831–7843, Oct. 2016.

[36] J. Choi, V. Va, N. Gonzalez-Prelcic, R. Daniels, C. Bhat, and R. Heath. Millimeter Wave Vehicular Communication to Support Massive Automotive Sensing, *IEEE Communications Magzine*, 54(12):160–167, Dec. 2016.

[37] L. Liang, H. Peng, G. Li, and X. Shen. Vehicular Communications: A Physical Layer Perspective, *IEEE Transactions on Vehicular Technology*, 2017.

[38] S. Yang and L. Hanzo. Fifty Years of MIMO Detection: The Road to Large-Scale MIMOs, in *IEEE Communications Surveys & Tutorials*, 17(4):1941–1988, 2015.

[39] F. Jiang and C. Li. Soft Input Soft Output MMSE-SQRD Based Turbo Equalization for MIMO-OFDM Systems under Imperfect Channel Estimation, in *Proceedings of IEEE Global Communications Conference*, San Diego, CA, Dec. 2015, DOI: 10.1109/GLOCOM.2015.7417541.

[40] A. Rafati, H. Lou, and C. Xiao. Low-Complexity Soft-Decision Feedback Turbo Equalization for MIMO Systems With Multilevel Modulations. *IEEE Transactions on Vehicular Technology*, 60(7):3218–3227, 2011.

[41] W. Zhong, A. Lu, and X. Gao. MMSE SQRD based SISO Detection for Coded MIMO-OFDM Systems. *Science China Information Science*, 57(4):1–10, Apr. 2014.

[42] C. Studer and H. Bolcskei. Soft-Input Soft-Output Single Tree-Search Sphere Decoding. *IEEE Transactions on Information Theory*, 56(10):4827–4842, Oct. 2010.

[43] S. Ahmed, T. Ratnarajah, M. Sellathurai, and C. Cowan. Iterative Receivers for MIMO-OFDM and Their Convergence Behavior. *IEEE Transactions on Vehicular Tech-*

nology, 58(1):461–468, Jan. 2009.

[44] N. Miridakis and D. Vergados. A Survey on the Successive Interference Cancellation Performance for Single-Antenna and Multiple-Antenna OFDM Systems. *IEEE Communications Surveys & Tutorials*, 15(1):312–335, Feb. 2013.

[45] W. Zhang, X. Xia, and K. Letaief. Space-Time/Frequency Coding for MIMO-OFDM in Next Generation Broadband Wireless Systems. *IEEE Transactions on Signal Processing*, 4(3):32–43, Jun. 2007.

[46] J. Tao. Single-Carrier Frequency-Domain Turbo Equalization With Various Soft Interference Cancellation Schemes for MIMO Systems. *IEEE Transactions on Communications*, 63(9):3206–3217, 2015.

[47] J. Zhang and Y. Zheng. Frequency-domain Turbo Equalization with Soft Successive Interference Cancellation for Single-Carrier MIMO Underwater Acoustic Communications. *IEEE Transactions on Wireless Communications*, 10(9):2872–2882, 2011.

[48] C. Shah, C. Tsimenidis, B. Sharif, and J. Neasham. Low-Complexity Iterative Receiver Structure for Time-Varying Frequency-Selective Shallow Underwater Acoustic Channels Using BICM-ID: Design and Experimental Results. *IEEE Journal of Oceanic Engineering*, 36(3):406–421, Jul. 2011.

[49] D. Liu and M. Fitz. Low Complexity Affine MMSE Detector for Iterative Detection-decoding MIMO-OFDM Systems. *IEEE Transactions on Communications*, 56(1):150–158, Jan. 2008.

[50] X. Wang and V. Poor. Iterative (Turbo) Soft Interference Cancellation and Decoding for Coded CDMA. *IEEE Transactions on Communications*, 47:1046–1061, 1999.

[51] J. Wu and Y. Zheng. Low Complexity Soft-input Soft-output Block Decision Feedback Equalization, *IEEE Journal on Selected Areas in Communications*, 26(2), Feb. 2008, DOI: 10.1109/JSAC.2008.080205.

[52] M. Hajjar and L. Hanzo. EXIT Charts for System Design and Analysis. *IEEE Communications Surveys & Tutorials*, 16(1):127–153, Feb. 2014.

[53] L. Hanzo, J. Woodard, and P. Robertson. Turbo Decoding and Detection for Wireless Applications. *Proceedings of the IEEE*, 95(6):1178–1200, Jul. 2007.

[54] X. Li and J. Ritcey. Bit-interleaved Coded Modulation with Iterative Decoding. *IEEE Communications Letters*, 1(6):169–171, Nov. 1997.

[55] M. Tuchle, A. Singer, and R. Koetter. Minimum mean squared error equalization using a priori information. *IEEE Transactions on Signal Processing*, 50(3):673–683, Mar. 2002.

[56] L. Xiao, P. Yang, Y. Xiao, J. Liu, S. Fan, B. Dong, and S. Li. Reduced-Complexity Iterative-Detection-Aided Generalized Space-Time Shift Keying. *IEEE Signal Processing Letters*, 23(1):30–34, Jan. 2016.

[57] S. Sugiura, C. Xu, S. Ng, and L. Hanzo. An Improved Soft-Input Soft-Output Detector for Generalized Spatial Modulation. *IEEE Transactions on Vehicular Technology*, 61(8):3656–3664, Oct. 2012.

[58] B. Dhivagar, K. Kuchi, and K. Giridhar. An Iterative DFE Receiver for MIMO SC-FDMA Uplink. *IEEE Communications Letters*, 18(12):2141–2144, Oct. 2014.

[59] S. Cheng and R. Narasimhan. Soft Interference Cancellation Receiver for SC-FDMA Uplink in LTE. *IEEE Wireless Communication and Networks Conference (WCNC)*, pp. 3318–3322, Shanghai, China, 2013.

[60] F. Jiang, Y. Zhang, and C. Li. A New SQRD-Based Soft Interference Cancelation Scheme in Multi-User MIMO SC-FDMA System. *IEEE Communications Letters*, 21(4):821–824, Apr. 2017. DOI:10.1109/LCOMM.2016.2642189.

[61] S. Feng, H. Mao, W. Feng, N. Ge, and J. Lu. Iterative Soft QRD-M Detection and Decoding for Single Carrier Block Transmission Systems. *IEEE Wireless Communication and Networks Conference (WCNC)*, pp. 618–623, Istanbul, Turkey, Apr. 2014.

[62] D. Wubben, R. Bohnke, V. Kuhn, and K. Kammeyer. MMSE extension of V-BLAST based on sorted QR decomposition. in *Proceedings of the IEEE 58th Vehicular Technology Conference*, pages: 508–512, Orlando, USA, Oct. 2003.

[63] H. Senol, E. Panayirci, and H. Poor. Near-Optimal Channel Estimation for OFDM in Fast-Fading Channels, *IEEE Transactions on Signal Processing*, 60(8):4236–4253, August 2012.

[64] P. Wan, M. McGuire, and X. Dong. Nondata-Aided Joint Channel Estimation and Equalization for OFDM Systems in Very Rapidly Varying Mobile Channels, *IEEE Transactions on Vehicular Technology*, 60(8):3780–3791, October 2011.

[65] T. Weber, A. Sklavos, and M. Meurer. Imperfect Channel-state Information in MIMO Transmission, *IEEE Transactions on Communications*, 54(3):543–552, Mar. 2006.

[66] M. Jiang, G. Yue, N. Prasad, and S. Rangarajan. Link Adaptation in LTE-A Uplink with Turbo SIC Receivers and Imperfect Channel Eestimation. in *Proceedings of the 45th Annual Conference on Information Sciences and Systems (CISS)*, pages: 1–6, Baltimore, USA, Mar. 2011.

[67] C. Wang, E. Au, R. Murch, W. Mow, R. Cheng, and V. Lau. On the Performance of the MIMO Zero-Forcing Receiver in the Presence of Channel Estimation Error. *IEEE Transactions on Wireless Communications*, 6(3):805–810, Mar. 2007.

[68] J. Wang, O. Wen, and S. Li. Soft-Output MMSE V-BLAST Detector under ML Channel Estimation and Channel Correlation. *IEEE communications Letters*, 13(2):103–105, Feb. 2009.

[69] F. Jiang, C. Li, C. Meng, and Z. Gong. A New Turbo Equalizer Conditioned on Estimated Channel for MIMO MMSE Receiver. *IEEE communications Letters*, 21(4):957–960, Apr. 2017. DOI:10.1109/LCOMM.2016.2638823.

[70] J. Wang, O. Wen, and S. Li. Soft-Output MMSE MIMO Detector under Imperfect

Channel Estimation. *IEEE Wireless Communication and Networks Conference (WCNC)*, pages: 1334–1338, Las Vegas, NV, Mar. 2008.

[71] L. Jalloul, S. Alex, and M. Mansour. Soft-Output MIMO Detectors with Channel Estimation Error. *IEEE Signal Processing Letters*, 22(7):993–997, Jul. 2015.

[72] K. Vardhan, S. Mohammed, A. Chockalingam, and B. Rajan. A Low-Complexity Detector for Large MIMO Systems and Multicarrier CDMA Systems. *IEEE Journal on Selected Areas in Communications*, 26(3):473–485, Apr. 2008.

[73] S. Mohammed, A. Zaki, A. Chockalingam, and B. Rajan. High-Rate Space-Time Coded Large-MIMO Systems: Low-Complexity Detection and Channel Estimation. *IEEE Journal of Selected Topics in Signal Processing*, 3(6):958–974, Dec. 2009.

[74] T. Datta, N. Srinidhi, A. Chockalingam, and B. Rajan. A Novel Monte-Carlo-Sampling-Based Receiver for Large-Scale Uplink Multiuser MIMO Systems. *IEEE Transactions on Vehicular Technology*, 62(7):3019–3038, 2013.

[75] L. Bai, T. Li, J. Liu, Q. Yu, and J. Choi. Large-Scale MIMO Detection Using M-CMC Approach With Blockwise Sampling. *IEEE Transactions on Communications*, 64(9):3697–3707, 2016.

[76] P. Som, T. Datta, N. Srinidhi, A. Chockalingam, and B. Rajan. Low-Complexity Detection in Large-Dimension MIMO-ISI Channels Using Graphical Models. *IEEE Journal of Selected Topics in Signal Processing*, 5(8):1497–1511, Dec. 2011.

[77] S. Wu, L. Kuang, Z. Ni, J. Lu, D. Huang, and Q. Guo. Low-Complexity Iterative Detection for Large-Scale Multiuser MIMO-OFDM Systems Using Approximate Message Passing. *IEEE Journal of Selected Topics in Signal Processing*, 8(5):902–915, Oct. 2014.

[78] Y. Cai, R. Lamare, B. Champagne, B. Qin, and M. Zhao. Adaptive Reduced-Rank Receive Processing Based on Minimum Symbol-Error-Rate Criterion for Large-Scale Multiple-Antenna Systems. *IEEE Transactions on Communications*, 63(11):4185–4201,

Nov. 2015.

[79] R. Lamare. Adaptive and Iterative Multi-Branch MMSE Decision Feedback Detection Algorithms for Multi-Antenna Systems. *IEEE Transactions on Wireless Communications*, 12(10):5294–5308, Oct. 2013.

[80] D. Zhu, B. Li, and P. Liang. On the Matrix Inversion Approximation Based on Neumann Series in Massive MIMO Systems. in *Proceedings of the IEEE International Conference on Communications (ICC)*, pages: 1763–1769, London, UK, Jun. 2015.

[81] C. Tang, C. Liu, L. Yuan, and Z. Xing. High Precision Low Complexity Matrix Inversion Based on Newton Iteration for Data Detection in the Massive MIMO, *IEEE Communications Letters*, 20(3):490–493, Mar. 2016.

[82] F. Wang, C. Zhang, X. Liang, Z. Wu, S. Xu, and X. You. Efficient Iterative Soft Detection Based on Polynomial Approximation for Massive MIMO. in *Proceedings of the IEEE International Conference on Wireless Communications and Signal Processing (WCSP)*, pages: 1–5, Nanjing, China.

[83] X. Gao, L. Dai, Y. Ma, and Z. Wang. Low-complexity Near-optimal Signal Detection for Uplink Large Scale MIMO Systems. *Electronics Letters*, 50(18):1326–1328, Aug. 2014.

[84] L. Dai, X. Gao, X. Su, S. Han, C. I, and Z. Wang. Low-Complexity Soft-Output Signal Detection Based on Gauss-Seidel Method for Uplink Multiuser Large-Scale MIMO Systems. *IEEE Transactions on Vehicular Technology*, 64(10):4839–4845, Oct. 2015.

[85] X. Qin, Z. Yan, and G. He. A Near-Optimal Detection Scheme Based on Joint Steepest Descent and Jacobi Method for Uplink Massive MIMO Systems, *IEEE Communications Letters*, 20(2):276–279, Feb. 2016.

[86] B. Yin, M. Wu, J. Cavallaro, and C. Studer. Conjugate Gradient-based Soft-output Detection and Precoding in Massive MIMO Systems in *Proceedings of the IEEE Global Comunications Conference*, pages: 3696–3701, Austin, TX, USA, Dec. 2014.

[87] Z. Wu, C. Zhang, Y. Xue, S. Xu, and X. You. Efficient Architecture for Soft-output Massive MIMO Detection with Gauss-Seidel Method in *Proceedings of the IEEE 2016 International Symposium on Circuits and Systems (ISCAS)*, pages: 1886–1889, Montreal, CA, May 2016.

[88] J. Goldstein, I. Reed, and L. Scharf. A Multistage Representation of the Wiener Filter Based on Orthogonal Projections, *IEEE Transactions on Information Theory*, 44(7):2943–2959, Nov. 1998.

[89] M. Honig and J. Goldstein. Adaptive Reduced-rank Interference Suppression Based on the Multistage Wiener Filter, *IEEE Transactions on Communications*, 50(6):986–994, Jun. 2002.

[90] J. Burgerscentrum. Iterative Solutions Methods, *Applied Numerical Mathematics*, 51(4):437–450, 2011.

[91] H. Lou. Stair Matrices and Their Generalizations with Applications to Iterative Methods I: A Generalization of the Successive Overrelaxation Method, *SIAM Journal on Numerical Analysis*, 37(1):1–17, 2000.

[92] H. Li, T. Huang, Y. Zhang, X. Liu, and T. Gu. Chebyshev-type Methods and Preconditioning Techniques, *Applied Mathematics and Computation, Elsevier*, 218(2):260–270, 2011.

[93] L. Zhang, X. Xu, W. Feng, and Y. Chen. Multi-array Iterative Receiver for Underwater Acoustic OFDM Communications with EXIT Chart Evaluation, *Applied Acoustics*, 114:307–316, Dec. 2016.

[94] M. Bilodeau and D. Brenner. Theory of Multivariate Statistics, *New York: Springer*, 1999.

[95] L. Dai, Z. Wang, and Z. Yang. Spectrally Efficient Time-Frequency Training OFDM for Mobile Large-Scale MIMO Systems, *IEEE Journal on Selected Areas in Communications*,

31(2):251–263, Feb. 2013.

[96] Z. Gao, L. Dai, W. Dai, B. Shim, and Z. Wang. Structured Compressive Sensing-Based Spatio-Temporal Joint Channel Estimation for FDD Massive MIMO, *IEEE Transactions on Communications*, 64(2):601–617, Feb. 2016.

[97] A. Tulino and S. Verdu. Random Matrix Theory and Wireless Communications, *Now Publishers*, 2004.

[98] G. Stewart. Matrix Algorithms: Basic Decompositions, *SIAM*, 1998.

[99] Evolved Universal Terrestrial Radio Access (E-UTRA): Physical Channels and Modulation, *3GPP Std. TS 36.211.*, [Online]. Available: http://www.3gpp.org.

[100] J. Cook. Inverse Gamma Distribution, *[Online]:http://www.johndcook.com/ inverse gamma.pdf, Tech. Rep.*, 2008.

[101] A. Osseiran, F. Boccardi, V. Braun, K. Kusume, P. Marsch, M. Maternia, O. Queseth, M. Schellmann, H. Schotten, H. Tullberg, M. Uusitalo, B. Timus, and M. FallgrenView. Scenarios for 5G mobile and wireless communications: the vision of the METIS project, *IEEE Communnications Magzine*, 52(5):26–35, May 2014.

[102] Y. Wang, K. Venugopal, A. Molisch, and R. Heath. MmWave Vehicle-to-infrastructure Communication: Analysis of Urban Microcellular Networks, *arXiv:1702.08122.*

[103] F. Jiang, C. Li, and Z. Gong. Block Gauss-Seidel Method Based Detection in Vehicle-to-Infrastructure Massive MIMO Uplink, in *Proceedings of the IEEE Global Communications Conference*, Singapore, Dec. 2017.

[104] Y. Huo, X. Dong, and W. Xu. 5G Cellular User Equipment: From Theory to Practical Hardware Design, *IEEE Access*, 5:13992–14010, 2017.

[105] X. Gao, L. Dai, Y. Hu, Y. Zhang, and Z. Wang. Low-Complexity Signal Detection for Large-Scale MIMO in Optical Wireless Communications, *IEEE Communications Surveys & Tutorials*, 33(9):1903–1912, 2015.

[106] A. Krishnamoorthy and D. Menon. Matrix Inversion Using Cholesky Decomposition, in *Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*, pages: 70–72, Poznan, Poland, 2013.

[107] D. Bernstein. Matrix Mathematics: Theory, Facts, and Formulas (2nd ed.), *Princeton, NJ, USA: Princeton University Press.*, 2009.

[108] W. Duan, J. Tao, and Y. Zheng. Efficient Adaptive Turbo Equalization for Multiple-Input Multiple-Output Underwater Acoustic Communications, *IEEE Journal of Oceanic Engineering*, DOI: 10.1109/JOE.2017.2707285, 2017.

[109] F. Socheleau, C. Laot, and J. Passerieux. Stochastic Replay of Non-WSSUS Underwater Acoustic Communication Channels Recorded at Sea, *IEEE Transactions on Signal Processing*, 59(10):4838–4849, 2011.

[110] I. Akyildiz, P. Wang, and Z. Sun. Realizing Underwater Communication through Magnetic Induction, *IEEE Communications Magzine*, 53(11):42–48, Nov. 2015.

[111] F. Rusek, D. Persson, B. Lau, E. Larsson, T. Marzetta, O. Edfors, and F. Tufvesson. Scaling Up MIMO: Opportunities and Challenges with Very Large Arrays, *IEEE Signal Processing Magzine*, 30(1):40–60, Jan. 2013.