UNIVERSIDADE LUSÓFONA

CENTRO UNIVERSITÁRIO DE LISBOA

ESCOLA DA COMUNICAÇÃO, ARQUITETURA, ARTES E TECNOLOGIAS DA INFORMAÇÃO

DOUTORAMENTO EM INFORMÁTICA – NOVOS MEDIA E SISTEMAS UBÍQUOS

EFFICIENT SIGNAL DETECTION IN MASSIVE MIMO SYSTEMS USING LOW-COMPLEXITY TECHNIQUES

Tese apresentada a provas públicas para a obtenção do grau de Doutor em Informática – Novos Media e Sistemas Ubíquos orientada pelo Prof. Doutor Rui Dinis Co-orientador Prof. Doutor Marko Beko.

Salah Berra /N° 22110824

2024

www.ulusofona.pt

Salah Berra - 22110824

Efficient Signal Detection in Massive MIMO Systems Using Low-Complexity Techniques

Tese defendida em provas públicas na Universidade Lusófona, Centro Universitário de Lisboa no dia 5/12 2024, despacho de nomeação de Júri nº 1115/2024 de 03 de Dezembro de 2024, para obtenção do grau de Doutor em Informática perante o júri, com a seguinte composição:

Presidente:
Prof. Dr. Paulo Jorge Tavares Guedes, Professor, Universidade Lusófona

Vogais:
Prof. Dr. Luis Bernardo, Full Professor, Universidade Nova de Lisboa
Prof. Dr. João Carlos Marques Silva, Professor, Instituto Universitário de Lisboa
Prof. Dr. Slavisa Tomic, Professor, Universidade Lusófona
Prof. Dr. João Canto, Professor, Universidade Lusófona
Prof. Dr. João Pedro Matos Carvalho, Professor, Universidade Lusófona

Orientador:
Prof. Dr. Rui dinis, Professor, Universidade FCT Nova de Lisboa

Universidade Lusófona – Centro Universitário de Lisboa

Escola de Comunicação, Arquitectura, Artes e Tecnologias da Informação (ECATI)
Lisboa

2024

Efficient Signal Detection in Massive MIMO Systems Using Low-Complexity Techniques

Copyright © Salah Berra, Departamento de Engenharia Informática e Sistemas de Informação, Universidade Lusófona.

The Departamento de Engenharia Informática e Sistemas de Informação and the Universidade Lusófona have the right, perpetual and without geographical boundaries, to file and publish this dissertation through printed copies reproduced on paper or on digital form, or by any other means known or that may be invented, and to disseminate through scientific repositories and admit its copying and distribution for non-commercial, educational or research purposes, as long as credit is given to the author and editor.

Acknowledgements

This doctoral thesis was conducted at COPELABS and the Department of Computer Engineering and Information Systems at Lusofona University, located in Lisbon, Portugal, from November 2021 to July 2024. The Fundação para a Ciência e Tecnologia supported the research under project UIDB/04111/2020.

I am thrilled to have this opportunity to express my gratitude to all whose kindness and support made this thesis possible.

First, I would like to thank my supervisor, Prof. Dr. Rui Dinis and Prof. Dr. Marko Beko, for their support, for giving me the opportunity to pursue my Ph.D. thesis on such interesting research topics, and for providing a very stimulating research environment at the Copelabs, Lusófona University, and the Telecommunications Laboratory at FCT NOVA University.

I am also pleased to thank Dr. Sourav Chakraborty, who kindly helped me in many aspects of my Ph.D. thesis and acted as an external co-examiner.

Lastly, I would like to thank my family for all their love and encouragement, especially my parents, who raised me with love and supported me in all my pursuits.

Abstract

In next-generation wireless communication systems such as 5G and 6G, massive multiple-input multiple-output (MIMO) is already established as a key technology due to its ability to enhance capacity and efficiency. Despite these advantages, massive MIMO suffers from high complexity on the detector side. In this thesis, we explore and propose advanced iterative detection algorithms for massive MIMO systems, focusing on efficiency, low complexity, and enhanced performance. The research is divided into chapters, each addressing different aspects of iterative massive MIMO detection.

Firstly, we introduce an efficient iterative detection algorithm based on the Modified Accelerated Over-Relaxation (MAOR) method. We derive the optimal values for the acceleration and relaxation parameters that control the convergence speed. Additionally, we apply an initial solution based on eigenvalues instead of the common diagonal matrix approximation, improving the convergence rate. By incorporating Chebyshev polynomial acceleration, we achieve near-minimum Mean Squared Error (MMSE) performance with reduced complexity, outperforming existing massive MIMO detectors.

Next, we extend the application of Chebyshev acceleration to various low-complexity detection methods for massive MIMO schemes. Our proposed technique demonstrates superior error rate performance and reduced computational complexity compared to state-of-the-art methods.

Furthermore, we develop two iterative detectors, AC-AOR and AC-SOR, enhanced through the accelerated Chebyshev method, and propose Deep Unfolding Networks (DUNs) for optimizing algorithm parameters via training. The DUNs show significant performance improvements and scalable, fast training processes, outperforming conventional detectors across various channel scenarios, including non-ideal conditions.

Additionally, we consider a massive MIMO system with single carrier frequency domain equalization (SC-FDE) modulations employing Iterative Block Decision Feedback Equalization (IB-DFE) receivers. We introduce a novel IB-DFE receiver structure compatible with existing iterative methods to avoid matrix inversion, incorporating a Chebyshev accelerated Richardson iterative method. Our complexity analysis indicates substantial computational savings, with simulation results showing BER performance close to that of exact receivers, even under imperfect channel state information.

In general, our proposed methods demonstrate significant advancements in iterative detection for massive MIMO systems, offering improved performance, reduced complexity, and robustness under various channel conditions.

Keyword

Massive MIMO, Signal detection, Iterative methods, SC-FDE, IB-DFE, Low-complexity detection, deep learning.

Resumo

As técnicas massive MIMO (Mulit-Input, Multi-Output) serão fundamentais nas comunicações sem fio de próxima geração, como 5G e 6G, pois permitem aumentar a capacidade e as eficiências espectrais e de potência. Contudo os sistemas massive MIMO requerem alta complexidade de implementação, nomeadamente devido aos requisitos de processamento de sinal para a receção. Nesta tese exploramos e propomos algoritmos de deteção iterativa avançados para sistemas massive MIMO, com foco na eficiência de potência e redução da complexidade de implementação.

Começamos por introduzir um algoritmo de detecção iterativa eficiente baseado no método de Aceleração por Sobrerrelaxação Modificada (MAOR), sendo obtidos os valores ótimos para os parâmetros de aceleração e relaxamento que controlam a velocidade de convergência. Além disso, aplicamos uma solução inicial baseada em autovalores, em vez da comum aproximação por matriz diagonal, melhorando a taxa de convergência. Ao incorporar a aceleração por polinómios de Chebyshev, conseguimos um desempenho próximo do obtido por recetores MMSE (Minimum Mean-Squared Error) com complexidade muito mais reduzida, superando os recetores massive MIMO existentes.

De seguida, estendemos o uso da aceleração de Chebyshev a vários métodos de deteção de baixa complexidade para esquemas massive MIMO. Quando comparadas com os métodos mais avançados atuais, as técnicas propostas apresentam um melhor desempenho em termos de probabilidade de erro, em conjunto com menor complexidade computacional.

Além disso, desenvolvemos dois detetores iterativos, AC-AOR (....) e AC-SOR (...) que empregam o método acelerado de Chebyshev, e propomos redes DUN (Deep Unfolding Networks) para otimização dos parâmetros dos algoritmos propostos. As DUNs mostram melhorias significativas de desempenho e os processos de treino rápido e escalável superam os recetores convencionais em múltiplos canais, incluindo em condições não ideais.

Adicionalmente, consideramos um sistema massive MIMO com técnicas SC-FDE (Single-Carrier with Frequency Domain Equalizaiotn) empregando recetores IB-DFE (Iterative Block Decision-Feedback Equalization). Introduziu-se uma nova estrutura de recetor IB-DFE compatível com os métodos iterativos existentes para evitar a inversão de matrizes, incorporando um método iterativo de Richardson acelerado por Chebyshev. Estas técnicas têm uma complexidade substancialmente reduzida quando comparadas com técnicas MMSE, apresentando desempenho similar, mesmo sob condições imperfeitas de informação do estado do canal.

De forma geral, nossos métodos propostos demonstram avanços significativos na detecção iterativa para sistemas massive MIMO, oferecendo melhor desempenho, menor complexidade e robustez em várias condições de canal.

Palavras-chave: Massive MIMO, Detecção de sinais, Métodos iterativos, SC-FDE, IB-DFE, Detecção de baixa complexidade, aprendizagem profunda.

Contents

Appendices

List of Figures

List of Tables

Acronyms

List of Notations

$(\cdot)^{-1}$ Matrix inversion

$(\cdot)^{H}$ Matrix Hermitian (conjugate transpose)

$(\cdot)^{T}$ Matrix transpose

$(\mathbf{x}^{(i)})_{i=0}^{\infty}$ The sequence

$[\mathbf{X}]_{n,k}$ (n, k)-th element of matrix $\mathbf{X}$

$\lambda(\mathbf{X})$ Eigenvalues of matrix $\mathbf{X}$

$\lambda_{\max}(\mathbf{X})$ Largest eigenvalue of matrix $\mathbf{X}$

$\lambda_{\min}(\mathbf{X})$ Smallest eigenvalue of matrix $\mathbf{X}$

$\lambda_i(\mathbf{X})$ i-th eigenvalue of matrix $\mathbf{X}$

$\mathbb{A}, \mathfrak{S}$ Set of M complex QAM symbols

$\mathbb{C}^{R\times T}$ Complex matrix with R rows and T columns

$\mathbb{E}[\cdot]$ Expectation operator

$\mathbb{R}^{2R\times 2T}$ Real-valued matrix with $2R$ rows and $2T$ columns

$\mathbf{0}_{N\times M}$ $N \times M$ matrix with all elements equal to 0

$\mathbf{1}_{N\times M}$ $N \times M$ matrix with all elements equal to 1

$\mathbf{D}$ Diagonal component of matrix

$\mathbf{I}_N$ Identity matrix of size $N \times N$

$\mathbf{L}$ Lower component of the triangle matrix

$\mathbf{P}^i$ Polynomial of degree i

$\mathbf{R}_{RX}$ Receive correlation matrix

$\mathbf{R}_{TX}$ Transmit correlation matrix

$\mathbf{U}$ Stringently upper of the triangle matrix

$\mathbf{X}$ Uppercase boldface Roman letter matrix

$\mathbf{x}$ Lowercase boldface Roman letter matrix

$\mathbf{x}^{(i+1)}$ Vector $\mathbf{x}$ at iteration i

$\mathbf{X}_k$ Frequency-domain block

$\mathbf{x}_n$ Time-domain block

$C(.)$ Ergodic capacity

$\mathcal{N}(0, \sigma^2)$ Gaussian distribution with mean 0 and variance σ^2

$O()$ Big-O notation

$\rho(\mathbf{X})$ Spectral radius of matrix $\mathbf{X}$

σ^2 Noise variance in complex received signal

$\mathrm{Tr}(.)$ Matrix trace

ξ The ratio between receiver antenna and users.

ζ The correlation parameter between two adjacent antennas

$diag(.)$ Matrix diagonal

List of Abbreviations

O Constellation symbols

3GPP 3rd Generation Partnership Project

6G Sixth Generation

$1G$ The first generation

$2G$ The second generation

$3G$ The third generation

$4G$ The fourth generation

$5G$ The fifth generation

BER Bit error rate

GSM The global system for mobile communications

LTE The long-term evolution

$MIMO$ Multiple-input multiple-output

$MISO$ Multiple-input single-output

NR New radio

$SIMO$ Single-input multiple-output

$SISO$ Single-input single-output

WCN wireless communication network

AC-AOR Accelerated Chebyshev AOR

AC-AORNet Accelerated Chebyshev AOR Network

AC-SOR Accelerated Chebyshev SOR

AC-SORNet Accelerated Chebyshev SOR Network

ADAM Adaptive Moment Estimation

AI Artificial Intelligence

AMP Approximate Message Passing

AOR Accelerated Overrelaxation

AORNet Accelerated Over-Relaxation Network

AWGN Additive White Gaussian Noise

BP Belief Propagation

BS Base Station

CAGR Mean annual growth rate of an investment over a period longer than one year

CG Conjugate Gradient

cmWave Centimeter Wave

CP Cyclic prefix

CR Cognitive Radio

CSI Channel State Information

D2D Device-to-Device

DetNet Detection Network

DFT Discrete Fourier Transform

DL Deep Learning

DNNs Deep Neural Networks

DUN Deep Unfolding Network

EGC Equal Gain Combining

EGD Equal Gain Detector

eNB evolved NodeB

EPNet Expectation Propagation Network

FFT Fast Fourier Transform

FLOPs Floating-Point Operations

FWA Fixed Wireless Access

GS Gauss-Seidel

HSPA+ High-Speed Packet Access Plus

IB-DFE Iterative Block Decision Feedback Equalizer

IDFT Inverse Discrete Fourier Transform

IEEE Institute of Electrical and Electronics Engineers

IFFT Inverse Fast Fourier Transform

IIC Iterative Interference Cancellation

IoT Internet of Things

ISI Inter-symbol interference

JA Jacobi

LLR Log-Likelihood Ratio

lS-MIMO large-scale MIMO

MAOR Modified Accelerated Overrelaxation

MCM Multi-Carrier Modulation

MEP Modified Expectation Propagation

MEPNet Modified Expectation Propagation Network

MF Matched Filter

MFB Matched Filter Bound

ML Maximum Likelihood

MMSE Minimum mean square error

mmWave Millimeter Wave

MPD Message Passing-based Detection

MRC Maximum-Ratio Combining

MRD Maximum Ratio Detector

MSE Mean Squared Error

MsNet Multi-Segment Mapping Network

NFV Network Functions Virtualization

NI Newton Iteration

NS Neumann Series

OAMPNet Orthogonal approximate message passing network

OFDM Orthogonal Frequency Division Multiplexing

OFDMA Orthogonal Frequency Division Multiple Access

OTFS Orthogonal Time Frequency Space

PAM Pulse Amplitude Modulation

PAPR Peak-to-Average Power Ratio

PHY Physical Layer

PIC Parallel Interference Cancellation

PSK Phase-Shift Keying

QAM Quadrature Amplitude Modulation

QoS Quality of Service

RI Richardson Iteration

RxD Receive Diversity

SC-FDE Single-Carrier Frequency-Domain Equalization

SC-FDMA Single-carrier frequency division multiple access

ScNet Sparsely Connected Neural Network

SD Sphere Decoding

SDMA Space Division Multiple Access

SDN Software-Defined Network

SDR Semidefinite Relaxation

SE Spectral Efficiency

SIC Successive Interference Cancellation

SINR Signal-to-Interference-and-Noise Ratio

SNR Signal-to-Noise Ratio

SOR Successive Overrelaxation

SORNet Successive Over-Relaxation Network

SPD Symmetric Positive Definite

STC Space-Time Coding

STD Steepest Descent

SVD Singular Value Decomposition

TPE Truncated Polynomial Expansion

TPG-detector Trainable Projected Gradient-Detector

TxD Transmit Diversity

UDN Ultra-Dense Network

UL Uplink

UTs the user terminals

VLC Visible Light Communication

Wi-Fi Wireless Fidelity

WiMAX Worldwide Interoperability for Microwave Access

WLAN Wireless Local Area Network

ZF Zero-Forcing

Chapter

1 Introduction

The field of wireless communications is undoubtedly one of the most significant technological advances in history and a pinnacle of modern engineering. Its influence on the operations of today's society is immense. Imagining the current world without the countless wireless-connected devices is nearly inconceivable.

During the decades since their inception, cellular systems have experienced five generational transitions. The first generation (1G) emerged in the 1980s, providing only analog telephony. The second generation (2G) launched in the 1990s with the global system for mobile communications (GSM), as its dominant standard, transitioned to digital and introduced text messaging. The third generation (3G), around the 2000s, included data and multimedia applications. Following this, the fourth generation (4G) not only delivered faster bit rates and lower latencies but also fully adopted packet switching as a platform for the mobile Internet, unifying all standards worldwide into the Long-Term Evolution (LTE) system. LTE subsequently evolved into the fifth generation (5G), enhanced by an additional radio access interface known as New Radio (NR), which enables operation across a much broader range of frequencies [1]. The sixth generation 6G wireless communication network (WCN) is envisioned to be expansive, targeting ambitious future goals. It aims to provide intelligently enabled seamless connectivity everywhere, with reduced energy consumption to enhance society and living overall. The increasing demands for connecting devices and data traffic are the fundamental driving forces behind the evolution of 6G. Key motivating and expected trends in 6G include ultra-high data rates of 1 Tbps, extremely low latency (one-tenth of that in 5G), data throughput 50 times faster than 5G, as well as double the energy efficiency and spectrum efficiency [2].

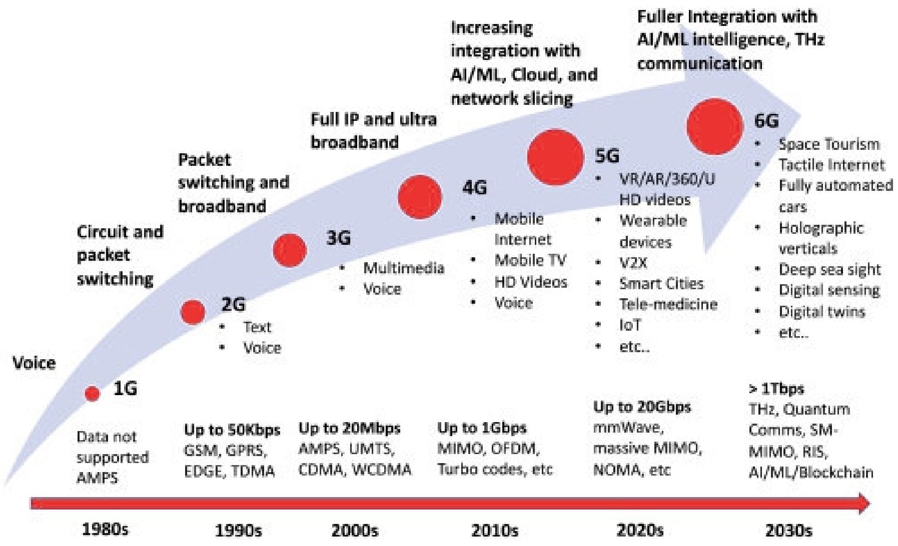


Figure 1.1: The evolution of wireless communication [3].

1.1 Multi-antenna systems

The wireless data transmission can be done through various antenna configurations. In general, based on the number of antennas on the transmitter and receiver side, we can classify wireless systems into four categories [4]: Single-input single-output (SISO), single-input multiple-output (SIMO), multiple-input single-output (MISO), and multiple-input multiple-output (MIMO).

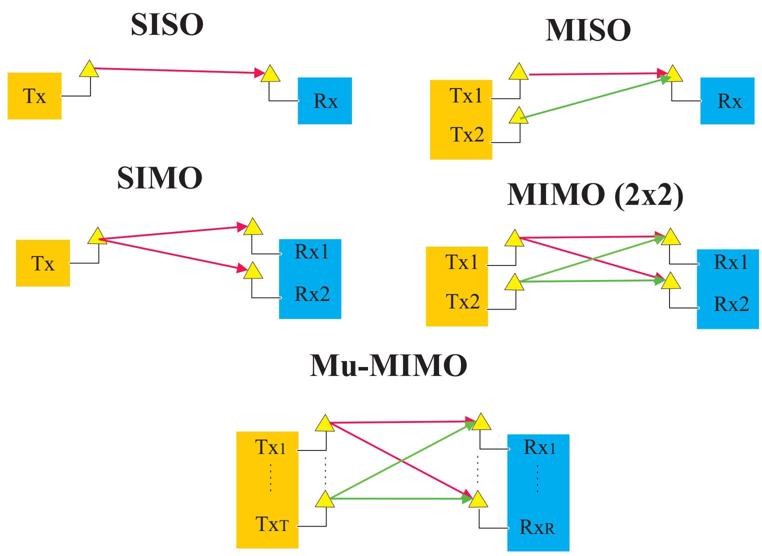


Figure 1.2: Different Multi-antenna configurations.

Fig. 1.2 illustrates different antenna configurations used in defining space-time systems. SISO used a single transmitter with single receiver system, SIMO uses a single transmitting antenna and multiple (R) receive antennas, MISO has multiple (T) transmitting antennas and one receive antenna, MIMO has multiple (T) transmitting antennas and multiple (R) receive antennas, and finally, multiuser-MIMO (MU-MIMO), which refers to a configuration that comprises a base station with multiple transmit/receive antennas interacting with multiple users, each with one or more antennas. The table 1.1 below provides the advantages and limitations of the configurations of multiple antenna systems, as considered from this reference [5].

Table 1.1 The advantages and disadvantages of various cases of multiple antenna systems

System Type	Advantages	Limitations
SISO	- Simple design - Cheaper than other systems	- Limited diversity gain, leading to poorer performance in terms of signal reliability - Lower bit error rate (BER) performance compared to more advanced systems like SIMO, MISO, and MIMO
SIMO	- Improved diversity - Better BER performance than SISO	- Increased complexity in receiver design - Higher cost and power consumption compared to SISO systems
MISO	- Overcomes issues with multiple signals arriving at different times - Shifts coding redundancy to the transmitter - it's easier to have many multiple antennas at the BS than the user terminals (UTs)	- Increased complexity and cost at the transmitter side - Performance depends on the receiver's position and movement
MIMO	- Best throughput and efficiency in signal transmission - Uses multiple antennas at both ends for improved performance - Increased diversity, spacial multiplexing, multi-user detection	- Significantly higher design complexity and cost - Increased power consumption - Higher hardware requirements, less suitable for low-power applications

MIMO technology has garnered interest in wireless communications due to its ability to significantly enhance data throughput and link range without needing extra bandwidth or transmit power. This is accomplished by offering higher spectral efficiency (more bits

per second per Hertz of bandwidth) and better link reliability or diversity (less fading) [6]. These characteristics make MIMO a crucial component of contemporary wireless communication standards like IEEE 802.11ax (Wi-Fi 6) [7].
In MIMO, the system configuration typically contains T antennas at the transmitter and R antennas at the receiver front end as illustrated in the following Fig.1.3.

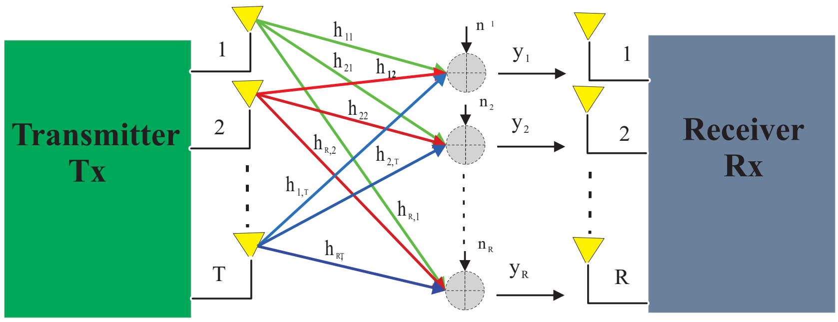


Figure 1.3: Example of 4×4 MIMO system.

Here, each receiver antenna receives not only the direct signal intended for it, but also receives a fraction of signal from other propagation paths.

The received vector $\mathbf{y}$ is expressed in terms of the channel transmission matrix $\mathbf{H}$, the input vector $\mathbf{x}$ and noise vector $\mathbf{n}$ as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{1.1}$$

and the matrix form is given by

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_R \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_T \end{bmatrix}, \quad \mathbf{H} = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1T} \\ h_{21} & h_{22} & \cdots & h_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ h_{R1} & h_{R2} & \cdots & h_{RT} \end{bmatrix}, \quad \mathbf{n} = \begin{bmatrix} n_1 \\ n_2 \\ \vdots \\ n_R \end{bmatrix} \tag{1.2}$$

MIMO represents one of the various types of smart antenna technologies. Among the four types, MIMO systems provide the highest capacity and offer a wide range of applications. In particular, MIMO signals can be transmitted using spatial diversity, enhancing performance and reliability [8].

1.2 Diversity and multiplexing

Multiple antenna channels offer spatial diversity to enhance link reliability by delivering the receiver with several independently faded versions of the same information symbol, thereby lowering the likelihood of simultaneous fading [9]. Multipath fading poses a major challenge in communications. In a fading channel, the signal strengths fluctuate, leading to periods in which the signal power significantly drops, known as fades. These fades result in high BER. To mitigate fading, we use diversity, which involves transmitting multiple replicas of the signal at different times, frequencies, or spatial paths. Wireless communications employ three main types of diversity schemes.

Space diversity is achieved by deploying multiple antennas at the transmitter and/or receiver [10]. When multiple antennas are positioned at the transmitter, a technique known as space-time coding (STC) [11] is typically employed to obtain diversity gain, referred to as transmit diversity. Conversely, if multiple antennas are installed at the receiver, diversity gain can be realized through combining techniques [12] such as maximum-ratio combining (MRC), equal gain combining or selection combining. These techniques are known as receive diversity. A SIMO system utilizes multiple receive antennas to achieve receive diversity, whereas a MISO system employs multiple transmit antennas to achieve transmit diversity.

1.2.1 Receiver diversity

To attain receive diversity, signals received from multiple antennas must be appropriately combined using a suitable combining method [13].

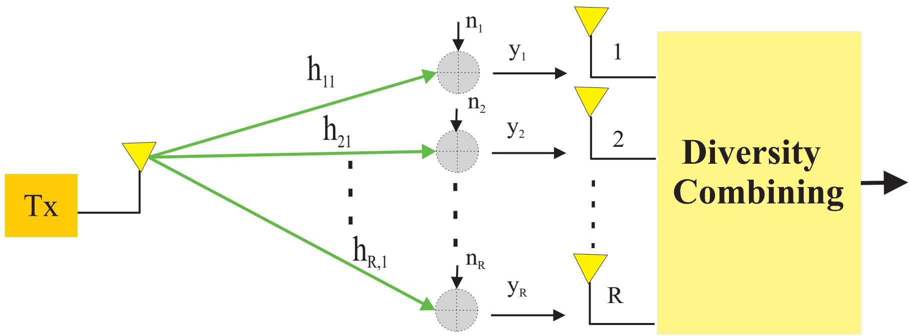


Figure 1.4: Spatial diversity provided by using multiple receive antennas.

At the receiver, MRC is a technique applied in diversity schemes to improve signal quality. This method adjusts the received signal at each antenna based on the signal strength and synchronizes the phases of the signals before combining them to enhance the output signal-to-noise ratio (SNR)[14].

1.2.2 Transmit diversity

In MISO system, transmit-diversity is achieved by sending signal copies over T channels, connecting T transmit antennas to a single receive antenna. This setup is more appropriate for scenarios such as the downlink in a mobile cellular network, where the base station can have multiple antennas, while the mobile device (due to size constraints) uses only one antenna. Achieving diversity in an MISO system is more challenging than in a SIMO system because the signals from the multiple transmit antennas are combined at the single receive antenna.

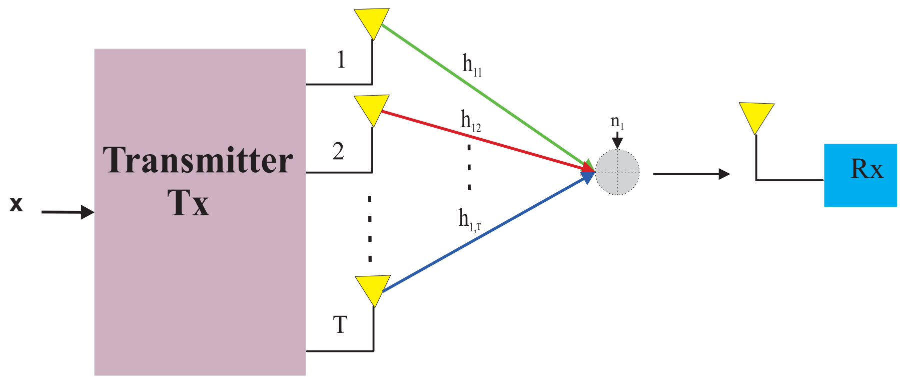


Figure 1.5: Spatial diversity provided by using multiple transmit antennas.

A widely recognized method for achieving transmit diversity is STC, with the Alamouti scheme being the most renowned example. Alamouti offers an exceptionally straightforward approach to attain transmit diversity using two transmit antennas without sacrificing bandwidth efficiency [15].

1.2.3 Gain diversity

Another diversity called the gain diversity, which is described as the SNR exponent of the error probability, which corresponds to the number of independently faded paths that a symbol traverses. For example, consider uncoded binary phase-shift keying (PSK) signals in a single-antenna fading channel (SISO model). It is well established [16] that the error probability at high SNR (averaged over the fading gain **H** and the additive noise) is given by

$$\mathbf{P}_e = \frac{1}{4}SNR^{-1}. \tag{1.3}$$

When the same signal is transmitted to a receiver with two antennas, the error probability is given by

$$\mathbf{P}_e = \frac{3}{16}SNR^{-2}. \tag{1.4}$$

Hence, the BER analysis of a Rayleigh fading channel is conducted by

$$\text{BER} = \mathbf{Q}\left(\sqrt{\mathbf{P}_e}\right) \tag{1.5}$$

Here, we notice that with the additional receive antenna, the error probability decreases with SNR at a rate of SNR^{-2}. Similar outcomes can be achieved if we switch the binary

PSK signals to other constellations. Since the performance improvement at high SNR is determined by the SNR exponent of the error probability, this exponent is referred to as the diversity gain. Essentially, it represents the number of independently faded paths a symbol traverses; in other words, the number of independent fading coefficients that can be averaged to detect the symbol. In a general system with T transmit and R receive antennas, there are $R \times T$ random fading coefficients to be averaged; thus, the maximum (full) diversity gain offered by the channel is RT.

1.2.4 Spatial multiplexing

Besides providing diversity to improve reliability, multiple-antenna channels can also support a higher data rate than single-antenna channels. Spatial multiplexing provides a linear increase in transmission rate (or capacity) proportional to the number of transmit-receive antenna pairs or min (R, T), for the same bandwidth and without extra power consumption. This is feasible in MIMO channels [10, 17]. As evidence of this, consider an ergodic block-fading channel where each block follows the model in (1.2) and the channel matrix is i.i.d. across blocks. The ergodic capacity (b/s/Hz) of such a channel is well documented in [18], [19] as

$$C(SNR) = \mathbb{E}\left[\log\det(\mathbf{I} + \frac{SNR}{R}\mathbf{H}\mathbf{H}^{H})\right]. \tag{1.6}$$

At high SNR

$$C(SNR) = \min\{R,T\} \log\frac{SNR}{r} + \sum_{i=|R-T|+1}^{\max\{R,T\}} \mathbb{E}\big[\log \mathcal{X}_{2i}^{2}\big] + o(1), \tag{1.7}$$

where $\mathcal{X}_{2i}^{2}$ follows a chi-square distribution with $2i$ degrees of freedom. It is observed that at high SNR, the channel capacity scales with SNR as $\min\{R,T\}\log SNR$ (b/s/Hz), unlike the $\log SNR$ for single-antenna channels. This indicates that the multiple-antenna channel can be considered as $\min\{R,T\}$ parallel spatial channels; thus, $\min\{R,T\}$ represents the total number of communication degrees of freedom. Consequently, independent information symbols can be transmitted in parallel through these spatial channels.

Spatial multiplexing in multiple-antenna channels enables higher data rates by transmitting independent information symbols in parallel through spatial channels. The spatial multiplexing gain refers to the number of independent data streams that can be transmitted simultaneously, resulting in a data rate increase proportional to the number of antennas.

1.3 Multi-carrier and SC-FDE MIMO system

Recently, significant attention has been directed towards physical layer (PHY) techniques suitable for high data rate MIMO wireless communication systems. Particularly noteworthy are techniques aimed at reducing the complexity of the multi-dimensional equalization process required in wideband MIMO systems. Two such techniques has considered: Orthogonal frequency division multiplexing (OFDM) and single-carrier frequency-domain equalization (SC-FDE) methods. In single-antenna systems, OFDM is well-established and widely accepted in both academic and industrial circles [20]. Meanwhile, SC-FDE, a PHY technique applies signal processing in the frequency domain to offer relatively low-complexity solutions to equalization challenges [21]. In [21], a comparison of OFDM and SC-FDE technologies is presented, focusing on single-antenna systems used in fixed broadband wireless applications (IEEE 802.16), where the channel multipath spread consists of only a few non-zero discrete taps. However, a performance comparison of OFDM and SC-FDE for MIMO systems used in wireless local area networks (WLANs), where the effects of multipath propagation can be significantly greater, has not yet been carried out [22].

1.4 From MIMO to massive MIMO systems

MIMO is now widely regarded as a significant technology in (4G) wireless communication networks [23].The development of MIMO wireless communication technology originated from the idea of point-to-point MIMO communication where the two communicating devices have multiple antennas and the idea of multi-user MIMO (MU-MIMO) communication where a base station (BS) equipped with multiple antennas serves end users of single antennas [24]. Multi-User MIMO (MU-MIMO) occurs when an evolved NodeB (eNB) with multiple antennas connects to numerous UTs using the same time-frequency resources [25]. MU-MIMO can improve BE or reliability by increasing multiplexing or diversity gains [26]. In addition, Marzetta proposed the large-scale MIMO concept in [27], which is also known as the massive MIMO scheme and is frequently associated with the terminologies of large-scale antenna systems, very large MIMO, very large multi-user MIMO, full-dimensional MIMO, hyper-MIMO, and others. To be more specific, a large-scale MIMO system is one that employs hundreds of antennas to service dozens of UEs at the same time [28]. Both theoretical and experimental data show that an LS-MIMO can greatly improve BE while simultaneously lowering transmit power [29], [30]. As a result, LS-MIMO is considered a suitable approach for next-generation wireless communication systems designed to improve both bandwidth efficiency and energy efficiency. However, a conventional MU-MIMO link suffers from inter-user interference, which considerably limits the performance of the system. In Fig. 1.6, the different cases of multiple antenna to massive MIMO systems are presented.

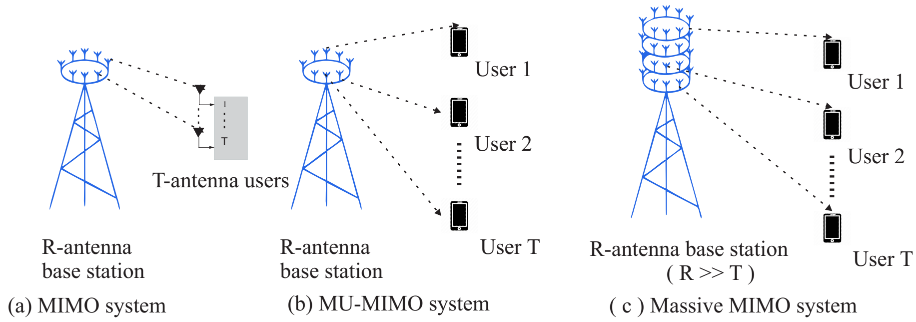


Figure 1.6: Multiple antenna technology: From MIMO to massive MIMO.

1.5 Massive MIMO systems

The increasing number of mobile users demands faster Internet access, instant multimedia services, and efficient data exchange in smart cities. As depicted in Fig. 1.7, the growth rate of mobile network data traffic is expected to continue declining, from 21 percent in 2024 to 16 percent in 2030. This represents a CAGR of 19 percent over the full forecast period [31].

Total global mobile data traffic excluding traffic generated by Fixed Wireless Access (FWA) is expected to grow by a factor of approximately 2.5, reaching 303 EB per month by 2030. When FWA is included, total mobile network data traffic is anticipated to grow by a factor of around three, rising to 473 EB per month by the end of the forecast period [31].

The share of 5G in mobile data traffic is projected to reach an estimated 34 percent by the end of 2024, up from 25 percent at the end of 2023. This share is forecast to grow to 80 percent by 2030.

Additionally, several large regions, such as India, Latin America, Southeast Asia, and Africa, are expected to significantly transition their subscriber bases to newer generations of mobile technologies in the coming years. Globally, an addition of 1.3 billion mobile broadband subscriptions is anticipated during the forecast period. Future traffic patterns of these users will depend on network capabilities, tariff plans, and available services. This trend is accelerating the increase in the average number of devices and connections per household and per capita. Each year, various new devices are introduced and adopted in the market in different form factors with enhanced capabilities and intelligence [31].

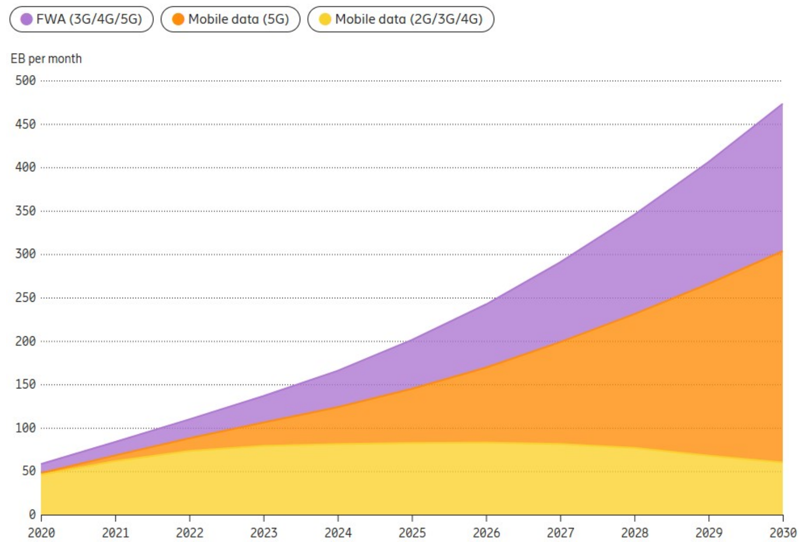


Figure 1.7: Global mobile network data traffic [31].

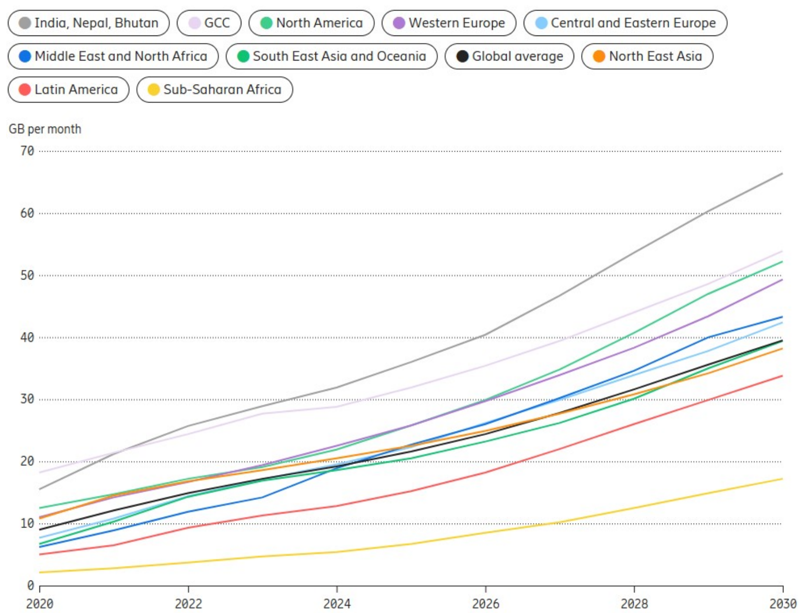


Figure 1.8: Mobile data traffic per active smartphone [31].

To achieve this, technology has advanced in each generation of communication architecture, in order to improve network capacity, efficiency, and reliability.As one of the most promising wireless communication solutions, massive MIMO technology has been widely adopted in next-generation 5G networks [32]. The growing number of users and the increasing demand for high data rates have forced researchers to enhance the capacity and spectral efficiency of next-generation wireless communications. The anticipated data rate for (5G) cellular communication is projected to reach 10 Gbps [33]. Various technologies have been proposed to increase 5G capacity, including small cell communications, millimeter wave communications, and massive MIMO technology [34].

Small cell communication improves spectral efficiency by amplifying frequency reuse capabilities, employing small cells such as femto cells and pico cells [35]. In contrast, millimeter-wave communication allocates a broad frequency band ranging from 30 GHz to 300 GHz to increase the capacity of 5G [36]. Massive MIMO technology deploys an extensive array of antennas in BS, increasing spatial resolution and consequently enhancing the spectral and power efficiency of communication systems [37]. Each of these technologies has the potential to significantly increase the capacity of wireless communication systems independently, and their synergistic application can further increase the capacity of the system [34]. However, these innovative technologies face challenges that must be addressed to achieve efficient implementation.

Researchers advocate for 5G and innovative technologies such as device-to-device (D2D) communication [38], ultra-dense networks (UDNs) [39], spectrum sharing, centimeter-wave (cmWave) or millimeter-wave (mmWave) [40], the Internet of Things (IoT) [41], and massive multiple-input multiple-output (MIMO) [42] to meet these demands.

Massive MIMO systems are the most exciting wireless access technology to meet the demands of 5G and beyond networks. Massive MIMO is an expansion of MIMO technology that uses hundreds or even thousands of antennas connected to a base station to improve spectral efficiency and throughput [24]. This technology combines antennas, radios, and airwaves to increase capacity and speed for the next 5G network [30].

Recently, massive MIMO systems are promising key technologies for 5G and beyond cellular networks [43] due to their improved coverage and at least 3 times higher spectral efficiency than MIMO techniques used in 4th generation networks. Occasionally, a BS may have a large number of antennas (e.g., 128, 256, etc.) that serve multiple user antennas (e.g., 16,32, etc.) in the same time-frequency period. Massive MIMO systems have shown notable improvements in spectrum utilization, data rates, reliability, and overall data provision [32]. This technology is well suited for next-generation and beyond communication in the 5G era [44]. Massive MIMO has been selected for deployment in cell-free architectures, which signifies its versatility and potential for future wireless communication systems [45]. Classical massive MIMO technology has found its application in 5G communication systems operating below 6 GHz, where radio channels exhibit rich scattering and multipath propagation [46]. However, as we venture into higher carrier frequencies, such as the cmWave or mmWave bands, and beyond, extending to the THz band, very large antenna arrays become essential. At these elevated frequencies, the propagation channels become more directive, creating distinct interference conditions. Consequently, the term "massive MIMO" has not traditionally been universally applied to these communication concepts; its usage varies from one paper to another. Large arrays are more manageable to implement and pack in higher frequencies due to the smaller size of the antennas. Despite this, the propagation characteristics of the channels in cmWave or mmWave systems differ, presenting unique challenges in the multi-user interference scenario.

Why massive MIMO?

Massive MIMO technology improves spatial multiplexing and diversity by increasing the number of antennas at the base station. This allows for easy signal processing across all antennas [47, 48]. The possible benefits of massive MIMO are listed below:

- Massive MIMO improves the stability of the connection by increasing diversity and preventing fading [49]. Using multicell minimal mean square error (MMSE) precoding/combining and spatial channel correlation, the number of antennas may enhance capacity without limit, even with pilot contamination [50].

- Massive MIMO improves spectral efficiency (SE) in cellular networks by spatially multiplexing several user devices per cell [27]. Multiple antennas increase spatial data streams, throughput, and multiplexing gain, resulting in excellent spectral efficiency [51]. Massive MIMO may achieve 10 times greater spectral efficiency compared to traditional MIMO, which serves many customers at the same time using the same resources [29].

- Coherent combining results in an inverse relationship between transmitted power and the number of antennas [52]. As the number of transmit antennas increases, the transmit power decreases significantly.

- Improving security and robustness is crucial for current wireless communication systems, as they are vulnerable to interference and jamming. The enormous number of antenna terminals [29] creates many degrees of freedom, which can be utilized to cancel out signals from purposeful jammers [53].

The massive MIMO capable of improving throughput and spectrum efficiency has made it a critical technology for the development of wireless standards [29]. The key here is the significant gain in the array achieved by massive MIMO using a large number of antennas [54]. Massive MIMO is a critical enabling technology for 5G and beyond networks. Because intelligent sensing systems rely heavily on 5G and beyond networks for operation, massive MIMO systems integrated with intelligent sensing systems are inextricably intertwined [55]. In Fig. 1.9, a cell-free integrated sensing and communication (ISAC) MIMO system is illustrated, where distributed MIMO access points jointly serve the communication users and sense the targets.

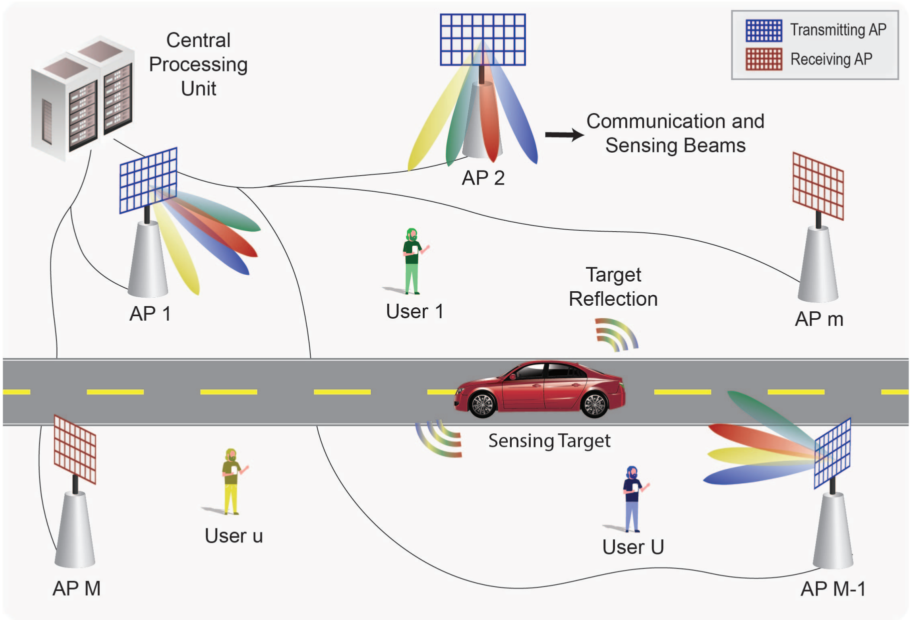


Figure 1.9: Integrated sensing and communication in cell-free massive MIMO systems [55].

More specifically, the benefits and advantages of the massive MIMO system are listed in Table 1.2. Data gathering from a large number of smart sensors using typical multi-access systems is difficult due to high latency, low data rate, and poor dependability. Massive MIMO with high multiplexing gain and beamforming capabilities can detect data from concurrent sensor transmissions with considerably shorter latency, allowing sensors to communicate at higher rates and with greater reliability [65].

Massive MIMO systems will play an important role in allowing the information collected by smart sensors to be transmitted in real time to central monitoring locations for smart sensor applications such as autonomous vehicles, remote healthcare, smart grids, smart antennas, smart highways, smart buildings, and intelligent environmental monitoring [66].

Despite the above advantages of the massive MIMO system, it still faces practical implementation challenges, notably in signal detection, channel estimation, precoding, hardware imprimante, etc., due to elevated multi-user interferences. Some of the fundamental challenges in massive MIMO systems are shown in Fig. 1.10.

Table 1.2 Benefits of Massive MIMO for 5G Networks and Beyond

Benefit	**Description**
Spectral Efficiency	Massive MIMO enhances spectral efficiency by directing narrow beams toward users using its antenna array. This technology can deliver a spectral efficiency that is more than ten times higher than that of the current MIMO systems employed in 4G/LTE [56].
Energy Efficiency	Due to the antenna array is concentrated in a narrowly defined area, it demands less radiated power, thereby lowering the energy consumption in massive MIMO systems[56].
High Data Rate	The data rate and capacity of wireless systems are enhanced by the array gain and spatial multiplexing offered by massive MIMO[57].
User Tracking	As massive MIMO employs narrow signal beams directed at the user, tracking the user becomes more precise and reliable [58].
Low Power Consumption	By using beamforming, the BS is able to focus its transmission power on each user, reducing the transmission power required to serve each user. This results in lower power consumption and increased energy efficiency for the network [59].
Less Fading	With numerous antennas in the receiver, massive MIMO demonstrates robustness against fading [60].
Low Latency	Massive MIMO reduces the latency on the air interface [61].
Robustness	Massive MIMO systems exhibit robustness against unintentional interference and internal jamming. Furthermore, these systems can withstand some antenna failures due to the presence of numerous antennas [62].
Reliability	The use of numerous antennas in massive MIMO systems enhance diversity gain, thereby improving link reliability [27, 63].
Enhanced Security	Massive MIMO enhances physical security through the use of orthogonal channels for mobile stations and focused beams [64].
Signal processing	channel hardening simplifies the signal processing and resource allocation at the base station.

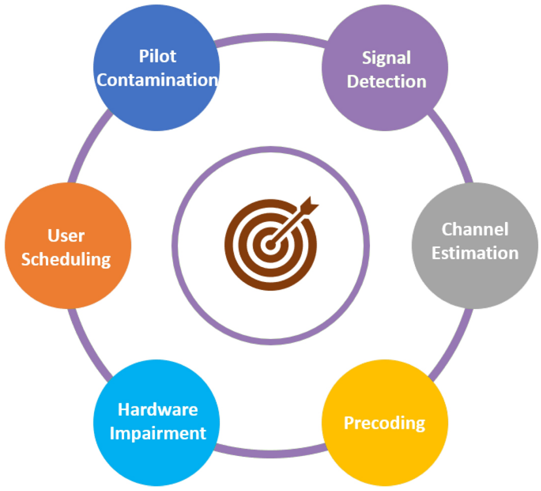


Figure 1.10: The challenges of massive MIMO deployment.

1.6 Objectives

The objective of this thesis is to design a receiver/detector for a massive MIMO system that reduces computational complexity compared to state-of-the-art detectors. Furthermore, the thesis explores receiver test scenarios under various channel types and conditions.

1.7 Thesis organization and contribution

In this work, low-complexity detectors are presented to address the issues related to massive MIMO systems, detailed in the following chapters:

- **Chapter 2**: This chapter introduces the massive MIMO system model and its challenges. It examines conventional detectors and recent advancements, emphasizing detection issues as the number of antennas increases. The chapter explores low-complexity detection schemes for massive MIMO and proposes an acceleration technique based on iterative methods. The reviewed methods are shown to outperform state-of-the-art approaches in both error-rate performance and computational complexity.

- **Chapter 3**: In this chapter, we present multi-carrier and single-carrier modulation based on the Iterative Block Decision Feedback Equalizer (IB-DFE) receiver. We outline the receiver design and address the issues related to high complexity in the inversion matrix.

- **Chapter 4**: This chapter introduces a low-complexity massive MIMO signal detection algorithm based on a modified accelerated overrelaxation (MAOR) method. The MAOR-based iterative detection is further improved by deriving optimal values for the acceleration and relaxation parameters. An eigenvalue-based initial solution is applied to achieve a better convergence rate for the MAOR. Finally, the Chebyshev polynomial acceleration method is applied to MAOR iterations to reach near-MMSE performance.

- **Chapter 5**: This chapter presents an acceleration method combined with recent iterative methods to improve convergence. Our method achieves linear MMSE performance with significantly lower computational complexity. We employ Chebyshev acceleration, which rewrites the iterative method with a new vector combination using the spectral radius of the iteration matrix and part of the iterative method. To enhance initialization, we propose using the stair matrix as the starting point for iterative methods. Our performance results demonstrate that Chebyshev acceleration, combined with iterative methods, can achieve near-linear MMSE performance, surpassing state-of-the-art methods in terms of error rate performance and computational complexity.

- **Chapter 6**: This chapter introduces Chebyshev acceleration for AOR and SOR algorithms, optimized through a deep unfolded network structure and offline training. The proposed methods, AC-AOR and AC-SOR, improve the detection performance of conventional AOR and SOR with minimal complexity increase. Additionally, the AC-AORNet and AC-SORNet networks, which are deep unfoldings of the AC-AOR and AC-SOR algorithms, feature cascaded layers with trainable parameters and an optimized training scheme. Performance testing demonstrates that AC-SORNet and AC-AORNet exhibit superior performance and robustness compared to conventional methods and current state-of-the-art techniques.

- **Chapter 7**: This chapter focuses on massive MIMO systems using SC-FDE modulations and an IB-DFE receiver. To reduce receiver complexity, we design an alternative structure that incorporates the Richardson iterative method to avoid the calculation of the inversion matrix. Furthermore, we apply the Chebyshev polynomial to accelerate the Richardson method.

- **Chapter 8**: This chapter offers a brief summary of the main contributions presented in this thesis and identifies possible directions for future research.

1.8 Published articles

The work associated with this thesis led to the publication of the following journal articles.

- Berra, S., Dinis, R., Rabie, K., & Shahabuddin, S. (2022). Efficient iterative massive MIMO detection using Chebyshev acceleration. Physical Communication, 52, 101651.

- Kansal, Lavish, Salah Berra, Mohamed Mounir, Rajan Miglani, Rui Dinis, and Khaled Rabie. "Performance analysis of massive MIMO-OFDM system incorporated with various transforms for image communication in 5G systems." Electronics 11, no. 4 (2022): 621.

- Berra, S., Dinis, R., & Shahabuddin, S. (2022). Fast matrix inversion based on Chebyshev acceleration for linear detection in massive MIMO systems. Electronics Letters, 58(11), 451-453.

- Berra, S., Chakraborty, S., Dinis, R., & Shahabuddin, S. (2023). Deep Unfolding of Chebyshev Accelerated Iterative method for Massive MIMO Detection. IEEE Access.

- Berra, S., Chakraborty, S., Dinis, R., (2024). A Low Complexity Iterative SC-FDE Receiver for Massive MIMO Systems. IEEE Transactions on Vehicular Technology.

- Berra, S., Benchabane, A., Chakraborty, S., Maruta, K., Dinis, R., & Beko, M. (2024). A Low Complexity Linear Precoding Method for Extremely Large-Scale MIMO Systems. IEEE Open Journal of Vehicular Technology.

CHAPTER

2 Massive MIMO system model and detection challenges

2.1 Introduction

Massive MIMO technology has emerged as a cornerstone of modern wireless communication systems, particularly in the context of 5G and beyond. By equipping BS with a large array of antennas, massive MIMO systems can simultaneously serve multiple users, significantly enhancing spectral efficiency, data throughput, and overall network capacity. This paradigm shift promises to meet the growing demands for higher data rates and more reliable communication in increasingly congested wireless environments.

The principle behind massive MIMO is the spatial multiplexing of signals, which allows for the concurrent transmission and reception of multiple data streams. This is achieved by taking advantage of advanced signal processing techniques to exploit the spatial dimensions of the wireless channel. However, the benefits of massive MIMO come with substantial challenges, particularly in the areas of system modeling and signal detection.

Accurate system modeling is crucial for designing and optimizing massive MIMO systems. It involves characterizing the wireless channel, accounting for hardware imperfections, and understanding the impact of various environmental factors. Moreover, detecting transmitted signals in a massive MIMO configuration is intrinsically challenging because of the high-dimensional aspects of the issue, the existence of noise, and the interference of numerous users.

This chapter introduces the model of the massive MIMO system and discusses its associated challenges. It examines conventional detectors and recent advancements, highlighting the difficulties in detection as the number of antennas increases. The chapter explores low-complexity detection schemes for massive MIMO and proposes an acceleration technique based on iterative methods. The reviewed technique is demonstrated to be applicable across various methods, outperforming state-of-the-art approaches in both error-rate performance and computational complexity.

2.2 System model

The massive MIMO system illustrated in the Figs. 2.1, and 2.2, the BS has employed R antennas to serve simultaneously T selected single antennas from UTs in the same frequency band, where $R >> T$ [30]. The distributed parallel bit streams of each user T are converted to constellation symbols O by using values from a constellation of energy normalization modulation (e.g. 64-QAM) and then distributed by different transmission antennas R. Let $\mathbf{x} = [x_1, x_2..., x_T]^T \in \mathbb{C}^{T\times 1}$ be the transmitted signal vector, and let the channel matrix $\mathbf{H} \in \mathbb{C}^{R\times T}$ denote the Rayleigh fading. The entries of $\mathbf{H}$ are independent and identically distributed (i.i.d.) with zero mean and unit variance [67].

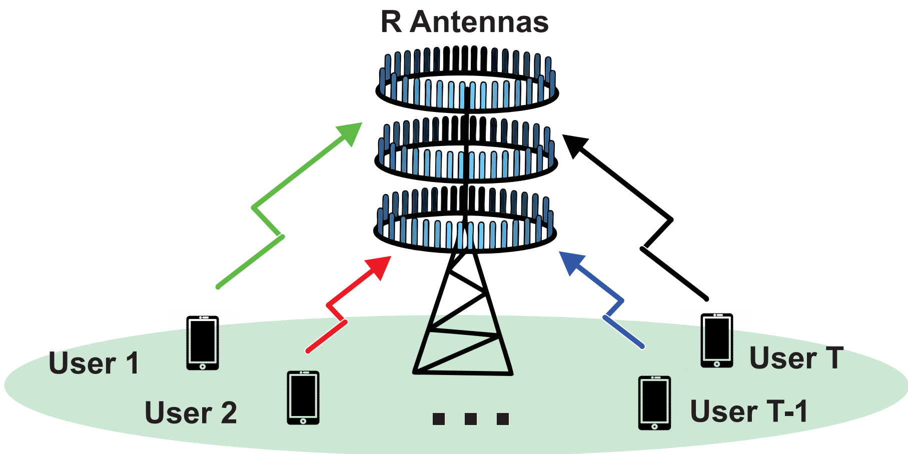


Figure 2.1: The system of massive MIMO system.

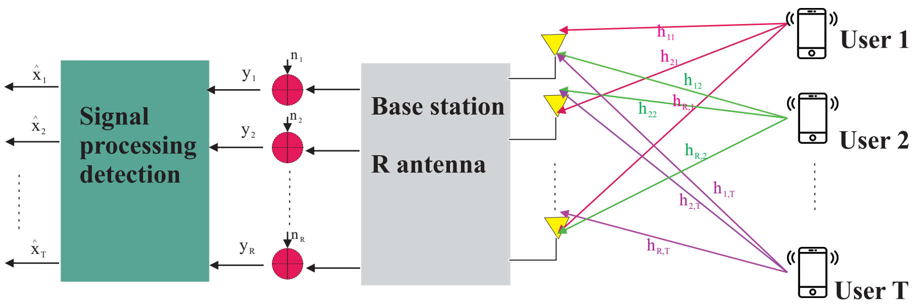


Figure 2.2: The signal detection massive MIMO system.

The received vector is presented as $\mathbf{y} = [y_1, y_2..., y_R]^T \in \mathbb{C}^{R\times 1}$ and it is presented as:

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

where $\mathbf{n} = [n_1, n_2..., n_R]^T \in \mathbb{C}^{R\times 1}$ denotes the additive white-Gaussian noise vector, whose entries follows the distribution $\mathcal{CN}\ (0, \sigma_n^2)$. σ_n^2 is the noise variance. The transmitted signal vector $\mathbf{x}$ is estimated by the signal detection algorithm based on $\mathbf{y}$.

2.2.1 The correlation channel model

In this work, we consider the Rayleigh fading channel and a spatially correlated channel, which are commonly employed in massive MIMO [68]. The correlated channel is presented as a Kronecker channel model [69], which incorporates spatial correlation between antenna elements. It is also assumed that the spatial correlations at the transmitter and receiver are separable. The Kronecker channel model $\check{\mathbf{H}}$ can be illustrated as

$$\mathbf{H} = \sqrt{\mathbf{R}_{Rx}}\check{\mathbf{H}}\sqrt{\mathbf{R}_{Tx}}, \tag{2.2}$$

where $\mathbf{R}_{Tx} \in \mathbb{C}^{T\times T}$ and $\mathbf{R}_{Rx} \in \mathbb{C}^{R\times R}$ denote the transmit correlation matrix and the receive correlation matrix and $\check{\mathbf{H}} \in \mathbb{C}^{R\times T}$ is a matrix with independent and identically distributed (i.i.d.) complex Gaussian random variables (RVs) with zero mean and unit variance (i.i.d.) $\mathcal{CN}\ (0, 1)$. We assume exponential correlation matrices [69], which means that the (p, q) element of $\mathbf{R}_{Tx}$ is given by

$$\mathbf{R}_{Tx}(q, p) = (\zeta_{Tx})^{q-p} \tag{2.3}$$

where ζ_{Tx} indicates a correlation parameter between two adjacent antennas ($0 \leq |\zeta_{Tx}| \leq 1$) and its phase can take any value in [0,2π]). For $|\zeta_{Tx}| = 0$ there is no correlation between transmit antennas. Similarly, the (p, q) term of the receive correlation matrix is

$$\mathbf{R}_{Rx}(q, p) = (\zeta_{Rx})^{q-p} \tag{2.4}$$

with the correlation parameter ζ_{Rx}.

2.3 The challenge of massive MIMO system

Although integrating several antennas improves communication, massive MIMO introduces additional signal processing issues, including Channel estimate, precoding, and detection. The channel estimate significantly affects the performance of the wireless system. The base station requires a precise assessment of the state of the channel. Information is required to fully utilize massive MIMO's benefits in practice the channel state information(CSI). Obtaining CSI for several channels might be challenging [70]. CSI must be exchanged quickly and with minimum latency between transmitters [71]. The pilot contamination effect might negatively impact the channel estimate [72]. To extract channel information

in the presence of inter-cell interference, finite-length pilot sequences are used. As a result, the pilot sequences of neighboring cells might contaminate one another. Therefore, channel estimation issues should be addressed in massive MIMO to provide a substantial improvement in the performance of RF transceivers for assistance, which is too expensive in the case of massive MIMO.
Another challenges Precoding, is a signal-processing approach that optimizes connection performance by using the transmitter's CSI. The BS must precode the downlink data to focus the spatial data streams on the user's location [73]. In the downlink, transmit precoding can direct each signal to the correct receiver. In a non-LOS situation, concentrating an antenna array on a given terminal might be challenging due to geographical considerations of multipath components. Small-scale MIMO precoding methods, such as Zero-Forcing (ZF), modify symbols' amplitudes and phases of symbols on the baseband and rely on specialized RF transceivers. Therefore, each antenna element requires a dedicated antenna. The BS requires accurate and immediate CSI for precoding (downlink) and detection (uplink). The Matched filter (MF) detector performs well in high-scattering channels with few users.

To improve spectral efficiency in spatially correlated channels, improved detectors are required. The complexity of large MIMO detection algorithms depends on the system size (number of antennas, transmitters, and receivers), matrix multiplication, and matrix inversion. It is important to strike a balance between performance and complexity [74]. However, realizing a large number of antennas at the base station can introduce issues, particularly in tasks such as matrix inversion and solving linear systems; making detection in practice is not a trivial task. In this chapter, we consider the challenge of detector issues in the uplink scenario of massive MIMO. In the next section, we provide a review of recent work on detectors.

2.4 Literature review of massive MIMO detectors

Developing low-complexity, high-performance data detection algorithms has been a long-standing challenge in research. The maximum likelihood (ML) detector is an optimal detection technique for the MIMO system. However, its complexity increases exponentially with modulation order and number of transmit antennas, making it inapplicable even for a small number of antennas [75]. Sphere Decoding (SD) has been proposed as a sub-optimal solution that minimizes the complexity of MLD by limiting the search space to a high-dimensional sphere. However, the complexity of SD is still very high, making it impractical for use in massive MIMO systems [76]. Furthermore, non-linear interference cancellation techniques, such as successive interference cancellation (SIC), parallel interference cancellation (PIC), and iterative interference cancellation (IIC), where the interfering signals can be regenerated and subsequently canceled from the desired signal, do not achieve the performance offered by the MLD. Their complexity is also very high [77, 78]. The linear ZF and MMSE detectors are recognized to have lower complexity compared to the techniques mentioned above. Still, they require the inversion of a matrix whose dimensions can be enormous in massive MIMO systems, making it challenging when the number of antennas and users increases.

Several linear detection methods have been developed in the literature to avoid the complexity issue in linear detection. Generally, two main categories of methods exist to overcome the complexity of inverting a matrix. The first category is based on approximate matrix inversion methods, which include Neumann series (NS) [79], truncated polynomial expansion (TPE) [80], and Newton iteration (NI)[81]. In [79], authors utilized the Neumann series method to reduce the complexity of linear detection by simplifying matrix inversion to a series of matrix multiplications. However, it suffers from a slower convergence rate if the ratio between the number of antennas at the BS and the users is not large enough. Further, NS requires more iterations due to its slower convergence rate. In [81], the Newton iteration (NI) method employs a matrix product to achieve the MMSE performance. It is especially useful for small antenna ratios where the NS method is not applicable. However, the complexity of NI is higher than that of the NS method. In [82], an improved version of NI is proposed in which selecting the initial solution plays an important role, requiring additional computations. The convergence rate of the TPE method depends on a finite number of terms J in the Taylor series expansion [80]. In addition, if the order of the polynomial changes, all polynomial coefficients must be recalculated, and the calculation of the corresponding polynomial coefficients is difficult [83]. The second category of detection techniques for massive MIMO is based on iterative methods, in which the transmitted vector can be approximated without the need for a matrix inverse. In general, the iterative method achieves a balance between complexity and performance than the approximation inversion matrix methods [84] Examples of these methods include Jacobi (JA) [85], Gauss-Seidel (GS) [86], successive over-relaxation (SOR) [87], accelerated over-relaxation (AOR) [88], Richardson iteration (RI) [89], and conjugate gradient (CG) [90] etc. The Jacobi method [85] is a linear detection technique that can estimate the transmitted vector without requiring computationally intensive matrix inversions. However, the convergence rate of the Jacobi method is low and requires a large number of iterations to get near-ZF results. In [91], a hybrid iterative scheme has been proposed that combines the Jacobi method with the steepest descent (STD) method to improve the convergence rate. The STD algorithm significantly improves the convergence speed by providing an efficient searching direction to the Jacobi method. On the other hand, the GS-based detectors [86] have a higher convergence rate than the JA method. However, since the internal iterations of the GS method require sequential operation, the parallel implementation of the GS method is difficult. In [90], the CG detector has been proposed. Still, it suffers from correlation issues and low parallelism. Like the JA method, the convergence rate of Richardson iteration is also low [89]. To further improve the convergence rate of the RI, an initial solution generation technique using the Newton method [92] has been proposed. Although a suitable initial solution can improve the conventional Jacobi and Richardson convergence rate and detection performance, the resulting detector complexity also increases. The SOR detector [87] can achieve good convergence by utilizing an appropriate matrix splitting pattern and a relaxation factor. In [93], authors proposed an AOR-based detection technique, which requires two parameters: relaxation coefficient and acceleration coefficient. However, selecting the optimum parameters remains challenging since they depend on the spectral radius and eigenvalues of the inversion matrix [94]. While iterative methods can be attractive due to their simplicity, the convergence rate can be further improved through the Chebyshev acceleration [95]. Recent studies have applied the Chebyshev acceleration technique to improve the convergence rates of iterative methods in massive MIMO detection [96, 97]. Recently, there has been a growing interest in using deep unfolding for MIMO detection to balance accuracy and complexity [98, 99]. In [100], the detection network (DetNet) was proposed, which unfolds a projected gradient descent method. Although DetNet has shown significant performance gains for low-order modulations such as BPSK and QPSK, its performance is poor for larger constellations. Moreover, it requires a large number of trainable parameters. In [101], an orthogonal approximate message passing network (OAMPNet) has been proposed, which unfolded the iterations of an OAMP detector. Using data-driven training, OAMPNet has outperformed conventional OAMP in overall detection performance. In addition, the deep unfolding technique was also proposed to improve multiuser interference cancellation schemes. For example, in [102], an iterative sequential detection scheme has been unfolded to enhance its performance. Furthermore, the trainable projected gradient-detector (TPG-detector) has been proposed to optimize the internal parameters using a deep unfolding network [103]. In [104], the modified expectation propagation (MEP) algorithm has been unfolded and named as MEPNet. Two learnable parameters have been optimized through the deep unfolded network. The resulting detector has shown significant improvement in error-rate performance compared to the original MEP. An expectation propagation network (EPNet) has recently been proposed for an orthogonal time frequency space (OTFS) system [105]. The parameters in the resulting detector were optimized using offline training and the detector demonstrated superior performance.

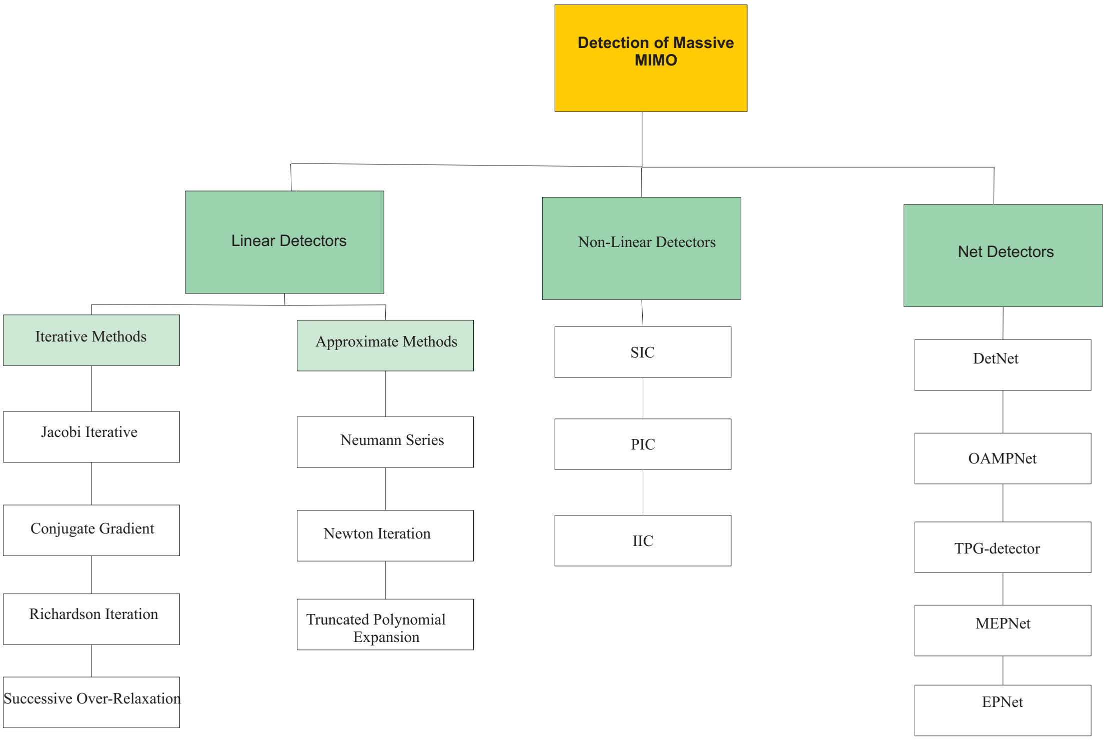


Figure 2.3: The methods detection of massive MIMO system.

2.5 Linear and non-linear detection in massive MIMO systems

The task of the signal detection process is to recover the vector of the transmitted signal $\mathbf{x}$ from the noisy detection signal $\mathbf{y}$. Signal detection is not an easy task, since massive MIMO systems employ a large number of antennas at the base station, which leads to high complexity and necessitates the design of lower complexity detection algorithms. Signal detection can consist of two categories: linear detectors and non-linear detectors, each with its own performance and accuracy characteristics.

2.5.1 Linear detector

The optimal ML detection minimizes the squared Euclidean distance between the actual received vector $\mathbf{y}$ and the hypothesized received signal $\mathbf{Hx}$, the ML solution is given by

$$\hat{\mathbf{x}}_{ML} = 2 \min_{\mathbf{x} \in O^T} ||\mathbf{y} - \mathbf{Hx}||_2^2. \tag{2.5}$$

However, the ML receiver has complexity which grows exponentially with the number of users and the constellation size, which precludes its use even in MIMO scenarios with a moderate number of users. MMSE equalization-based detection is widely used due to its good performance/complexity tradeoffs. The MMSE estimate of the transmitted signal vector $\hat{x}_{MMSE}$ is

$$\hat{\mathbf{x}}_{MMSE} = (\mathbf{H}^H\mathbf{H} + \sigma_n^2\mathbf{I}_R)^{-1}\mathbf{H}^H\mathbf{y} = (\mathbf{G} + \sigma_n^2\mathbf{I}_R)^{-1}\mathbf{H}^H\mathbf{y} = \mathbf{W}^{-1}\mathbf{y}^{MF}, \tag{2.6}$$

where $\mathbf{y}^{MF} = \mathbf{H}^H\mathbf{y}$ is the output of matched filter, $\mathbf{W}$ is the MMSE filter matrix and $\mathbf{G} = \mathbf{H}^H\mathbf{H}$ is the Gram matrix of the channel [106, 107]. From (2.6), it is clear that the MMSE receiver involves the inversion of a matrix whose dimensions grow with the number of users, which might be too complex for massive MIMO systems. The MMSE detection in (2.6) requires an inversion of $\mathbf{W}$ which has $O(T^3)$ complexity. Other linear detectors, such as ZF, have performed excellently in massive MIMO systems. The detected signal vector for the ZF detector [32] is given by

$$\hat{\mathbf{x}} = (\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\mathbf{y} = \mathbf{W}^{-1}\mathbf{H}^H\mathbf{y}, \tag{2.7}$$

where $\mathbf{W} = \mathbf{H}^H\mathbf{H}$ for the ZF detector, and $\mathbf{y}^{MF} = \mathbf{H}^H\mathbf{y}$ is the matched filter response. However, inverting the matrix $\mathbf{W}$ has a complexity of $O(T^3)$, which can be prohibitively high even for moderate values of the number of users T.

Lemma 1 *The MMSE and ZF filtering matrix* $\mathbf{W}$ *is symmetric positive definite for signal detection of massive MIMO systems.*

The column vectors of the channel matrix $\mathbf{H}$ in large-scale MIMO systems are asymptotically orthogonal (i.e., $\mathrm{rank}(\mathbf{H}) = R$) [30]. Then we have the equation $\mathbf{Hr} = 0$ when and only when $\mathbf{r}$ is a $R \times 1$ zero vector. Thus, for an arbitrary non-zero $R \times 1$ vector $\mathbf{r}$, we have

$$(\mathbf{Hr})^H\mathbf{Hr} = \mathbf{r}^H(\mathbf{H}^H\mathbf{H})\mathbf{r} > 0, \tag{2.8}$$

which indicates that the Gram matrix $\mathbf{G} = \mathbf{H}^H\mathbf{H}$ is positive definite. The column vectors of the channel matrix are asymptotically orthogonal [30], but even for smaller dimensions, a Gram matrix is always Hermitian and (at least) positive semi-definite (i.e., the eigenvalues are greater or equal to 0). This means that $\mathbf{G} + \sigma_n^2\mathbf{I}_R$ is positive definite, since $\sigma_n^2 > 0$.

Lemma 2 *for M-MIMO, the distribution of the eigenvalues can be well approximated by the Marchenko-Pastur distribution.*

Consider a random matrix $\mathbf{H}$ of dimensions $R \times T$, where each entry is an i.i.d. Gaussian random variable (RV). Denote the element in the i-th row and j-th column of $\mathbf{H}$ as $[\mathbf{H}]_{i,j} \sim \mathcal{CN}(0, \frac{1}{N})$. In a massive MIMO system, $\mathbf{H}$ represents the small-scale Rayleigh fading channel matrix between T user terminals and N base station antennas. As both $R, T \to \infty$, while maintaining the ratio $R/T = \xi$, the empirical cumulative distribution function of the eigenvalues of $\mathbf{H}$ (the spectrum) and its functionals exhibit interesting convergence properties. Specifically, the spectrum of $\mathbf{H}$ becomes deterministic in this asymptotic scenario. This fundamental aspect of asymptotic random matrix theory asserts that the empirical distribution of the moments of $\mathbf{H}$'s eigenvalues and its functionals becomes deterministic, regardless of the matrix entries' distribution. Specifically, the spectrum of $\mathbf{H}^H\mathbf{H}$ almost surely converges to a non-random distribution known as the Marchenko-Pastur law, whose density function [108, 109]

$$f_\xi(x) = \left(1 - \frac{1}{\xi}\right)^+ \delta(x) + \frac{\sqrt{(x-a)^+(b-x)^+}}{2\pi\xi x} \tag{2.9}$$

where $(z)^+ = \max(0, z)$ and

$$a = (1 - \sqrt{\xi})^2, \quad b = (1 + \sqrt{\xi})^2. \tag{2.10}$$

Analogously, the empirical distribution of the eigenvalues of $\mathbf{H}^H\mathbf{H}$ converges almost surely to a non-random limit whose density function is

$$\bar{f}_\xi(x) = (1 - \xi)\delta(x) + \xi f_\xi(x) \tag{2.11}$$

$$= (1 - \xi)^+\delta(x) + \frac{\sqrt{(x-a)^+(b-x)^+}}{2\pi\xi x}. \tag{2.12}$$

These results are particularly useful given that eigenvalues of random matrices are used to characterize the performance of communication links (for e.g., MIMO links).

When $\mathbf{H}^H\mathbf{H} = \mathbf{I}_R$, the limiting distribution aligns with the Marčenko-Pastur law, originally explored by Marčenko and Pastur in [109]. This distribution is illustrated in Fig. 2.4, where it is compared with the empirical eigenvalue distribution of a large-dimensional matrix $\mathbf{H}^H\mathbf{H}$. The figure illustrates that the empirical eigenvalue distribution closely aligns with the limiting distribution when R and T are large. Moreover, it is important to observe that no eigenvalue seems to lie outside the bounds of the limiting distribution, even when R and T have finite values. This notable feature can be proven for any matrix $\mathbf{H}^H\mathbf{H}$ with independent entries that have a zero mean and unit variance, given that certain mild conditions on higher-order moments are satisfied.

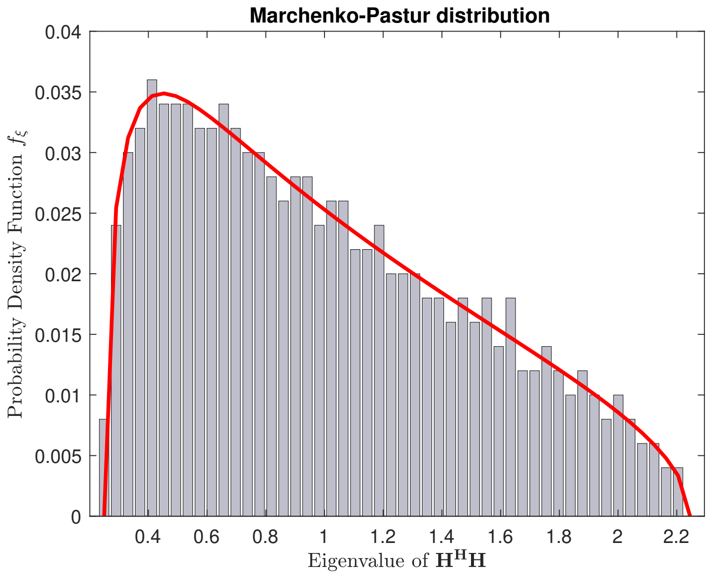


Figure 2.4: Histogram of the eigenvalues of $\mathbf{H}^H\mathbf{H}$ for T = 500, R = 2000.

2.5.2 Linear detectors based on the iterative method

In this section, we will provide some low-complexity detectors that have been presented in recent years. We start by addressing the issue of conventional detectors and then explore methods to reduce complexity. As observed in a recent section, the inversion of the matrix $\mathbf{W}^{-1}$ required by (2.6) and (2.7) involve the operations $O(T^3)$ (where T denotes the number of users), significantly impacting the overall complexity of the receiver's signal processing. Given that $\mathbf{H}$ is the full rank column and asymptotically orthogonal [30], and the matrix $\mathbf{W}$ is Hermitian positive definite [110], we can use these properties in iterative methods to estimate the signal without matrix inversion. The iterative method is a powerful tool for solving linear equation systems ($\mathbf{AX} = \mathbf{b}$), where the vector $\mathbf{X}$ is a solution in dimension $(R \times 1)$, $\mathbf{A}$ is a non-singular and Hermitian positive definite matrix $(R \times R)$, and $\mathbf{b}$ is a vector of measurements $(R \times 1)$. To avoid exact computation of $(\mathbf{A}^{-1}\mathbf{b})$, the iterative method is applied [111]. The MMSE weighting matrix $\mathbf{W}$ can be decomposed as follows:

$$\mathbf{W} = \mathbf{D} - \mathbf{U} - \mathbf{L}, \tag{2.13}$$

where $\mathbf{D}$, $\mathbf{U}$, and $\mathbf{L}$ represent the diagonal matrix, the upper triangular matrix, and the components of the lower triangular matrix, respectively. A low-complexity, near-optimal signal detection algorithm is proposed for large-scale MIMO systems using the GS method [112]. This algorithm leverages the Hermitian positive definite property of the MMSE filtering matrix, eliminating the need for matrix inversion. The resultant GS-based detector is introduced as:

$$\mathbf{x}^{(i+1)} = (\mathbf{D} - \mathbf{L})^{-1}\mathbf{U}\mathbf{x}^{(i)} + \mathbf{U}\mathbf{y}^{MF} \tag{2.14}$$

The detector based on the GS method iteratively resolves the matrix inversion, with further optimizations for accelerated performance. In a related work [113], a refinement of the conventional GS method is introduced based on the concept of the band matrix, in order to accelerate convergence and reduce overall complexity while ensuring near-optimal BER performance. An efficient hardware architecture is proposed for the detection of massive MIMO in soft output based on GS, introducing a novel initial solution based on the NS [114]. The GS algorithm employs Taylor expansion and group inversion to improve efficiency [115]. Another algorithm based on Jacobi is proposed, eliminating the need for $\mathbf{W^{-1}}$:

$$\mathbf{x}^{(i+1)} = \mathbf{D}^{-1}(\mathbf{D} - \mathbf{W})\mathbf{x}^{(i)} + \mathbf{D}^{-1}\mathbf{y}^{MF}. \tag{2.15}$$

Efforts to enhance the acceleration and performance of this detector are discussed in [113]. Another iterative method is the SOR-based approach:

$$\mathbf{x}^{(i+1)} = (\mathbf{D} - \omega\mathbf{L})^{-1}((1-\omega)\mathbf{D} + \omega\mathbf{U})\mathbf{x}^{(i)} + (\mathbf{D} - \omega\mathbf{L})^{-1}\mathbf{y}^{MF}. \tag{2.16}$$

In [116], a stable and efficient SOR-based detection, nearly unaffected by ω, is proposed. In [87], two detectors, non-adaptive SOR (NA-SOR) and adaptive SOR (A-SOR), are introduced, the latter achieving better performance, especially in highly correlated channels. An efficient pipelined hardware architecture for the A-SOR detector is proposed. A Richardson-based approach is also presented.

$$\mathbf{x}^{(i+1)} = (\mathbf{I}_R - \omega\mathbf{W})\mathbf{x}^{(i)} + \omega\mathbf{y}^{MF}. \tag{2.17}$$

In these equations, ω of SOR belongs to the interval $[0, 2]$, and ω of Richardson belongs to the interval $[0, 2/\lambda]$, where λ is the largest eigenvalue [84]. The Richardson method achieves near-ML performance but requires numerous iterations [110]. In [117], the authors propose a detection algorithm with a high convergence rate and efficient hardware architecture based on second-order Richardson iteration (SORI), which a pre-iteration-based initialization the method is presented to accelerate the convergence without extra complexity. Those algorithmic methods can reduce the inversion matrix by one order, but the convergence of those methods still faces problems and requires further development to find optimal parameters and achieve lower complexity based on fewer iterations.

2.5.3 Linear detectors based on the approximate matrix inversion

After discussing the iterative methods above, in this subsection, we will review some approximation methods that avoid calculating the matrix inversion. These methods can convert the MMSE filtering matrix into matrix-vector multiplication, which reduces the complexity of the inversion matrix. The NS is used to arrive at an approximation of the matrix inverse, which subsequently reduces the complexity of the linear detector and provides an advantage in hardware implementation [118], [119].

The NS algorithm deals with the non-diagonal matrix by decomposing $\mathbf{W}$ into the sum of diagonal elements matrix $\mathbf{D}$ and non-diagonal elements matrix $\mathbf{E}$, which needs more iterations to achieve a certain performance [120]. The NS expands the inverse matrix $\mathbf{W}^{-1}$ as

$$\mathbf{W}^{-1} \approx \sum_{i=0}^{\infty} \left(\mathbf{I}_R - \mathbf{X}^{-1}\mathbf{W}\right)^i \mathbf{X}^{-1}, \tag{2.18}$$

where $\mathbf{X}$ is the matrix of an initial approximation of $\mathbf{W}^{-1}$. The matrix $\mathbf{X}$ must satisfy

$$\lim_{i\to\infty} \left(\mathbf{I}_R - \mathbf{X}^{-1}\mathbf{W}\right)^i \approx \mathbf{0}_R. \tag{2.19}$$

Then, $\mathbf{P}$ can be decomposed into $\mathbf{P} = \mathbf{D} + \mathbf{L}$, where $\mathbf{D}$ is the main diagonal matrix and $\mathbf{L}$ is the non-diagonal matrix [121]. The NSE expansion of $\mathbf{P}$ is given by

$$\mathbf{P}^{-1} = \sum_{i=0}^{\infty} (-\mathbf{D}^{-1}\mathbf{L})^i \mathbf{D}^{-1}. \tag{2.20}$$

The polynomial expansion in (2.20) converges to the matrix inverse $\mathbf{P}^{-1}$ if

$$\lim_{i\to\infty} (-\mathbf{D}^{-1}\mathbf{L})^i = 0. \tag{2.21}$$

In practice, a finite number of terms is utilized, and, thus, a fixed number of iterations of (2.20) is performed. As the number of iterations i increases, high precision of the matrix inverse will be achieved at the expense of extra complexity.

More precisely, [122] derives that $\xi = R/T > 5.83$ can offer a high convergence probability for the Neumann method in massive MIMO systems. Generally, the truncated NS[123] is more concerned due to its lower computational complexity. The complexity of the i-term Neumann series is $\mathcal{O}(T^2)$ for $i \leq 2$. And for $i = 3$, the complexity of $\mathbf{W}^{-1}$ scales with $\mathcal{O}(T^3)$. For larger T, computing the NSE can be of higher complexity than the exact matrix inverse.

The Newton iteration (NI) is also known as the Newton-Raphson method [124] is an approximate method to calculate the matrix inversion. It is derived from the first-order Taylor series and can improve precision by iteration. If $\mathbf{X}_0^{-1}$ is the rough and original estimate of $\mathbf{W}^{-1}$, then the estimation of the $(i + 1)$ th iteration is

$$\mathbf{X}_i^{-1} = \mathbf{X}_{i-1}^{-1}(2\mathbf{I} - \mathbf{W}\mathbf{X}_{i-1}^{-1}). \tag{2.22}$$

For convergence, the initial input $\mathbf{X}_0^{-1}$ must satisfy the condition [125]

$$||\mathbf{I}_R - \mathbf{W}\mathbf{X}_0^{-1}|| < 1. \tag{2.23}$$

Then, (2.23) converges quadratically to $\mathbf{W}^{-1}$. For an arbitrary matrix, it usually requires a complex calculation to obtain an appropriate initial input $\mathbf{X}_0^{-1}$. The Newton iteration algorithm simplifies the matrix inversion process. However, the Newton iteration algorithm has quadratic convergence, and its convergence rate is slow. To achieve cubic convergence, the Newton iteration algorithm can be modified to obtain the Chebyshev iteration [82] as,

$$\mathbf{X}_i^{-1} = \mathbf{X}_{i-1}^{-1}(3\mathbf{I}_R - \mathbf{W}\mathbf{X}_{i-1}^{-1}(3\mathbf{I}_R - \mathbf{W}\mathbf{X}_{i-1}^{-1})). \tag{2.24}$$

The Chebyshev iteration also requires an initial value and certain conditions for Newton iterative convergence rate.

2.5.4 Non-linear detection

The non-linear algorithms, such as expectation propagation (EP)-based methods [126] can achieve near-optimal performance at the expense of huge hardware resources due to the excessive complexity. Numerous non-linear signal detectors [127], such as K-best [128], fixed-complexity sphere decoding [129], and tabu search [130], have been proposed to attain the near-optimal performance with reduced complexity in traditional MIMO system. However, these detectors are unfeasible for massive systems because the excessive computing loads of these detectors would significantly de-escalate the data processing speed of the entire system. Nonlinear MIMO detection schemes such as successive interference cancelation (SIC) can help mitigate this effect, particularly in per-space-layer coded MIMO systems [131]. In the Parallel Interference Cancellation (PIC) technique also, all the CWs are detected in parallel, interleaved, re-modulated and sent back to the interference canceller [132]. The gains achieved thereby stem from an increased post-detection signal-to-interference-and-noise ratio (SINR). Since SIC is known to be sensitive to error propagation, a careful adjustment of the data rate at each spatial layer is mandatory, and decoding of already detected spatial layers is required prior to an interference cancelation step. Thus, performance gains come at the price of an increased computational complexity and detection delay at the receiver, and it is important to quantify the achievable gains under certain operation conditions in order to decide in advance on the more complex nonlinear MIMO detection scheme.

2.5.5 Deep learning detector

In the past ten years, deep learning (DL) has brought significant advancements to various fields, including computer vision [133] and speech recognition [134]. Motivated by these achievements, DL has been recently utilized in the design of communication systems, encompassing physical layer processing [135] (such as channel estimation [136], [137] and symbol detection [138]), and resource management [139], [140], among others. Among all DL-based applications, MIMO detection stands out as one of the most essential and fundamental challenges. DL-based detectors can learn to map received signals to transmitted symbols using training data, thereby achieving superior performance compared to traditional detection algorithms [141]. In general, DL-based detectors can be classified into two types: data-driven DL detectors that rely on deep neural networks (DNNs) and model-driven DL detectors derived from iterative detection algorithms. Data-driven DL detectors utilize DNN architectures for symbol detection [142], [143]. These DNN-based architectures are independent of specific models and can accurately recover transmitted symbols in various scenarios if adequately trained. Nevertheless, this capability comes with the cost of a large number of trainable parameters and training samples. Conversely, model-driven DL detectors are developed from traditional iterative detection algorithms, with each network layer representing a single iteration and incorporating some trainable variables [144]. These detectors generally exhibit superior performance and faster convergence compared to the original iterative detection algorithms [145]. However, current model-driven DL detectors assume a linear channel model and the availability of CSI, which restricts their use in complex environments.

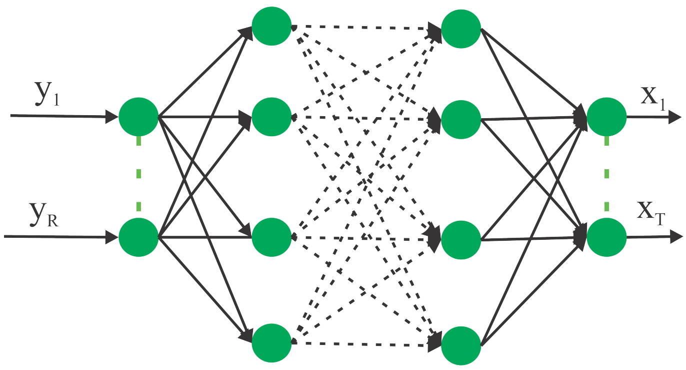


Figure 2.5: The structure of DNN.

Various learning-based massive MIMO detection methods, such as OAMPNet [145] and MMNet [100], [146], have been developed by unfolding iterative detectors [147]. These methods incorporate a small number of trainable parameters into the iterative detectors, which are optimized by minimizing an objective function over datasets. Since they build upon existing detection algorithms, these methods are often referred to as model-driven approaches. Although these model-driven methods demonstrate enhancements over their respective iterative detectors, their performance is heavily dependent on the original iterative techniques. Data-driven methods, on the other hand, extract features directly from vast datasets without depending on relevant models for problem-solving. The detection network (DetNet), exemplifying data-driven methods, is introduced in [144], [145] by transforming a projected gradient descent algorithm into a fully connected deep neural network (DNN). Given the DNN's strong learning capability from data and nonlinear representation, DetNet achieves performance comparable to semidefinite relaxation (SDR) and approximate message passing (AMP) detectors, albeit with significant training time. Subsequently, an enhanced network presented in [148], the sparsely connected neural network (ScNet), primarily simplifies DetNet's connection structure, significantly reducing network complexity and training time. Moreover, to enhance ScNet's detection performance in high-order modulation scenarios, the multi-segment mapping network (MsNet), which constructs a staircase function based on the sigmoid activation function, is introduced in [149]. These studies' findings highlight unprecedented performance improvements due to the DNN framework. Unlike model-driven approaches, the DNN network designed by the data-driven method exhibits greater nonlinearity and flexibility due to deeper network layers and a sufficient number of learnable parameters, enabling the resolution of more complex and larger problems, albeit at the cost of increased memory usage due to the rise in model parameters [150].

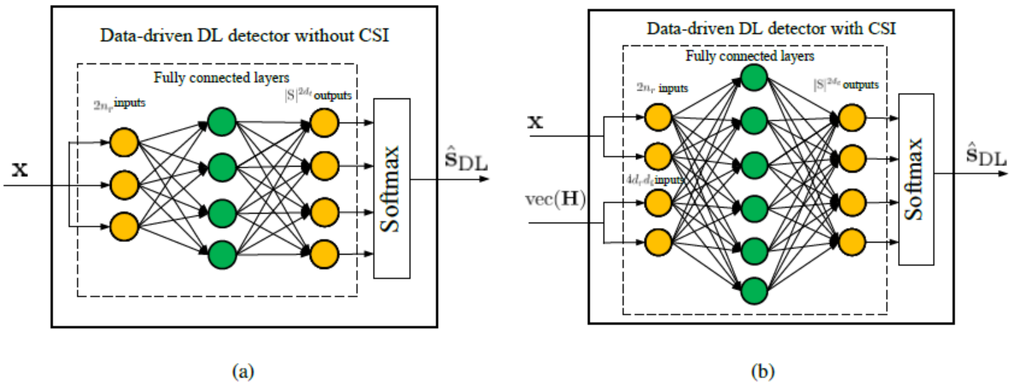


Figure 2.6: Network structures for the data-driven DL detectors with CSI and without CSI [151].

2.6 Applications of massive MIMO detection techniques in other areas

Massive MIMO detection methods can also be applied in more sophisticated scenarios. For instance, although massive MIMO-based underlay cognitive radio (CR) networks are an encouraging idea for future wireless networks to enhance spectral efficiency, the reuse of identical pilot sequences in both networks results in pilot contamination, which leads to residual interference [152]. Massive MIMO overlay CR systems were employed for bidirectional relaying, where the licensed spectrum was partitioned into several sub-channels in either the time or frequency domains. This enabled the Secondary Users

(SUs) to relay the spectrum of the Primary Users (PUs) and transmit secondary data through distinct sub-channels [153]. Furthermore, massive MIMO techniques can also be integrated with artificial intelligence (AI) enabled integration of massive MIMO techniques with virtualization of network functions (NFV) and software-defined network (SDN) architectures to support customized services over wireless mobile networks of 6G [154]. In addition to their dominant applications in wireless communications, massive MIMO detection techniques also significantly benefit a range of other research areas. For example, the idea of massive MIMO signal processing was extended to radar design, a target detection problem, and a robust Wald-type test that guarantees certain detection performance, regardless of the unknown statistical characterization of the disturbance [155]. Additionally, MIMO signal processing techniques are also instrumental in visible-light communication (VLC). For more details, the massive MIMO-VLC system employs a space division multiple access (SDMA) to support multi-users in addition to increasing the data rate of a single user and improving the received SNR through use of the spatial multiplying (SM), MRC and transmit/receive diversity (TxD, RxD) techniques, respectively [156]. Moreover, massive MIMO can be applied to media applications as shown in Fig. 2.9. Based on the detection, in [157], this work analyzes the performance of a massive MIMO-OFDM system incorporated with various transforms for image communication in 5G systems.

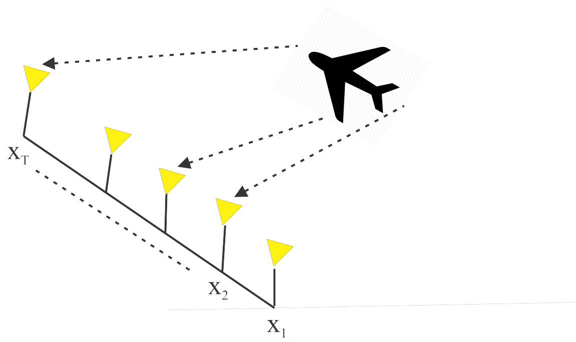


Figure 2.7: An example of MIMO radar systems.

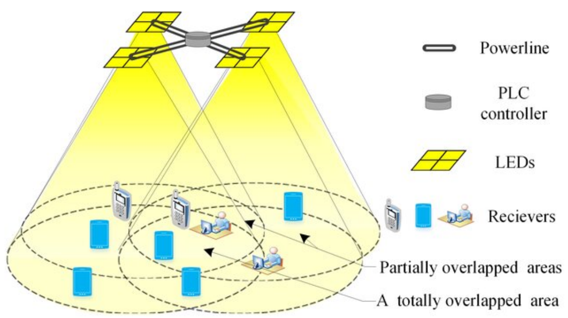


Figure 2.8: An example of multi-user multi-cell MIMO VLC systems [158].

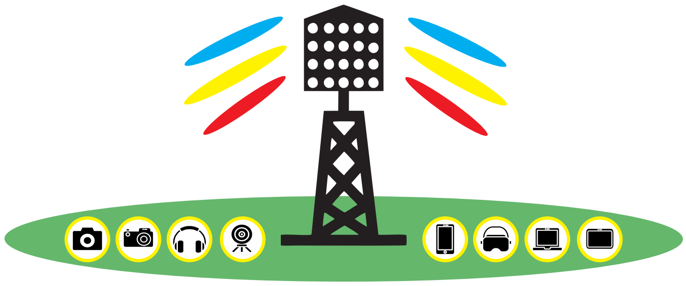


Figure 2.9: Employing Massive MIMO for Media Applications.

2.7 Summary

In this chapter, we review the latest receiver and detector designs in massive MIMO systems. We discuss most of the detection methods that have been published and designed, including deep learning detectors, which are hot topics today. Finally, we mention other detection applications that can be used in this field. However, the design of detectors or receivers remains a hot topic, and researchers are still looking for suitable receivers for the next generation.

Chapter

3 Receiver Techniques for Multi-Carrier and Single-Carrier Transmission Schemes massive MIMO Systems

Multi-carrier transmission methods are employed in wireless mobile systems to achieve nearly error-free high data rates. To accomplish this goal, multi-carrier modulation (MCM) has gained popularity. MCM works by splitting the transmitted data stream into multiple symbol streams with significantly lower symbol rates, which are then used to modulate various sub-carriers. MCM offers relative resistance to multipath fading and is less prone to interference from impulse noise compared to single-carrier systems. Furthermore, it shows improved immunity to interference [159]. One widely used multi-carrier system

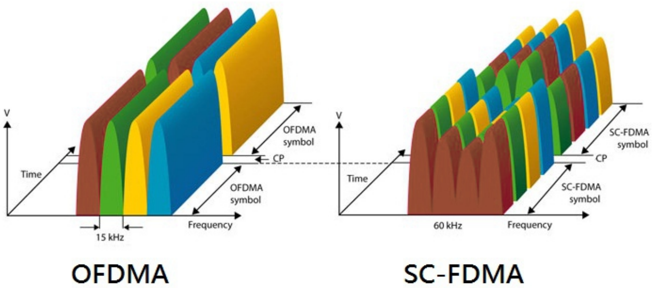


Figure 3.1: The comparison spectrum between OFDM and SC-FDMA.

in wireless communication is OFDM, known for its ability to cope with severe channel conditions such as attenuation of high frequencies in long copper wires, narrowband interference, and frequency-selective fading due to multipath. However, OFDM suffers from Peak-to-Average Power Ratio (PAPR), which is an issue for transmission efficiency.

To address this, SC-FDE modulation can avoid the PAPR effect and provide better results. A comparison of their spectrum is shown in Fig.3.1.

This chapter provides a brief overview of OFDM modulation and its integration with massive MIMO systems. Subsequently, we introduce SC-FDE modulation, highlighting its differences from OFDM and its advantages. Finally, we discuss the receiver design for SC-FDE in massive MIMO systems, including the associated complexities and optimization parameters.

3.1 OFDM Techniques

The most prominent example of MCM transmission is OFDM [160], [161]. In this technique, a large number of orthogonal sub-carriers are used to transmit information from several parallel streams modulated with a digital modulation scheme. If these data streams belong to different terminals or users, OFDM becomes orthogonal frequency division multiple access (OFDMA) [161] in which different data signals are transmitted through a common physical media that is divided into frequency resources units. An OFDM transmitter and receiver can be described using the Inverse Discrete Fourier Transform (IDFT) and Discrete Fourier Transform (DFT) blocks [162] and also the IFFT and FFT algorithms are then used to dramatically reduce the system's complexity [163] [164]. Fig. 3.2 shows the block diagram of the OFDM system model [165]. At the OFDM transmitter, the input binary data is mapped to constellations in the frequency domain using typically quadrature amplitude modulation (QAM) or PSK. The data is then converted from serial to parallel (S/P). After that, the data are converted to a time-domain signal using an IDFT as the equation below.

$$x_n = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X_k e^{\frac{j2\pi nk}{N}} \tag{3.1}$$

Afterward, the last part of the signal is appended to the beginning of the signal, known as the cyclic prefix (CP), in part to combat inter-symbol interference (ISI) and also to convert the linear convolution of the signal with the channel impulse response to a circular convolution. Lastly, the data is converted from parallel to serial and sent through a multipath communication channel with additive white Gaussian noise (AWGN). At OFDM receiver, the reverse operation of the OFDM transmitter is achieved and DFT can be represented as:

$$X_k = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x_n e^{\frac{-j2\pi nk}{N}}. \tag{3.2}$$

One important feature of the OFDM receiver is its ability to equalize in the frequency domain. There are many different ways of equalizing, two of which are the ZF and the MMSE equalizers. ZF equalizer causes noise enhancement and the MMSE equalizer causes time distortion in the signal. The equalizer-weight matrix $\tilde{X}_k = \mathbf{W}_k \mathbf{Y}_k$ may be expressed as

$$\mathbf{W}_k = \begin{cases} \mathbf{H}_k, & \text{MF-OFDM} \\ \mathbf{H}_k{}^{-1}, & \text{ZF-OFDM} \\ (\mathbf{H}_k{}^{H}\mathbf{H}_k + \sigma_n^2 \mathbf{I}_N)^{-1}\mathbf{H}_k, & \text{MMSE-OFDM,} \end{cases} \tag{3.3}$$

One benefit of equalization using any frequency domain technique is that applying it is simple, an element-wise multiplication in the frequency domain.

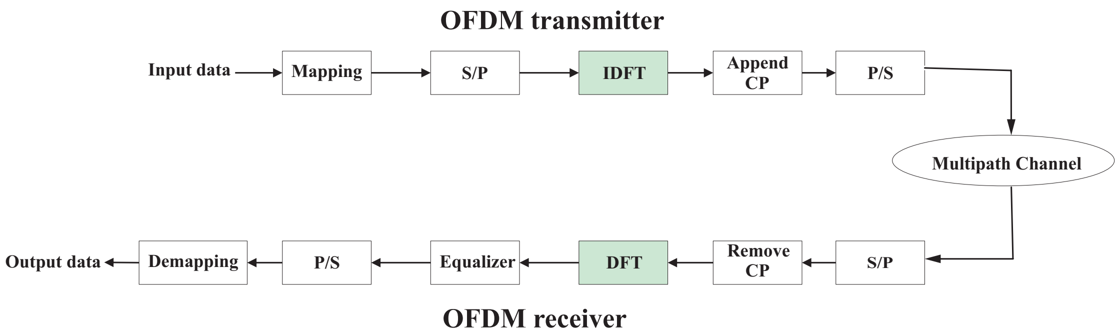


Figure 3.2: Block diagram of baseband OFDM system model.

OFDM and OFDMA suffer from distortion that may be particularly troublesome in uplink transmissions where excessive complexity in the user terminal is an issue.

3.1.1 OFDM-based massive MIMO Systems

To combine OFDM into massive MIMO in the uplink transmission where a BS with R antennas concurrently serves T single antenna user (the generalization for multi-antenna UEs is straightforward), and we assume $R \geq T$ (typically, $R >> T$). Each user transmits their data to an N-point IDFT block in the uplink, which inserts a CP into the data block before sending it through the channel. CP is a component of the block's end that is added to the block's beginning. The received signal from each BS antenna is sent via an N-point DFT block at the BS after the CP is removed, and the resulting signals are given to detection methods such as zero forcing or minim square error MMSE. The kth component of the tth vector received by the user, after removing the CP can be expressed as [166]

$$\mathbf{x}_k = \frac{1}{\sqrt{N}} \sum_{t=1}^{T} \sum_{n=0}^{N-1} \mathbf{X}_{t,n} e^{\frac{j2\pi nk}{N}}, \tag{3.4}$$

where $\mathbf{x}_k$ is the nth sub-carrier from the tth antenna. After adding CP the frequency domain representation of the received signal can be expressed as:

$$\begin{aligned}\hat{\mathbf{x}}_k &= \frac{1}{\sqrt{N}}\sum_{r=1}^{R}\sum_{t=1}^{T}\sum_{n=0}^{N-1}\mathbf{Y}_{t,r,n}e^{\frac{-j2\pi nk}{N}} \\ &= \frac{1}{\sqrt{N}}\sum_{r=1}^{R}\sum_{t=1}^{T}\sum_{n=0}^{N-1}\mathbf{H}_{t,r,n}\mathbf{x}_{r,n}e^{\frac{-j2\pi nk}{N}} + \eta_{r,k},\end{aligned} \tag{3.5}$$

where $\eta_{r,k}$ is the kth element of the zero-mean complex white Gaussian noise with variance added to the rth antenna of BS.

Furthermore, simplification of the implementation of the equalization process in OFDM [167] has ensured that 4G and 5G mobile networks will also largely rely on OFDM for providing high-speed data access to end users. LTE wireless systems have been implemented based on OFDM in 4G, which provide various types of traffic, fair scheduling, and balanced Quality of Service (QoS). LTE offers good QoS delivery for different traffic classes. The QoS scheduling rule strategy is suggested in an LTE downlink to provide high QoS delivery for different traffic classes using MU-MIMO [168]. However, OFDM affects PAPR. To address this limitation, single-carrier frequency division multiple access (SC-FDMA) [169] was selected as the uplink multiple access scheme because of its lower power distortion. SC-FDMA employs a modulation technique called SC-FDE.

3.2 SC-FDE Techniques

Although OFDM has been successful, it is hindered by several well-known issues such as high PAPR, sensitivity to amplifier nonlinearities, and high susceptibility to carrier frequency offsets. A low-complexity alternative to address ISI is the use of FDEs in SC communications [170]. Systems that use FDE are quite similar to OFDM systems. In both scenarios, digital transmission is performed in blocks and relies on DFT/IDFT operations. Consequently, SC systems with FDE have a complexity comparable to OFDM systems but without the strict requirements for precise frequency synchronization and linear power amplification, allowing for the use of more cost-effective components in user terminals with high efficiency. Additionally, FDEs offer significantly lower computational complexity compared to their time domain (TD) equalization counterparts.
Fig 3.3 shows the block diagram of the SC-FDE system model [170]. SC-FDE is a variation of OFDM, a modulation scheme that contains all the same blocks but moves the IDFT from the transmitter to the receiver. The difference between this system and OFDM is that the constellation mapping takes place in the time domain, and the CP is the last time samples are appended to the beginning of the signal. Equivalently, an SC-FDE symbol is

a group of $N + N_{CP}$ modulated (QAM, PSK) sub-symbols, where N is the length of the DFT/IDFT used in the receiver, and N_{CP} is the length of the cyclic prefix. This makes for a very simple transmitter. By moving the IDFT to the receiver, equalization is done with a simple multiplication in the frequency domain like in an OFDM receiver.

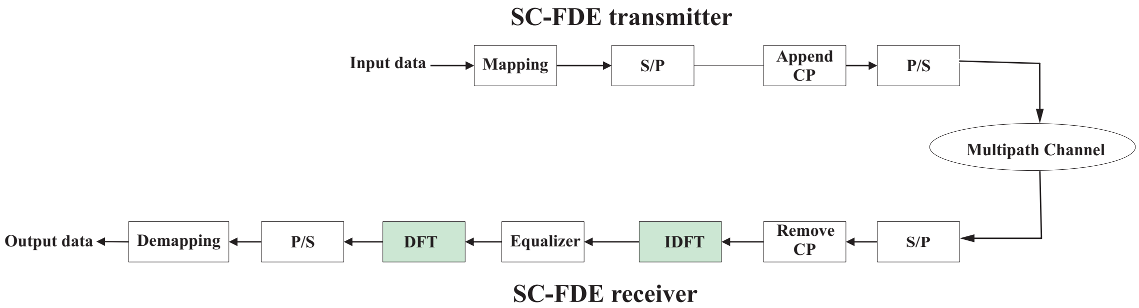


Figure 3.3: Block diagram of baseband SC-FDE system model .

The received SC signal can be converted and equalized in the frequency domain using IDFT and DFT operations. This setup is known as SC modulation with FDE, as illustrated in Fig. 3.3. The use of SC modulation and FDE by processing the DFT of the received signal has several attractive features:

- SC modulation has a low peak-to-average ratio compared with OFDM, allowing the use of less costly power amplifiers.
- Its performance with FDE is similar to that of OFDM, even for very long channel delay spread.
- Frequency domain receiver processing has a similar complexity reduction advantage to that of OFDM: complexity is proportional to the log of multipath spread.
- Coding, while desirable, is not necessary for combating frequency selectivity as it is in non-adaptive OFDM. In fact, SC schemes can have acceptable uncoded performance, contrary to OFDM schemes that require suitable channel coding schemes.
- SC modulation is a well-proven technology in many existing wireless and wireline applications, and its RF system linearity requirements are well-known.

3.2.1 Advanced techniques in receiver design

Designing a receiver for SC-FDE is not an easy task, as it requires addressing performance and complexity issues. Most conventional linear receivers, such as MMSE and ZF, are adapted for detection in SC-FDE systems, but they require matrix inversion. In the majority of scenarios, a linear FDE is implemented at the receiver, despite the fact that it has been demonstrated that nonlinear (NL) equalizers can achieve considerably superior performance compared to linear equalizers. Therefore, it is advantageous to develop NL equalizers for SC-FDE systems. Although nonlinear receivers outperform linear ones, they are prone to error propagation, particularly with long feedback filters. To mitigate this issue, an Iterative FDE (IFDE) technique for SC-FDE, known as Iterative Block Decision Feedback Equalizer (IB-DFE), was introduced in [171]. This method was subsequently expanded to cover diversity scenarios and layered space-time schemes [172].

3.3 System and signal characterization

3.3.1 System model

We consider a massive MIMO system employing an SC-FDE modulation in the uplink transmission where a BS with R antennas concurrently serves T single antenna user (the generalization for multi-antenna UEs is straightforward), and we assume $R \geq T$ (typically, $R >> T$). The total time domain block data to be transmitted is defined as $\mathbf{x} = [x_0, x_1, \ldots, x_{N-1}]$ and the associated data with t transmitted vector can be expressed as $\mathbf{x}_n = [x_{n,1}, x_{n,2}, ..., x_{n,T}]^T$ where $n = 1, 2, \ldots, N$ and N are the block data of the symbols. The discrete Fourier transform (DFT) of the time-domain block associated with the t-th transmit antenna $[x_{0,t}, x_{1,t}, ..., x_{N-1,t}]$ yields the corresponding frequency-domain block $[X_{0,t}, X_{1,t}, ..., X_{N-1,t}]$, which are aggregated in the frequency-domain vector $\mathbf{X} = [X_0, X_1, \ldots, X_{N-1}]$ where the k-th column $\mathbf{X}_k = [X_{k,1}, X_{k,2}, ..., X_{k,T}]^T$ denotes the size-T vector with the set of frequency-domain samples for the k-th subcarrier.

In this section, we consider the use of generic M-QAM constellations. This means that the symbols $x_{n,t}$ to be transmitted by the t-th antenna are selected from a specific alphabet $\mathfrak{S}$ that has a dimension of M, i.e., $x_{n,t} \in \mathfrak{S}$. The symbols from $\mathfrak{S}$ are being chosen following the bits that correspond to the mapping $x_{n,t} = f(\beta_{n,t}^{(1)}, \beta_{n,t}^{(2)}, ..., \beta_{n,t}^{(\mu)})$, with $\mu = \log_2(M)$ denote the number of bits per constellation symbol. In that case, the modulation M-QAM can be presented as the combination of two $\sqrt{M}$-Pulse Amplitude Modulation ($\sqrt{M}$-PAM) constellations, each with $\mu/2$ bits per symbol. This means that, for square constellations, the alphabet $\mathfrak{S}$ can be regarded as the vector combination of two subsets, $\mathfrak{S}^I$ and $\mathfrak{S}^Q$, and the symbols of a transmitter can be expressed as

$$x_{n,t} = x_{n,t}^I + jx_{n,t}^Q, \tag{3.6}$$

where $x_{n,t}^I \in \mathfrak{S}^I$ and $x_{n,t}^Q \in \mathfrak{S}^Q$ are the in-phase and the quadrature symbol's components, respectively.

It can be shown (see [173]) that each constellation symbol can be written as a linear function of the corresponding bits, i.e.,

$$x_{n,t} = \sum_{\nu=0}^{M-1} g_\nu \prod_{l=0}^{\mu-1} (b_n^{(l)})^{\Omega_{l,\nu}}, \tag{3.7}$$

where $(\Omega_{\mu-1,\nu}, \Omega_{\mu-2,\nu}, ..., \Omega_{1,\nu}, \Omega_{0,\nu})$ is the binary representation of ν (with μ bits) and $b_n^{(l)} = 2\beta_n^{(l)} - 1$ are the bits in the polar form, i.e., $\beta_n^{(l)}$= 0 or 1 for $b_n^{(l)}$=-1 or 1, respectively (if we decompose the QAM constellation in two PAM constellations, we have two decompositions similar to (3.7), each with sums up to $\sqrt{M}$ terms and products of up to $\mu/2$ bits). The coefficients g_ν can be obtained from the inverse Hadamard transform of the vector of constellation points [174]. Although a general size-M constellation (3.7) might require up to M terms, the number of non-zero terms in many practical constellations can be much lower. For instance, for conventional 16-QAM with Gray mapping, we have

$$x_{n,t} = 2b_{n,t}^{(2)} + b_{n,t}^{(1)} b_{n,t}^{(2)} + j(2b_{n,t}^{(4)} + b_{n,t}^{(3)} b_{n,t}^{(4)}), \tag{3.8}$$

and for conventional 64-QAM with Gray mapping, we have

$$\begin{aligned} x_{n,t} = {} & 4b_{n,t}^{(3)} + 2b_{n,t}^{(3)} b_{n,t}^{(2)} + b_{n,t}^{(3)} b_{n,t}^{(2)} b_{n,t}^{(1)} \\ & + j(4b_{n,t}^{(6)} + 2b_{n,t}^{(6)} b_{n,t}^{(5)} + b_{n,t}^{(6)} b_{n,t}^{(5)} b_{n,t}^{(4)}). \end{aligned} \tag{3.9}$$

Each block is padded with an appropriate CP longer than the overall channel impulse response associated with the link between each transmit and receive antenna to deal with the frequency-selective wireless channel. The CP samples are removed at the receiver, and the resulting time-domain block associated with each receive antenna is submitted to a DFT operation, leading to the frequency domain

$$\mathbf{Y}_k = \mathbf{X}_k \mathbf{H}_k + \mathbf{W}_k, \tag{3.10}$$

where $\mathbf{Y}_k = [Y_{k,1}, Y_{k,2}, ..., Y_{k,R}]^T \in \mathbb{C}^{R\times 1}$ is the received vector signal, with $Y_{k,r}$ denoting the signal received by the r-th antenna, $\mathbf{W}_k = [W_{k,1}, W_{k,2}, ..., W_{k,R}]^T \in \mathbb{C}^{R\times 1}$ denotes the AWGN vector with entries $\mathcal{CN}(0, \sigma_{W_k}^2)$. $\mathbf{H}_k \in \mathbb{C}^{R\times T}$ denotes the channel matrix between the T transmit antennas and R receive antennas for the k-th sub-carrier (which is assumed invariant during the transmission of a giving block).

3.3.2 Iterative block decision-feedback equalizer

When we employ an IB-DFE as shown in Fig. 3.4 in (a), the frequency domain samples at the output of the FDE at the i-th iteration are given by [175]

$$\tilde{\mathbf{X}}_k^{(i)} = \mathbf{F}_k^{(i)} \mathbf{Y}_k - \mathbf{B}_k^{(i)} \hat{\mathbf{X}}_k^{(i-1)}, \text{where } k = 1, 2, ..., N, \tag{3.11}$$

where $\hat{\mathbf{X}}_k^{(i-1)}$ denotes the frequency domain data estimates (i.e., $\{\hat{X}_{0,t}^{(i)}, \hat{X}_{1,t}^{(i)}, ..., \hat{X}_{N-1,t}^{(i)}\}$ is the DFT of the block $\{\hat{x}_{0,t}^{(i)}, \hat{x}_{1,t}^{(i)}, ..., \hat{x}_{N-1,t}^{(i)}\}$, with $\hat{x}_{n,t}^{(i)}$ denoting "hard decision" estimate of $x_{n,t}$ from previous FDE iteration; for the first iteration they are assumed to be 0). $\mathbf{F}_k^{(i)} \in \mathbb{C}^{T\times R}$ denotes the feedforward coefficients associated with the i-th iterations numbers and $\mathbf{B}_k^{(i)} \in \mathbb{C}^{T\times T}$ denotes the feedback coefficients. The feedforward and feedback coefficients $\mathbf{F}_k^{(i)}$ and $\mathbf{B}_k^{(i)}$, respectively, are optimized under the MMSE criterion, which is known to be able to have good performance by minimizing the overall noise at the detection level (which includes the Gaussian noise term, the residual ISI and the residual inter-user interference. The computation of the optimum coefficients is described in the next section.

3.3.3 Optimization of feedforward and feedback coefficients

To obtain the optimal values for the coefficients $\mathbf{F}_k^{(i)}$ and $\mathbf{B}_k^{(i)}$ of the i-th iteration, we will proceed as follows. The mean squared error (MSE) of the frequency-domain samples associated with the k-th subcarrier and the set of t transmit antennas at iteration i is given by [176]

$$\Theta_{k,t}^{(i)} = \mathbb{E}\left[|\hat{X}_{k,t}^{(i)} - X_{k,t}|^2\right]. \tag{3.12}$$

Accordingly, to minimize the MSEs of all transmitters concurrently, we need to minimize the sum of all MSEs, that is,

$$\begin{aligned} \min \Theta_k^{(i)} &= \min \sum_{t=1}^{T} \Theta_{k,t}^{(i)} \\ &= \min \sum_{t=1}^{T} \mathbb{E}\left[|\ \hat{X}_{k,t}^{(i)} - X_{k,t}\ |^2\right] \\ &= \mathrm{Tr}\left\{\mathbb{E}\left[(\hat{\mathbf{X}}_k^{(i)} - \mathbf{X}_k)(\hat{\mathbf{X}}_k^{(i)} - \mathbf{X}_k)^H\right]\right\}, \end{aligned} \tag{3.13}$$

where $\Theta_{k,t}^{(i)}$ denotes the MSE for each transmit antenna at the k-th subcarrier. To simplify the equation (3.13) and from the IB-DFE structure, we have

$$\begin{aligned} \Theta_k^{(i)} = \mathrm{Tr}\Big\{\mathbb{E}\big[&(\mathbf{F}_k^{(i)}\mathbf{Y}_k - \mathbf{B}_k^{(i)}\hat{\mathbf{X}}_k^{(i-1)} - \mathbf{X}_k) \\ &(\mathbf{F}_k^{(i)}\mathbf{Y}_k - \mathbf{B}_k^{(i)}\hat{\mathbf{X}}_k^{(i-1)} - \mathbf{X}_k)^H\big]\Big\}. \end{aligned} \tag{3.14}$$

To ensure that the receiver output is unbiased, we add the normalizing constraints [176]

$$\frac{1}{N}\sum_{k=0}^{N-1}[\mathbf{F}_k^{(i)}\mathbf{H}_k]_{t,t} = 1, \text{where } t = 1, 2, ..., T. \tag{3.15}$$

Therefore, we have the optimization problem

$$\begin{aligned} \min_{\mathbf{F}_k^{(i)},\mathbf{B}_k^{(i)}} \quad & \Theta_k^{(i)} \\ \text{s.t.} \quad & \frac{1}{N}\sum_{k=0}^{N-1}[\mathbf{F}_k^{(i)}\mathbf{H}_k]_{t,t} = 1, \text{where } t = 1,2,...,T. \end{aligned} \tag{3.16}$$

The solution of (3.16) is shown in Appendix B.1 and is given by

$$\mathbf{F}_k^{(i)} = \mathbf{\Gamma}_k^{(i)}\left(\frac{\sigma_{W_k}^2}{\sigma_{X_k}^2}\mathbf{I}_T + \mathbf{H}_k^H\mathbf{H}_k(\mathbf{I}_T - \mathbf{\Upsilon}_k^{(i-1)2})\right)^{-1}\mathbf{H}_k^H, \tag{3.17}$$

and

$$\mathbf{B}_k^{(i)} = \mathbf{F}_k^{(i)}\mathbf{H}_k - \mathbf{I}_T, \tag{3.18}$$

where $\mathbf{\Gamma}_k^{(i)} = diag(\gamma_{k,1}^{(i)}, \gamma_{k,2}^{(i)}, ..., \gamma_{k,T}^{(i)})$ are the normalizing conditions to meet (3.15) [177], $\sigma_{X_k}^2$, and $\sigma_{W_k}^2$ are the signal and noise variances, respectively, and $\mathbf{\Upsilon}_k^{(i-1)}$ is given by [174]

$$\mathbf{\Upsilon}_k^{(i-1)} = \frac{E[\hat{\mathbf{x}}_n^{(i-1)}\mathbf{x}_n^H]}{E[|\mathbf{x}_n|^2]} = \frac{E[\hat{\mathbf{X}}_k^{(i-1)}\mathbf{X}_k^H]}{E[|\mathbf{X}_k|^2]}. \tag{3.19}$$

Since the data streams associated with the different transmit antennas are uncorrelated, for high SNR (and the optimum $\mathbf{F}_k$), $\mathbf{\Upsilon}_k^{(i-1)}$ is approximately diagonal. The t-th element of its diagonal corresponds to the reliability of the estimates of the t-th data stream at the i-th iteration. For the first iteration, since there is no information about $\hat{\mathbf{X}}_k^{(i)}$, we have $\mathbf{\Upsilon}_k^{(0)} = \mathbf{0}_{T\times T}$, and $\mathbf{B}_k^{(1)} = \mathbf{0}_{T\times T}$ in (3.17), leading to

$$\mathbf{F}_k^{(1)} = \left(\alpha\mathbf{I}_T + \mathbf{H}_k^H\mathbf{H}_k\right)^{-1}\mathbf{H}_k^H, \tag{3.20}$$

where $\alpha = \frac{\sigma_{W_k}^2}{\sigma_{X_k}^2}$. Then, the first iteration of $\tilde{\mathbf{X}}_k^{(1)}$

$$\tilde{\mathbf{X}}_k^{(1)} = \mathbf{F}_k^{(1)}\mathbf{Y}_k, \text{where } k = 1,2,...,N. \tag{3.21}$$

After the first iteration, the feedback coefficients can be applied to reduce a significant part of the residual interference (for higher reliability of the data estimates, we remove more residual interference). Then, after several iterations and for a moderate to high signal-to-noise ratio (SNR), the correlation coefficient will be $\mathbf{\Upsilon}_k^{(i-1)} \approx \mathbf{I}_T$, and the residual ISI will be almost totally canceled. The IB-DFE receivers work well with different SNRs. For low SNR, the IB-DFE receiver reduces to a linear FDE. For moderate-to-high SNR, it reduces interference according to the reliability of its estimated transmitted signals, avoiding error propagation. At soft detector IB-DFE as shown in Fig. 3.4 in (b), The general characterization of the constellation symbols as a function of the corresponding bits given by (3.7) can be employed to obtain the average symbol conditioned to the FDE output and correlation coefficient used by the IB-DFE iterations [174]. For this purpose, we start by obtaining the log-likelihood ratio (LLR) of the l-th bit of the n-th transmitted symbol of the t-th user given by

$$\begin{aligned} a_{n,t}^{(l)} &= \log\left(\frac{Pr(\beta_{n,t}^{(l)} = 1|\tilde{x}_{n,t})}{Pr(\beta_{n,t}^{(l)} = 0|\tilde{x}_{n,t})}\right), \\ &= \log\left(\frac{\sum_{x_{n,t}\in\boldsymbol{\Psi}_1^{(l)}} \exp(-\frac{|\tilde{x}_{n,t}-x_{n,t}|^2}{2\sigma_t^2})}{\sum_{x_{n,t}\in\boldsymbol{\Psi}_0^{(l)}} \exp(-\frac{|\tilde{x}_{n,t}-x_{n,t}|^2}{2\sigma_t^2})}\right), \end{aligned} \tag{3.22}$$

where $\tilde{x}_{n,t}$ denotes the output of the FDE (i.e., $\{\tilde{x}_{0,t}^{(i)}, \tilde{x}_{1,t}^{(i)}, ..., \tilde{x}_{N-1,t}^{(i)}\}$, is the IDFT of the block $\{\tilde{X}_{0,t}^{(i)}, \tilde{X}_{1,t}^{(i)}, ..., \tilde{X}_{N-1,t}^{(i)}\}$), $\boldsymbol{\Psi}_1^{(l)}$ and $\boldsymbol{\Psi}_0^{(l)}$ are the subsets of $\mathfrak{S}$ set of constellations symbols, where $\beta_{n,t}^{(l)} = 1$ or 0, respectively (clearly $\boldsymbol{\Psi}_1^{(l)} \cup \boldsymbol{\Psi}_0^{(l)} = \mathfrak{S}$ and $\boldsymbol{\Psi}_1^{(l)} \cap \boldsymbol{\Psi}_0^{(l)} = \emptyset$). Since the approximation is defined as

$$\log(\sum_j \exp(-x_j^2)) \approx \max_j(-x_j^2). \tag{3.23}$$

(see [178]), then, the bit LLRs can be approximated by

$$a_{n,t}^{(l)} \approx \frac{1}{\sigma_t^2}\Big(\min_{x_{n,t}\in\boldsymbol{\Psi}_1^{(l)}} |\tilde{x}_{n,t} - x_{n,t}|^2 - \min_{x_{n,t}\in\boldsymbol{\Psi}_0^{(l)}} |\tilde{x}_{n,t} - x_{n,t}|^2\Big), \tag{3.24}$$

where σ_t^2 can be estimated by

$$\sigma_t^2 = \frac{1}{2}\mathbb{E}[|\tilde{x}_{n,t} - x_{n,t}|^2] \approx \frac{1}{2N}\sum_{n=0}^{N-1}[|\tilde{x}_{n,t} - \hat{x}_{n,t}|^2]. \tag{3.25}$$

By using the approximation in (3.23), which is accurate when the SNR is not too small, we avoid computing exponential and logarithmic functions to reduce the complexity. From the LLR values, we can obtain the average bit values conditioned to the FDE output (for a given IB-DFE iteration), which are given by [174]

$$\hat{b}_{n,t}^{(l)} = \tanh\left(\frac{a_{n,t}^{(l)}}{2}\right). \tag{3.26}$$

Using (3.7), the average symbol values conditioned to the FDE output are computed as

$$\hat{x}_{n,t} = \sum_{\nu=0}^{M-1} g_\nu \prod_{l=0}^{\mu-1}(\hat{b}_{n,t}^{(l)})^{\Omega_{l,\nu}}. \tag{3.27}$$

Next, the reliability of the symbol estimates that are used in the feedback loop can finally be determined by [173]

$$\Upsilon_{n,t}^{(i-1)} = \frac{E[\hat{x}_{n,t}^{(i-1)} x_{n,t}^H]}{E[|x_{n,t}|^2]} = \frac{\sum_{\nu=0}^{M-1} |g_\nu|^2 \prod_{l=0}^{\mu-1} |\hat{b}_{n,t}^{(l)}|^{\Omega_{l,\nu}}}{\sum_{\nu=0}^{M-1} |g_\nu|^2}, \tag{3.28}$$

and the overall block reliability is[1]

$$\bar{\Upsilon}_t^{(i-1)} = \frac{1}{N}\sum_{n=0}^{N-1}\Upsilon_{n,t}^{(i-1)}. \tag{3.29}$$

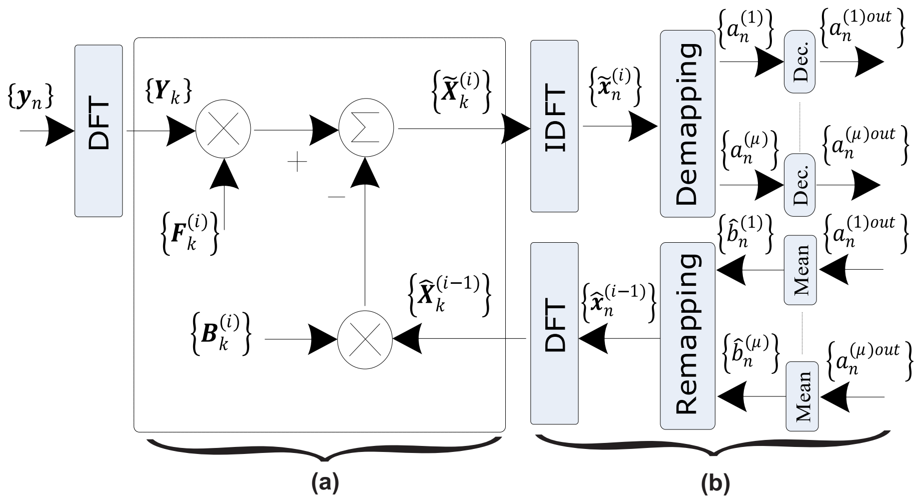


Figure 3.4: IB-DFE receiver structure with soft detector.

3.3.4 Complexity issues

In linear ZF and MMSE-based receivers, the complexity lies in the matrix inversion needed for each frequency, and the dimensions of these matrices can be very high in massive MIMO systems. For the IB-DFE receiver, the complexity of equalization primarily arises from matrix inversions required to calculate the feedforward coefficients $\mathbf{F}_k$ used in the process. It is important to note that even in non-iterative FDE, the computation of $\mathbf{F}_k$ necessitates an inverse matrix operation. Due to these matrix inversions, the IB-DFE receiver exhibits a computational complexity of $O(N^3)$.

To mitigate this issue, two low-complexity iterative frequency-domain receivers, namely Maximum Ratio Detector (MRD) and Equal Gain Detector (EGD), are proposed [179]. The MRD frequency-domain receiver processes each frequency based on the term $\mathbf{H}_k\mathbf{Y}_k$, and takes advantage of the fact that

$$\mathbf{H}_k^H\mathbf{H}_k \approx R\mathbf{I}, \tag{3.30}$$

which is accurate when $R \gg 1$ and the channels between different transmit and receive antennas have small correlation. As stated in [179], the proposed IB-DFE iterative receiver

[1]For QAM constellations where we have separate PAM constellations in the real and imaginary parts, we could obtain the reliabilities of the PAM symbols associated to them, $\Upsilon_{n,t}^I$ and $\Upsilon_{n,t}^Q$, and average them, leading to $\Upsilon_t^{(i-1)} = \frac{1}{2N}\sum_{n=0}^{N-1}(\Upsilon_{n,t}^{I(i-1)} + \Upsilon_{n,t}^{Q(i-1)})$.

based on the MRD receiver is characterized by

$$\tilde{\mathbf{X}}_k = \mathbf{\Gamma}_k \mathbf{H}_k \mathbf{Y}_k - \mathbf{B}_k \mathbf{X}_k, \tag{3.31}$$

The matrices $\mathbf{B}_k$ are employed to remove the residual ISI and inter-user interference. Then, the feedback coefficient can be presented as

$$\mathbf{B}_k = \mathbf{\Gamma}_k \mathbf{H}_k \mathbf{H}_k - \mathbf{I}. \tag{3.32}$$

Another possible implementation is the EGD receiver. This receiver takes advantage of the fact that for massive MIMO systems with $R \gg 1$ and small correlation between the channels associated with different transmit and receive antennas, the elements outside the main diagonal of

$$\mathbf{A}_k \mathbf{H}_k, \tag{3.33}$$

are much lower than the ones at its diagonal, where the $(q, q')^{\text{th}}$ element of the matrix $\mathbf{A}$ is

$$[\mathbf{A}]_{q,q'} = \exp(j2([\mathbf{A}]_{q,l})), \tag{3.34}$$

i.e., they have an absolute value of 1 and a phase identical to the corresponding element of the matrix $\mathbf{H}$. For SC-FDE, we could employ a frequency-domain receiver with EGC at each frequency, based on $\mathbf{A}_k \mathbf{Y}_k$. However, the residual interference levels can still be substantial, especially for moderate values of R/T. The proposed IB-DFE iterative receiver based on the EGC receiver is characterized by

$$\tilde{\mathbf{X}}_k = \mathbf{\Gamma}_k \mathbf{A}_k \mathbf{Y}_k - \mathbf{B}_k \mathbf{X}_k, \tag{3.35}$$

Therefore, the EGD consists of a frequency-domain receiver based on the input for $\mathbf{A}_k^H \mathbf{H}_k \mathbf{Y}_k$. The only difference from the MRD receiver is the matrix $\mathbf{B}_k$, whose values are calculated as

$$\mathbf{B}_k = \mathbf{\Gamma}_k \mathbf{A}_k^H \mathbf{H}_k - \mathbf{I}. \tag{3.36}$$

As we shall see, this receiver has similar or even better performance than the MRD iterative receiver when the number of communicating antennas is small. Although a few IB-DFE receivers like MRC [179] and Equal Gain Combining (EGC) [180] do not require matrix inversions, these methods exhibit poor detection performance for higher-order constellations. They are only suitable for massive MIMO systems where the constellation size is small and the number of receiving antennas is much larger than the number of users [181].

3.4 Summary

In this chapter, we review the differences between OFDM and SC-FDE. One key advantage of SC-FDE is its lower PAPR in the transmitter, making it very suitable for wireless communication. However, when SC-FDE is employed with massive MIMO, it suffers from increased complexity as the channel size increases. Although some solutions have been proposed by other researchers, the performance still falls short of the target. In the next chapter, we will address this issue and present techniques to reduce the complexity of detection.

CHAPTER

4 Efficient Iterative Massive MIMO Detection Using Chebyshev Acceleration

Berra, S., Dinis, R., Rabie, K., & Shahabuddin, S. (2022). Efficient iterative massive MIMO detection using Chebyshev acceleration. Physical Communication, 52, 101651.

4.1 Introduction

The BS of a massive MIMO system is equipped with hundreds of antennas to serve tens or hundreds of users in the same time-frequency resources, as well as providing substantial improvements in link reliability, spectral efficiency, energy efficiency and interference reduction [32, 67]. Despite all the advantages, massive MIMO suffers from a high computational complexity associated with a large number of antennas. A major challenge for baseband processing of massive MIMO originates from symbol vector detection in the uplink (UL) receiver [30] due to the substantially increased channel dimensions and multi-user interference. For this reason, to reduce the BSs' complexity and cost, it is required to relax further the computational complexities related to linear schemes. The optimum detection algorithm in multi-user scenarios is the ML detector. However, it suffers from exponential complexity in terms of the number of users (and the constellation size), which make the ML detector impractical for a large number of antennas [182]. Some nonlinear detectors such as fixed-complexity sphere decoding (SD) [183], tabu search [184], and the belief propagation (BP) algorithms [185] can be found in the literature. Although these solutions can approach the optimum performance ML, the computational complexity is still extremely high, precluding its use when the channel dimensions is very high, as in the massive MIMO case, especially for high modulation orders. Therefore, linear detection methods, such as ZF and MMSE are the natural choice due to the good performance/complexity tradeoffs. However, for massive MIMO systems even linear detectors can be too complex because they require the inversion of large matrices. For example, a 16-user massive MIMO system requires a 16×16 matrix inversion. It should be noted that linear detectors are near-optimal for certain massive MIMO configurations, namely when the ratio of numbers between BS antennas and users is high. Due to channel hardening, the uncorrelated noise and the small-scale fading can be eliminated. As a result, a linear signal detector such ZF and MMSE can achieve near-optimal performance. The linear detectors are easily implementable for a small-scale MIMO system but suffers from high complexity in a massive MIMO setting due to the high computational complexity associated with the matrix inversion when the number of users is high.

Therefore, a new class of detectors based on approximate matrix inversion have become popular over the past decade. For example, a Neumann series approximation scheme was proposed in [186] to convert the matrix inverse operation into a series of matrix-vector multiplications by utilizing the diagonal matrix. The successive over-relaxation (SOR) has been proposed in [187] to balance the tradeoff of performance and complexity by an optimal relaxation parameter to achieve better performance and avoid the inversion matrix iteratively, but still need larger iterations to achieve the same performance. The GS method basd on the detection signal proposed in [115] with a proper initial solution to attain the target performance with low computational complexity, but the internal parts of each GS iteration render parallel execution complicated, and hence, the computational complexity is raised. The error recovery-based detector [188], Richardson iteration [110] and conjugate gradient [189] are also iterative detection methods, which utilize the initial solution to achieve near-MMSE performance without a matrix inversion, but they suffer from SNR loss. Interested readers are invited to read through [190] to know more about the approximate inversion based detectors. The message passing-based detection (MPD) [191] used to simplify the calculation of matrix-inverse in massive MIMO system, but it contains many exponential computations caused by increasing the number of users and the order of modulation. In [192] and [193], the Chebyshev iteration and Newton iteration (NI) have been proposed to provide fast convergence while their complexities depend on the number of iterations. Furthermore, these iterative methods require a complex calculation of initial input to ensure convergence. However, designing efficient methods for accelerating the iterative method, which are needed to improve the convergence and reduce the complexity, can be challenging [95, 194]. The Chebyshev acceleration [95] uses a polynomial, involving a new sequence vector from linear combinations obtained from the basic iterative method, to achieve faster convergence. Therefore, Chebyshev acceleration based on SOR and symmetric successive over-relaxation (SSOR) approaches have been proposed to achieve near optimal detection for massive MIMO systems [194, 195]. In [196], a modified accelerated overrelaxation (MAOR) method is proposed for iterative solution of nonsingular linear systems, which is a generalization of accelerated overrelaxation (AOR) [82, 197]. MAOR essentially develops a new splitting matrix which depends on M-splittings of a singular matrix [198, 199]. The convergence of the MAOR method can be controlled by the selection of its acceleration and relaxation parameters. Therefore, the splitting matrix based on MAOR is able to present a polynomial acceleration such as Chebyshev. Also, the small spectral radius of MAOR makes the convergence rate better than conventional AOR and SOR when using acceleration and relaxation parameters [196]. In contrast to the studies above, in this chapter, we propose a low-complexity massive MIMO signal detection algorithm based on MAOR method [196]. The MAOR based iterative detection is improved further by deriving optimal values for two key parameters of the original MAOR algorithm, i.e., the acceleration and relaxation parameters. In addition, an eigenvalue based initial solution is applied instead of utilizing the diagonal matrix, to achieve a better convergence rate for the MAOR. Finally, the Chebyshev polynomial acceleration method is applied to MAOR iterations to reach near-MMSE performance. It is shown that Chebyshev acceleration has a relatively simple structure with the aid of error analysis.

The resulting Chebyshev accelerated MAOR, which we refer to as Chebyshev-MAOR, provides near MMSE performance with reduced complexity by an order of magnitude. For the sake of comparison, the performance of the Chebyshev-MAOR is compared with state-of-the-art massive MIMO detectors for similar channel conditions, iteration numbers, and MIMO configurations. The contributions of this chapter are as follows

- A novel low-complexity efficient data detection algorithm is proposed based on MAOR method iteration, which enables near-MMSE error rate performance for uplink M-MIMO systems.
- We present the optimal values of two key parameters of MAOR and develop a suitable and less complex initial solution to make the proposed algorithm more appealing for realistic implementations. This has shown that it achieves the fastest convergence of existing counterparts.
- We demonstrate that by using the Chebyshev acceleration technique, the proposed algorithm overcomes the susceptibility of the traditional MAOR detector to relaxation parameters while significantly accelerating the convergence rate.
- The computational complexity of the aforementioned various approaches are examined and numerical models are used to explain efficiency disparities.

4.2 Low-complexity signal detection for UL massive MIMO

In this section, we present the MAOR based massive MIMO detection algorithm. To accelerate the convergence rate and reduce the computational complexity, suitable values of acceleration and relaxation parameters are identified. The spectral radius of the Jacobi matrix is exploited to achieve a high convergence rate. In addition, a method for selecting the initial solution is proposed.

Finally, the Chebyshev polynomial acceleration is utilized to further improve the convergence rate.

4.2.1 Itearative detection based on the MAOR

The estimated signal in (2.6) can be obtained by the MAOR because of the symmetric positive definite (SPD) property. It can efficiently solve the linear system without exact computation of matrix inversion. Since the matrix $\mathbf{W}$ is SPD, it can be decomposed to a diagonal component $\mathbf{D}$, a strictly lower triangular component $\mathbf{L}$, and a strictly upper triangular $\mathbf{U}$. Then, $\mathbf{W}$ can be decomposed as:

$$\mathbf{W} = \mathbf{D} - \mathbf{L} - \mathbf{U}, \tag{4.1}$$

The MAOR is based on M-splittings nonsingular [199] as follows

$$\mathbf{W} = \mathbf{M} - \mathbf{P}, \tag{4.2}$$

where the singular M-matrix and P-Matrix are presented as

$$\mathbf{M} = \frac{1}{\alpha}(\mathbf{D} - \beta\mathbf{L}), \mathbf{P} = \frac{1}{\alpha}((1-\alpha)\mathbf{D} + (\alpha - \beta)\mathbf{L} + \alpha\mathbf{U}). \tag{4.3}$$

Then, the MAOR iterative method [196] can be presented as

$$\mathbf{x}^{(i+1)} = \mathbf{M}^{-1}\mathbf{P}\mathbf{x}^{(i)} + \mathbf{M}^{-1}\mathbf{y}^{MF}. \tag{4.4}$$

Consequently,

$$\mathbf{x}^{(i+1)} = (\frac{1}{\alpha}(\mathbf{D} - \beta\mathbf{L}))^{-1} \left(\frac{1}{\alpha}((1-\alpha)\mathbf{D} + (\alpha-\beta)\mathbf{L} + \alpha\mathbf{U})\mathbf{x}^{(i)}\right) + \left(\frac{1}{\alpha}(\mathbf{D} - \beta L)\right)^{-1} \mathbf{y}^{MF}, \tag{4.5}$$

where i, α, β and $\mathbf{x}^{(0)}$ are the number of iterations, acceleration parameter, relaxation parameter and initial solution, respectively [196]. For $\alpha = \beta$, we have a modified SOR (MSOR) converge for the some value of α [200]. The modified Gauss-Seidel (MGS) method ($\alpha = \beta = 1$) converges and the Modified Jacobi (MJ) method ($\beta = 0, \alpha = 1$) also converges [200]. The iteration matrix of MAOR, $\mathbf{M}^{-1}\mathbf{P}$ can be expressed as

$$\mathbf{B}_{MAOR} = (\frac{1}{\alpha}(\mathbf{D} - \beta\mathbf{L}))^{-1} \left(\frac{1}{\alpha}((1-\alpha)\mathbf{D} + (\alpha - \beta)\mathbf{L} + \alpha\mathbf{U})\right). \tag{4.6}$$

If the spectral radius of MAOR achieves $\rho(\mathbf{B}_{MAOR}) < 1$ or $\lim_{t\to\infty} \mathbf{B}^{t}_{MAOR} = 0$, the iterative method will converge. According to random matrix theory [201], the spectral radius of $\mathbf{B}_{MAOR}$ is given by

$$\rho(\mathbf{B}_{MAOR}) = \max |\lambda(\mathbf{B}_{MAOR})|, \tag{4.7}$$

where $\lambda(\mathbf{B}_{MAOR})$ denote the eigenvalues of matrix $\mathbf{B}_{MAOR}$.

It is clear that the MAOR performance and complexity profile depends on α and β. They can be selected based on the spectral radius of Jacobi iteration matrix, $\rho(\mathbf{B}_{Jacobi})$ [196] to accelerate convergence rate. The estimation of the optimum acceleration and relaxation parameters is done in the following subsection.

4.2.2 Optimum parameters

From (4.5), it is clear that the selection of the acceleration and relaxation parameters will influence the speed of convergence rate of the proposed MAOR technique. The eigenvalue λ of MAOR iteration matrix is defined according to [202] as

$$(\lambda + \beta - 1)^2 = (\alpha\lambda + \beta - \alpha)\beta\rho^2(\mathbf{B}_{Jacobi}). \tag{4.8}$$

According to (4.7), to achieve a fast convergence rate of the proposed method, we need to minimize the eigenvalue λ, to get the smallest spectral radius. Therefore, the optimal parameters of the proposed method are [196, 203]

$$\beta^{opt} = \frac{2}{1 + \sqrt{1 - \rho^2(\mathbf{B}_{Jacobi})}}, \tag{4.9}$$

$$\alpha^{opt} = \frac{\rho(\mathbf{B}_{Jacobi})^2 - \rho(\mathbf{B}_{Jacobi})^4}{\rho(\mathbf{B}_{Jacobi})^2(1 - \rho(\mathbf{B}_{Jacobi})^2)}, \tag{4.10}$$

where $\rho(\mathbf{B}_{Jacobi})$ indicates the spectral radius of Jacobi iteration matrix. To find the optimal acceleration α^{opt} and relaxation β^{opt} parameters we need to obtain the spectral radius of the Jacobi method. The Jacobi iterative method can be expressed as

$$\mathbf{x}^{(i+1)} = (\mathbf{I} - \mathbf{D}^{-1}\mathbf{W})\mathbf{x}^{(i)} + \mathbf{D}^{-1}\mathbf{y}^{MF}, \tag{4.11}$$

where $\mathbf{D}$ is diagonal matrix of $\mathbf{W}$ and $\mathbf{I}$ is identity matrix. The iteration matrix of Jacobi is defined as

$$\mathbf{B}_{(Jacobi)} = (\mathbf{I} - \mathbf{D}^{-1}\mathbf{W}), \tag{4.12}$$

and the corresponding spectral radius of MAOR is $\rho(\mathbf{B}_{MAOR}) = \beta^{opt} - 1$ [204]. Since computing the eigenvalues of matrix $\mathbf{W}$ is complicated, an approximation is adopted here. As the number of antennas R and users T increase then the approximate eigenvalues of the matrix can be calculated as

$$\rho(\mathbf{B}_{Jacobi}) = \rho| - \mathbf{D}^{-1}\mathbf{W} - \mathbf{I}_T| = \rho| - \mathbf{D}^{-1}\mathbf{W}| - 1. \tag{4.13}$$

Since $R >> T$, the the elements of diagonal matrix $\mathbf{D}$ will convergence to R. Thus, $\mathbf{W}^{-1} = (\mathbf{H}\mathbf{H}^H)^{-1} = \mathbf{D}^{-1} = \frac{1}{R}\mathbf{I}_T$, and (4.13) becomes

$$\rho(\mathbf{B}_{Jacobi}) = \frac{1}{R}\rho[-\mathbf{W}] - 1. \tag{4.14}$$

Based on the random matrix theory [201], the spectral radius of $\rho(\mathbf{W})$ and the largest and smallest eigenvalues of $\lambda(\mathbf{W})$ can be approximated as

$$\rho(-\mathbf{W}) = |\lambda_{max}(\mathbf{W})| = R\left(1 + \sqrt{\xi}\right)^2. \tag{4.15}$$

$$\rho(-\mathbf{W}) = |\lambda_{min}(\mathbf{W})| = R\left(1 - \sqrt{\xi}\right)^2. \tag{4.16}$$

Finally, by substituting (4.15) and (4.16) into (4.14), the spectral radius of the iteration matrix $\rho(\mathbf{B}_{Jacobi})$ can be approximated as

$$\rho(\mathbf{B}_{Jacobi}) = \overline{\eta} = \left(1 + \sqrt{\xi}\right)^2 - 1, \tag{4.17}$$

$$\rho(\mathbf{B}_{Jacobi}) = \underline{\eta} = \left(1 - \sqrt{\xi}\right)^2 - 1, \tag{4.18}$$

where $\xi = R/T$ is ratio. Since the number of antennas R is much larger than the number of users T, which makes the value of ξ almost constant, and indicates that with the increase of the loading factor $\xi = R/T$, the spectral radius of $\rho(\mathbf{B}_{Jacobi})$ decreases [204]. Thus, a faster convergence rate can be achieved. However, as the spectral radius is related to the value of the matrix's eigenvalues, α and β can be rewritten as

$$\alpha^{opt} = \frac{2}{1 + \sqrt{1 - |\overline{\eta}|^2}}, \tag{4.19}$$

$$\beta^{opt} = \frac{|\overline{\eta}|^2 - |\underline{\eta}|^4}{|\overline{\eta}|^2\left(1 - |\underline{\eta}|^2\right)}. \tag{4.20}$$

However, the condition

$$0 \leq \alpha \leq \beta, \tag{4.21}$$

should be satisfied. It should be pointed out that the careful selection of the optimum parameters is critical to a high performance and low complexity of the detection.

4.2.3 Approximate initial solution

Even with the the optimal parameters α and β, the proposed method still needs an iterative initial solution which provides a faster and desired result. Theoretically, any moderate initial solution can lead to the convergence of the iterative method, although it might require a substantially higher number of iterations. Since the channel matrix $\mathbf{H}$ is asymptotically orthogonal, for $R > T$ the condition of favorable channel can be described as

$$\frac{\mathbf{h}_j^H \mathbf{h}_t}{R} \rightarrow 0, j \neq t, j, t = 1, 2, ...T, \tag{4.22}$$

where $\mathbf{h}_j$ represents the mth column of $\mathbf{H}$. The inverse of matrix $\mathbf{W}$ is diagonally dominant, so when the ratio of R/T increases, the non-diagonal matrix $\mathbf{W^{-1}}$ and the diagonal one $\mathbf{D}^{-1}$ become increasingly similar, so that matrix $\mathbf{W}$ is a diagonally dominant matrix and satisfies [30] as

$$\mathbf{W}_{(}j, t) = \begin{cases} \frac{\lambda_{max}+\lambda_{min}}{2} & j = t, \\ 0, & j \neq t. \end{cases} \tag{4.23}$$

Therefore, we propose the initial solution proposed as

$$\mathbf{x}^{(0)} = \frac{2}{\lambda_{max} + \lambda_{min}} \mathbf{H}^H \mathbf{y}. \tag{4.24}$$

This initial value enables the proposed iterative method to achieve a faster convergence rate. The computational complexity of the initial value calculation is significantly small if we follow (4.14) and (4.15). In addition, the computation of the initial value can be executed in parallel, which is an important advantage from the hardware implementation point of view.

4.2.4 Convergence analysis

The proposed method achieves a sufficient convergence rate condition when the iteration matrix's spectral radius is less than 1. Additionally, the approximation error is an important element to determine the convergence rate. Then the approximation error of the proposed MAOR can be expressed as

$$\hat{\mathbf{x}}^{(i+1)} - \hat{\mathbf{x}} = \mathbf{R}_{\alpha,\beta}(\hat{\mathbf{x}}^{(i)} - \hat{\mathbf{x}}) = ... = \mathbf{R}_{\alpha,\beta}^{(i+1)}(\hat{\mathbf{x}}^{(0)} - \hat{\mathbf{x}}). \tag{4.25}$$

The approximation error can be expressed as

$$\left\|\hat{\mathbf{x}}^{(i+1)} - \hat{\mathbf{x}}\right\|_2 = \left\|\mathbf{R}_{\alpha,\beta}^{\mathbf{(i+1)}}\right\|_F \left\|\hat{\mathbf{x}}^{(0)} - \hat{\mathbf{x}}\right\|_2 \leq \left\|\mathbf{R}_{\alpha,\beta}\right\|_F^{(i+1)} \left\|\hat{\mathbf{x}}^{(0)} - \hat{\mathbf{x}}\right\|_2. \tag{4.26}$$

The determination of the approximation error of the proposed MAOR detection scheme is affected by the Frobenius norm of the iteration matrix $\mathbf{B}_{\alpha,\beta}$. A smaller $||\mathbf{B}_{\alpha,\beta}||_F$ provides a faster convergence for the proposed method. In massive MIMO, the number of antennas at the BS R grows to infinity with a fixed number of users T and $\beta = 1$, which means that the proposed method MAOR can have a faster convergence rate than Neumann series (NS) which define $2 - \sqrt{2} \leq \alpha \geq \sqrt{2}$.

Hence, we need to show that the Frobenius norms of the iteration matrices of the NS method are larger than that of the proposed MAOR method. It is clear that $||\mathbf{B}_{\alpha,\beta}||_F \geq ||\mathbf{B}_{\mathbf{NS}}||_F$, which $||\mathbf{B}_{\mathbf{NS}}||_F$ represents the iteration matrix of the NS $\mathbf{B}_{\mathbf{NS}} = \mathbf{D}^{-1}(\mathbf{L}+\mathbf{U})$, then the Frobenius norm is [205]

$$||\mathbf{B}_{NS}||_F = ||\mathbf{D}^{-1}(\mathbf{L}+\mathbf{U})|| = \left[\sum_{m=1}^{T}\sum_{m=1}^{T}\left|\frac{\mathbf{w}_{m,m}}{\mathbf{w}_{m,m}}\right|^2\right]^{\frac{1}{2}} = \sqrt{2}||\mathbf{D}^{-1}L||_F, \tag{4.27}$$

where $\mathbf{w}_{m,m}$ denotes the mth column of matrix $\mathbf{W}$ and $||\mathbf{D}^{-1}L||_F = ||\mathbf{D}^{-1}\mathbf{U}||_F$. From the iterative matrix $\mathbf{B}_{\alpha,\beta}$ when $\beta = 1$, we have

$$\begin{aligned}\mathbf{B}_{MAOR} &= (\frac{1}{\alpha}(\mathbf{D}-\mathbf{L}))^{-1}(\frac{1}{\alpha}((1-\alpha)D + (\alpha-1)\mathbf{L} + \alpha\mathbf{U}))\\ &= (\frac{\mathbf{D}^{-1}}{\alpha}(\mathbf{I}_T - \mathbf{D}^{-1}\mathbf{L}))^{-1}(\frac{1}{\alpha}((1-\alpha)\mathbf{D} + \alpha\mathbf{L} - \mathbf{L} + \alpha\mathbf{U}))\\ &= (\frac{1}{\alpha}(\mathbf{I}_T - \mathbf{D}^{-1}L))^{-1}\mathbf{D}^{-1}(\frac{1}{\alpha}((1-\alpha)\mathbf{D} + \alpha\mathbf{L} - \mathbf{L} + \alpha\mathbf{U}))\\ &= (\frac{1}{\alpha}(\mathbf{I}_T - \mathbf{D}^{-1}\mathbf{L}))^{-1}(\frac{1}{\alpha}((1-\alpha) + \mathbf{D}^{-1}\alpha\mathbf{L} - \mathbf{D}^{-1}L + \mathbf{D}^{-1}\alpha\mathbf{U})).\end{aligned} \tag{4.28}$$

The matrix $(\mathbf{I}_T - D^{-1}\mathbf{L}))^{-1}$ can be calculated by the following polynomial expansion

$$(\mathbf{I}_T - \mathbf{D}^{-1}\mathbf{L}))^{-1} = \sum_{t=0}^{\infty}(-1)^t(\mathbf{D}^{-1}\mathbf{L})^t = \mathbf{I}_T + \mathbf{D}^{-1}\mathbf{L} + \sum_{t=2}^{\infty}(-1)^t(\mathbf{D}^{-1}\mathbf{L})^t. \tag{4.29}$$

Assuming a massive MIMO where the number of BS antennas is much larger than UEs T i.e, $R \gg T$, the upper valued L goes to zero, and the diagonal part goes to R, which is defined as $\lim_{t\to\infty}\mathbf{D}^{-1}\mathbf{L} = 0$. Therefore, (4.28) can be approximated as

$$\begin{aligned}\mathbf{B}_{MAOR} &\approx (\frac{1}{\alpha}(\mathbf{I}_T + \mathbf{D}^{-1}\mathbf{L}))(\frac{1}{\alpha}((1-\alpha) + \mathbf{D}^{-1}\alpha\mathbf{L} - \mathbf{D}^{-1}\mathbf{L} + \mathbf{D}^{-1}\alpha\mathbf{U}))\\ &\approx (\frac{1}{\alpha}((\frac{1}{\alpha}((1-\alpha) + \mathbf{D}^{-1}\alpha\mathbf{L} - \mathbf{D}^{-1}\mathbf{L} + \mathbf{D}^{-1}\alpha\mathbf{U})).\end{aligned} \tag{4.30}$$

By using Frobenius norm theory to approximation $\mathbb{R}_{MAOR}$ can be selected as

$$\begin{aligned}||\mathbf{B}_{\alpha,\beta}||_F &= ||(\frac{1}{\alpha}((\frac{1}{\alpha}((1-\alpha) + \alpha\mathbf{L} - \mathbf{D}^{-1}\mathbf{L} + \mathbf{D}^{-1}\alpha\mathbf{D}^{-1}\mathbf{U}))||_F\\ &= \frac{1}{\alpha^2}\left(1 - \alpha + \alpha||\mathbf{D}^{-1}\mathbf{L}||_F - ||\mathbf{D}^{-1}L||_F + \alpha||\mathbf{D}^{-1}L||_F\right)\\ &= \frac{1}{\alpha^2}(1-\alpha) + \frac{1}{\alpha^2}(\alpha - 1 + \alpha)||\mathbf{D}^{-1}\mathbf{L}||_F\\ &= \sqrt{2}(\frac{1}{\alpha^2}(1-\alpha) + \frac{1}{\alpha^2}(\alpha - 1 + \alpha))/2||\mathbf{D}^{-1}\mathbf{L}||_F\\ &= \sqrt{2}(\frac{1}{\alpha^2}(1-\alpha) + \frac{1}{\alpha^2}(\alpha - 1 + \alpha))/2 \leq 1 \Rightarrow 2 - \sqrt{2} \leq \alpha \geq \sqrt{2}.\end{aligned} \tag{4.31}$$

Then, we conclude that $||\mathbf{B}_{\alpha,\beta}||_F \leq ||\mathbf{B}_{NS}||_F$ which shows that the Frobenius norms of the iterative matrix of the proposed method have smallest the spectral radius compared to the iterative method of NS method. To demonstrate that $||\mathbf{B}_{\alpha,\beta}||_F \leq ||\mathbf{B}_{MSOR}||_F$, where $||\mathbf{B}_{MSOR}||_F = (\mathbf{D} - \beta\mathbf{L}^{-1})(\mathbf{L}^H + (1-\beta)\mathbf{L})$, we can prove the following based on the spectral radius of MSOR:

$$\begin{aligned} ||\mathbf{B}_{MSOR}||_F &= (\mathbf{D} - \beta(\mathbf{L}^{-1}(\mathbf{L}^H + (1-\beta)\mathbf{L}) \\ &= \mathbf{D}^{-1}(\mathbf{I} - \beta(\mathbf{L}^{-1})(\mathbf{U} + (1-\beta)\mathbf{L}) \\ &= (\mathbf{I} - \beta(\mathbf{L}^{-1})(\mathbf{D}^{-1}\mathbf{L}^H + (1-\beta)\mathbf{D}^{-1}\mathbf{L}). \end{aligned} \tag{4.32}$$

From (4.29), the spectral radius of MSOR can be simplified as

$$\begin{aligned} \approx \mathbf{D}^{-1}\mathbf{L}^H + (1-\beta)\mathbf{D}^{-1}\mathbf{L}) \approx ||1-\beta||\mathbf{D}^{-1}\mathbf{L}^H + \mathbf{D}^{-1}\mathbf{L} \approx (||1-\beta|| + 1)\mathbf{D}^{-1}\mathbf{L} \\ \approx (||1-\beta|| + 1)||\mathbf{R}_{NS}||_F. \end{aligned} \tag{4.33}$$

In [194], the value of β in MSOR is limited to $1 < \beta < 2$, where MAOR has $0 < \alpha < \beta$, which makes the contain of $(||1-\beta||+1)||$ large. Thus, $||\mathbf{B}_{\alpha,\beta}||_F \leq ||\mathbf{B}_{MSOR}||_F$, as equation (4.33) contains NS which is large spectral radius then MAOR.

It has been demonstrated that the MAOR method converges faster than MSOR and can even converge in situations when the MSOR fails to converge [196]. To further speed up the convergence rate, we will apply the Chebyshev polynomial acceleration method described below.

4.3 Chebyshev acceleration

The convergence rate of our method can be further improved by adopting a polynomial acceleration scheme. Chebyshev is well-known acceleration method, which has been rigorously discussed in [95] and [206, 207]. Suppose, we convert $\mathbf{x}_{MMSE} = \mathbf{W}^{-1}\mathbf{y}^{MF}$ to the iteration following as [95]

$$\begin{aligned} \mathbf{x}^{(i+1)} &= \mathbf{B}_{MAOR}\mathbf{x}^{(i)} + \mathbf{c}, \\ \mathbf{c} &= (\frac{1}{\alpha}(\mathbf{D} - \beta\mathbf{L}))^{-1}\mathbf{y}^{MF}, \end{aligned} \tag{4.34}$$

where $\mathbf{B}_{MAOR} = \mathbf{M}^{-1}\mathbf{P}$ and $\mathbf{c} = \mathbf{P}^{-1}\mathbf{y}^{MF}$. Then, we use (4.34) to give all these approximations of signal vector $\mathbf{x}$. Consequently, the Chebyshev acceleration is good method to accelerate the convergence $(\mathbf{x}^{(i)})_{i=0}^{\infty}$. Based on the signal vector sequence $(\mathbf{x}^{(i)})_{i=0}^{\infty}$, we construct a new linear combination $(\mathbf{y}^{(m)})_{m=1}^{\infty}$, which features faster convergence towards the accurate solution $\mathbf{x}^{(i)}$ of (4.34) and is known as secondary iteration, as

$$\mathbf{y}^{(m)} = \sum_{i=1}^{m} \gamma_{m,i}\mathbf{x}^{(i)}, \tag{4.35}$$

where parameters $\gamma_{m,i}$ can be referenced in polynomials as,

$$p_m(\mathbf{z}) = \sum_{i=1}^{m} \gamma_{m,i}\mathbf{z}^m. \tag{4.36}$$

Note that the scalars $\gamma_{m,i}$ must satisfy $\sum_{i=0}^{m} \gamma_{m,i} = 1$. Since at convergence $\mathbf{x^{(0)}} = \mathbf{x^{(1)}} = ... = \mathbf{x}$ have been generated via the iteration the equation $p_m = \mathbf{x}$ should be applied. Then the error approximations of $\mathbf{x}$ and $\mathbf{y}^{(m)}$ provide as

$$\mathbf{y}^{(m)} - \mathbf{x} = \sum_{i=1}^{m} \gamma_{m,i}(\mathbf{x}^{(i)} - \mathbf{x}) = \sum_{i=1}^{m} \gamma_{m,i}\mathbf{B}^{i}_{MAOR}(\mathbf{x}^{(0)} - \mathbf{x}) = p_m(\mathbf{B}_{MAOR})(\mathbf{x}^{(0)} - \mathbf{x}), \tag{4.37}$$

where $p_m(\mathbf{B}_{MAOR}) = \sum_{i=1}^{m} \gamma_{m,i}\mathbf{B}^{i}_{MAOR}$ is polynomial of degree m with $p_m(1) = 1$, and $p_m(\mathbf{z}) = \sum_{i=1}^{m} \gamma_{m,i}\mathbf{z}^i$

The expression of Chebyshev polynomials are defined by the three-term recurrence relation as

$$T_{m+1}(z) = 2zT_m(z) - T_{m-1}(z), m \geq 1, \tag{4.38}$$

where initial parameters indicates to $T_0(z) = 1$ and $T_1(z) = z$. The Chebyshev polynomials have many interesting properties. A polynomial parameter indicated as

$$p_m(z) = \frac{T_m(z/\rho)}{T_m(1/\rho)}, \tag{4.39}$$

can provide minimum error, while ρ is spectral radius of matrix $\mathbf{B}_{MAOR}$ belonging to $[\lambda_{min}, \lambda_{max}]$. The optimal iteration polynomials $p_m(\mathbf{z})$ can be obtained by the Chebyshev polynomials as

$$p_m(z) = \frac{T_m(1 + 2\frac{z-\epsilon}{\epsilon-\eta})}{T_m(1 + 2\frac{r-\epsilon}{\epsilon-\eta})}, \tag{4.40}$$

where ϵ and η are satisfied as

$$\begin{aligned} \epsilon &= \frac{\lambda_{max} + \lambda_{min}}{\lambda_{max} - \lambda_{min}}, \\ \eta &= \frac{\lambda_{max} + \lambda_{min}}{2}. \end{aligned} \tag{4.41}$$

The ϵ and η give the optimal parameters of the polynomials and are defined by the smallest and largest eigenvalues of the iteration matrix $\mathbf{B}_{MAOR}$, respectively. Here, r is any scalar satisfying the coefficients of polynomials. Let us assume that $\mu_m = 1/T_m(1/\rho)$, so $p_m(\mathbf{R}_{MAOR}) = \mu_m/T_m(\mathbf{B}_{MAOR})/\rho)$, and $\gamma_{0,0} = 1, \gamma_{1,0} = 0, \gamma_{1,1} = 1$. Then, the properties of μ_m becomes

$$\frac{1}{\mu_{m+1}} = 2\frac{1}{\rho\mu_m} - \frac{1}{\mu_{m-1}}, \tag{4.42}$$

To express the Chebyshev polynomial method into the proposed iterative method, we rewrite the equation in (4.3) of $\mathbf{y}^{(m)}$, which also satisfies a three-term-recurrence (see Appendix A.1) as

$$\mathbf{y}^{(m+1)} - \mathbf{x} = 2\frac{\mu_m}{\rho\mu_{m+1}}\mathbf{B}_{MAOR}\mathbf{y}^{(m)} - \frac{\mu_{m-1}}{\mu_{m+1}}\mathbf{y}^{(m-1)} + \mathbf{d}_m. \tag{4.43}$$

From (4.42), the rest of the properties can be defined (see Appendix A.1) as follows.

$$\mathbf{d}_m = 2\frac{\mu_m}{\rho\mu_{m+1}}\mathbf{c}. \tag{4.44}$$

After some mathematical manipulations, the secondary iteration can be defined by

$$\mathbf{y}^{(m+1)} = 2\frac{\mu_m}{\rho\mu_{m+1}}\mathbf{B}_{MAOR}\mathbf{y}^{(m)} - \frac{\mu_{m-1}}{\mu_{m+1}}\mathbf{y}^{(m-1)} + 2\frac{\mu_m}{\rho\mu m+1}\mathbf{c}, \tag{4.45}$$

where $\mathbf{y}^{(0)} = \mathbf{x}^{(0)}$ and $\mathbf{y}^{(1)} = \mathbf{x}^{(1)}$. The Chebyshev polynomial can be easily handled using the MAOR method, which is based on the spectral radius ρ and the iteration matrix $\mathbf{B}_{MAOR}$. We present the complete Chebyshev accelerated MAOR, which we refer to as Chebyshev-MAOR, in Algorithm 1.

Algorithm 1 The Chebyshev-MAOR

Input:$\mathbf{y}, \mathbf{H}, \mathbf{D}, \mathbf{L}, \mathbf{U}, \sigma^2, n, \beta, \alpha$
Output: Estimated signal $\mathbf{y}^{(m)}$
Primary iteration:
$\mathbf{x}^{(i+1)} = (\frac{1}{\alpha}(\mathbf{D} - \beta\mathbf{L}))^{-1}(\frac{1}{\alpha}((1-\alpha)\mathbf{D} + (\alpha-\beta)\mathbf{L} + \alpha\mathbf{U})\mathbf{x}^{(i)}) + (\frac{1}{\alpha}(\mathbf{D} - \beta\mathbf{L}))^{-1}\mathbf{y}^{MF}$.
Initialization:
$\mathbf{x}^{(0)} = \mathbf{y}^{(0)} = \frac{2}{\lambda_{max}+\lambda_{min}}\mathbf{H}^H\mathbf{y}^{MF}$.
$\mathbf{y}^{(1)} = \mathbf{B}_{MAOR}\mathbf{X}^{(i)} + \mathbf{c}$.
$\mu_0 = 1; \mu_1 = \rho$.
$0 \leq \alpha \leq \beta$.
Iteration:
for $i = 1 : m$ **do**
    $\frac{1}{\mu_{m+1}} = 2\frac{1}{\rho\mu_m} - \frac{1}{\mu_{m-1}}$;
    $\mathbf{y}^{(m+1)} = 2\frac{\mu_m}{\rho\mu_{m+1}}\mathbf{R}_{MAOR}\mathbf{y}^{(m)} - \frac{\mu_{m-1}}{\mu_{m+1}}\mathbf{y}^{(m-1)} + 2\frac{\mu_m}{\rho\mu m+1}\mathbf{c}$.
end for
Return $\mathbf{y}^{(m)}$.

4.3.1 Error analysis

The error formula is critical to show the merits of the Chebyshev acceleration method [208]. Assume to the matrix $\mathbf{B} = \mathbf{M}^{-1}\mathbf{P}$ is symmetric and its eigenvalues satisfy

$$-1 < \lambda_{min} < ... < \lambda_{max} < 1. \tag{4.46}$$

Let us consider the error formula defined as

$$\mathbf{e}_i = \mathbf{x_0} - \mathbf{x}, \mathbf{e}_{\mathbf{y}_i} = \mathbf{y}_i - \mathbf{x}, i = 0, 1, 2, ... \tag{4.47}$$

The eigenpair of matrix $\mathbf{B}$ denotes as (λ_i, v_i) which v is an orthonormal basis of the real value matrix of $\mathbf{B}$ since is symmetric. Then $\mathbf{e_0}$ can be expressed as

$$\mathbf{e}_0 = \sum_{i=1}^{m} d_i v_i. \tag{4.48}$$

Let p_m be a polynomial with degree m and (λ_i, v_i) are eigenpair of matrix $\mathbf{B}$ and d_i scalar factor. Then

$$||\mathbf{e_y}^r||^2 = \sum_{i=1}^{m} [d_i^2 p_m^2(\lambda_i)], r = 1, 2, 3, ..., \tag{4.49}$$

noting that

$$\mathbf{x}_i - \mathbf{x} = (\mathbf{M}^{-1}\mathbf{P})^i(\mathbf{x}_0 - \mathbf{x}) = \mathbf{B}^i(\mathbf{x}_0 - \mathbf{x}). \tag{4.50}$$

Therefore, we have

$$\mathbf{y}_r - \mathbf{x} = \sum_{j=1}^{m} \gamma_j(\mathbf{x}_j - \mathbf{x}) = \sum_{j=1}^{m} \gamma_j \mathbf{B}^j(\mathbf{x}_0 - \mathbf{x}) = \mathbf{p}_m(\mathbf{B})(\mathbf{x}_0 - \mathbf{x}). \tag{4.51}$$

Thus,

$$||\mathbf{y}_r - \mathbf{x}||_2 \leq ||\mathbf{p}_m(\mathbf{B})||_2||\mathbf{x}_0 - \mathbf{x}||_2. \tag{4.52}$$

To simplify (4.52), we rewrite the equation as

$$\mathbf{ey}_r = \mathbf{p}_m(\mathbf{B})\mathbf{e}_0, ||\mathbf{ey}_0|| \leq ||\mathbf{p}_m(\mathbf{B})||_2||\mathbf{e}_0||. \tag{4.53}$$

The error estimation in (4.53) is not very accurate. Hence, the structure needs to exploit a relaxed version of (4.48) and (4.51). The error can be expressed as

$$\mathbf{e}_{\mathbf{y}_r} = p_m(\mathbf{B})\mathbf{e}_0 = \mathbf{p}_m(\mathbf{B}) \sum_{i=1}^{m} d_i v_i = \sum_i d_i \mathbf{p}_m(\mathbf{B}) v_i = \sum_i d_i (\sum_j \gamma_j \Lambda_j^j) v_i, \tag{4.54}$$

which leads to

$$||\mathbf{e}_0||^2 = \sum_{i=1}^{m} d_i^2, ||\mathbf{e}_{\mathbf{y}_r}||^2 = \sum_{j=1}^{m} d_i^2 \mathbf{p}_m^2(\lambda_i). \tag{4.55}$$

The improved approximation $\mathbf{y}_i$ then performs the same iteration and refinement to get another improved approximation $\mathbf{y}_{r+1}$,

$$||\mathbf{e}_{\mathbf{y}_{r+1}}||^2 = \sum_{i=1}^{m} [d_i^2 \mathbf{p}_m^4(\lambda_i)], \tag{4.56}$$

where d_i (i=1,2,3,...) are the original values. The above process is repeated r times. Then the error estimation can be indicated as

$$||\mathbf{e}_{\mathbf{y}_\mathbf{r}}||^2 = \sum_{\mathbf{i=1}}^{\mathbf{m}} [\mathbf{d}_\mathbf{i}^2 \mathbf{p}_\mathbf{m}^{\mathbf{2r}}(\lambda_\mathbf{i})], r = 1, 2, 3, ... \tag{4.57}$$

4.4 Computational complexity

In this section, we provide insights into the computational complexity of the proposed method Chebyshev-MAOR. We express the complexity in terms of complex-valued multiplications which is a fairly standard practice in the literature. The initial solution, which is based on a diagonally dominant of the filtering matrix MMSE $\mathbf{W}^{-1}$, requires $T+1$ operations. The overall complexity of the proposed method can be analysed from

$$\begin{aligned}\mathbf{y}_m^{(i+1)} = \frac{1}{\mathbf{w}_{m,m}}\Bigg[\beta\sum_{t=1}^{m+1}\mathbf{w}_{m,t}y_t^{(i+1)} + 2\frac{\mu_m}{\rho\mu_{m+1}}(1-\alpha)\mathbf{w}_{m,m}\mathbf{y}_m^{(i)} \\ + 2\frac{\mu_m}{\rho\mu_{m+1}}(\alpha-\beta)\sum_{t=1}^{m+1}\mathbf{w}_{m,t}y_t^{(i)} + 2\frac{\mu_m}{\rho\mu_{m+1}}\alpha\sum_{t=m+1}^{K}\mathbf{w}_{m,t}\mathbf{y}_t^{(i)} \\ - \frac{\mu_{m-1}}{\mu_{m+1}}\alpha\mathbf{w}_{m,m}\mathbf{y}_m^{(i-1)} + \beta\alpha\sum_{t=1}^{m+1}\mathbf{w}_{m,t}\mathbf{y}_t^{(i-1)} + 2\frac{\mu_m}{\rho\mu_{m+1}}\mathbf{y}^{MF}\Bigg].\end{aligned} \tag{4.58}$$

The required number of multiplications can be calculated as follows: The term outside the square braces is 1 and the terms inside the braces require m, 2, $m+1$, $T-m-1$, 2, m, and 1 operation, respectively. Hence, $\mathbf{y}_m^{i+1}$ requires $T+2m+6$ multiplications per iteration. The total complexity of the Chebyshev-MAOR is $2T^2+7T$ per iteration since m takes values from 1 to T. Hence, the analysis shows that the complexity order of the proposed method is $O(T^2)$, which is lower than the conventional MMSE procedure, which has an $O(T^3)$ complexity. Although other iterative methods can also reduce the complexity to $O(T^2)$, they still require a large number of iterations, which compromises the overall complexity. For example, the modified successive overrexlation (MSOR) and the modified accelerated overrexlation (MAOR) methods satisfy the complexity order which reduces $O(T^3)$ to $O(T^2)$. The NS complexity conducted $O(T^3)$, which more than two iterations (i.e., $n>2$). The proposed method provides a significant reduction in the signal processing complexity due to the smaller required number of iterations.

Table 4.1 shows the overall complexities of the proposed method in addition to other algorithms. The order magnitude of MAOR, MSOR and NS have the some complexity $O(T^2)$. However, the Neumann series based algorithms has $O(T^2)$ only for $i=2$, while the order magnitude growths to $O(T^3)$ when $i\geq 2$.

4.5 Results and discussions

In this section, The error-rate performance of Chebyshev-MAOR, compared with other state-of-the-art detection techniques, is presented. To be more specific, the proposed Chebyshev-MAOR is compared with different variants of MSOR, Chebyshev-MSOR-based detectors [194], Neumann series approximation [186], and exact inversion-based

Table 4.1 Computational complexity comparison

Method	Number of multiplications
MSOR [194]	$i\frac{3}{2}T^2 + i\frac{3}{2}T$
MAOR	$\frac{i}{2}(7T^2 + 3T)$
Chebyshev-MAOR	$i(2T^2 + 7T)$
Neumann Series [186]	$(i-2)T^3 + RT^2 + RT$
Chebyshev-MSOR [194]	$(16T + 8T^2)i$
MMSE	$iT^3 + iT^2$

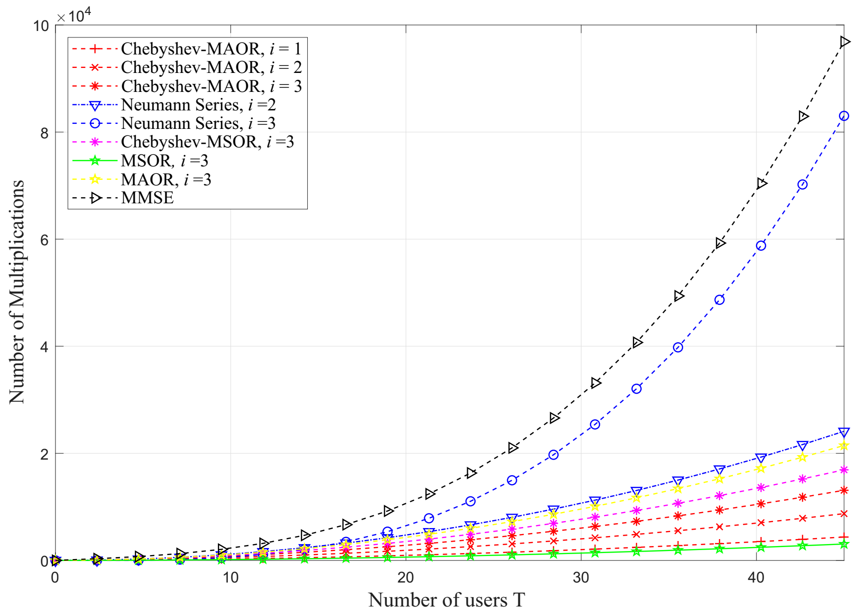
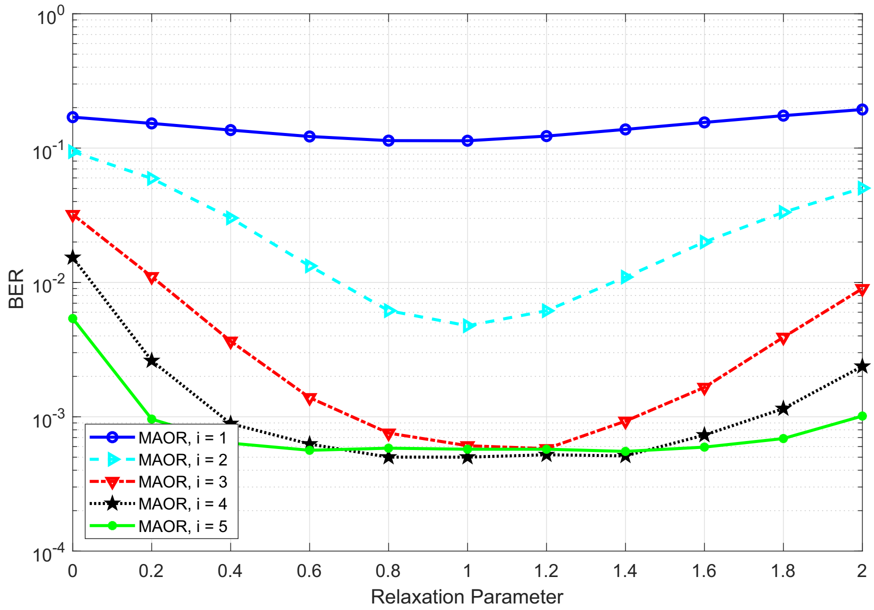


Figure 4.1: Computational complexity versus the number of users.

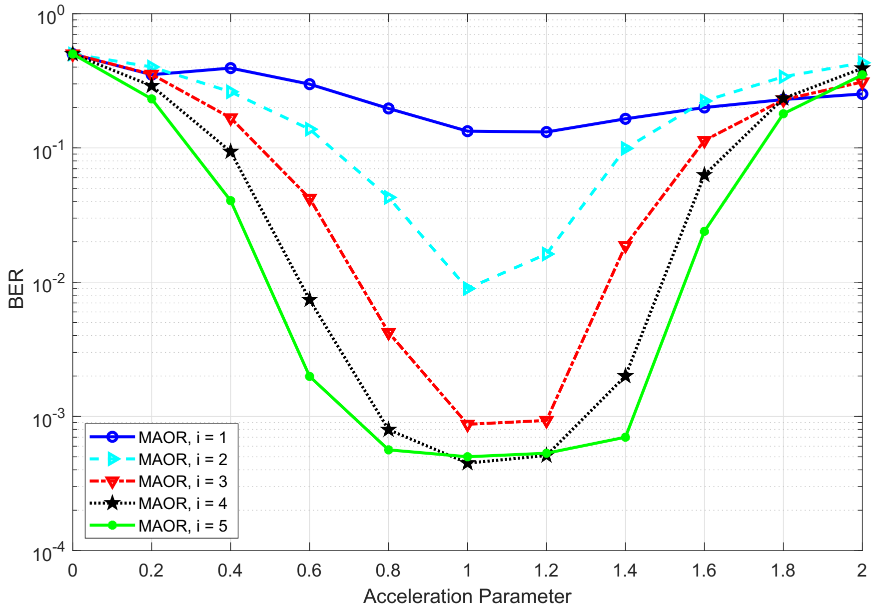
conventional MMSE detection. Considering massive MIMO scenarios corresponding to the UL transmission with T single antenna users and a BS with R antennas. Moreover, in the modulation 64-QAM and 256-QAM scheme is employed and also assume perfect synchronization of Monte Carlo simulations.

The correlated channel model of (2.2) is employed, where the Rayleigh fading is considered $\zeta_{Tx} = \zeta_{Rx} = 0$. As detection for UL massive MIMO, the BS-Correlated channel is considered to $\zeta_{Rx} = 0.4$ and $\zeta_{Rx} = 0.6$. The optimal relaxation parameter is β^{opt}= 1.14 and the optimum acceleration parameter is $\alpha^{opt} = 1.09$, which can be computed using (4.19) and (4.20).

Figs. 4.2 and 4.3 show the impact of the relaxation and acceleration range on the BER performance of the iterative method MAOR, for an SNR level of 15 dB with the Rayleigh fading channel and the Kronecker channel. Although there is some tolerance around the optimum value, the BER performance can be significantly affected by a poor choice of



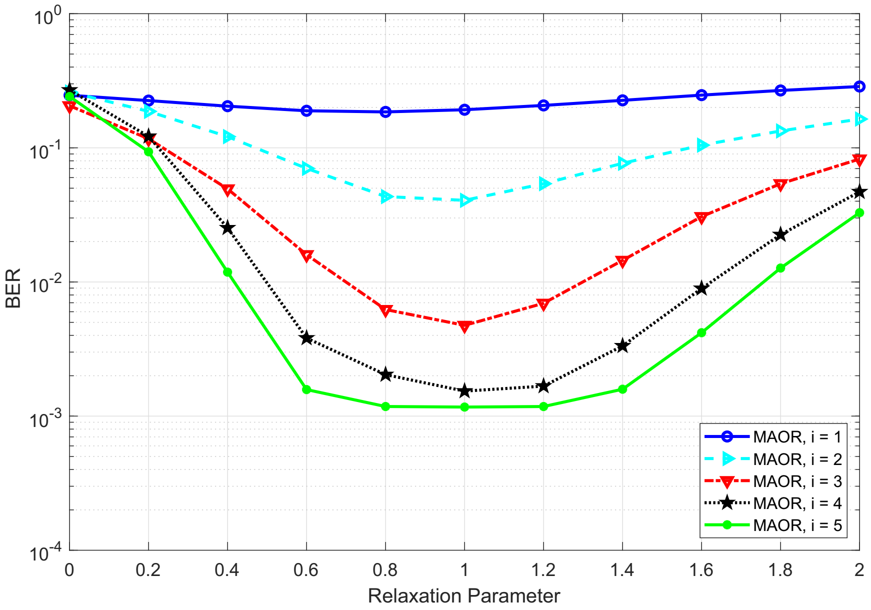
(a) $\alpha = 1.09$



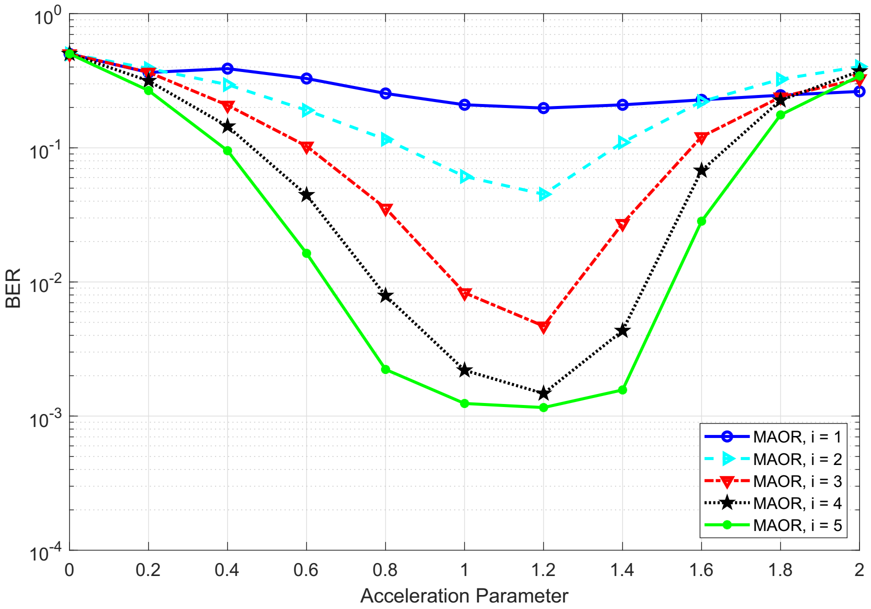
(b) $\beta = 1.14$

Figure 4.2: Efficient relaxation and acceleration range for the MAOR iterative method with configuration antenna R=128, T=16, 64 QAM, the Rayleigh fading channel adopted and with difference number of iterations i and SNR = 15dB: (a) impact of β for $\alpha = 1.09$; (b) impact of α for $\beta = 1.14$.

these parameters. We consider $R = 128$ and $T = 16$, for a 64-QAM constellation, although similar conclusions could be drawn for other constellations and antenna configurations. Hence, the value of relaxation plays an important role in the converge rate for the proposed



(a) $\alpha = 1.09$



(b) $\beta = 1.14$

Figure 4.3: Efficient relaxation and acceleration range for the MAOR iterative method with configuration antenna R=128,T=16, 64 QAM, The Kronecker channel adopt with BS-Correlated $\zeta_{Rx} = 0.4$ and with difference number of iterations i and SNR = 15dB: (a) impact of β for $\alpha = 1.09$; (b) impact of α for $\beta = 1.14$.

method MAOR, while the acceleration value is also effective in the convergence rate but it is much less sensitive compared with relaxation parameters. Moreover, the impact of the relaxation and acceleration parameters is more degradation BER since applying the Kronecker channel, which the Kronecker channel has a deeper degradation than the Rayleigh fading channel. Figs. 4.4 and 4.5 compare the BER performance of the proposed

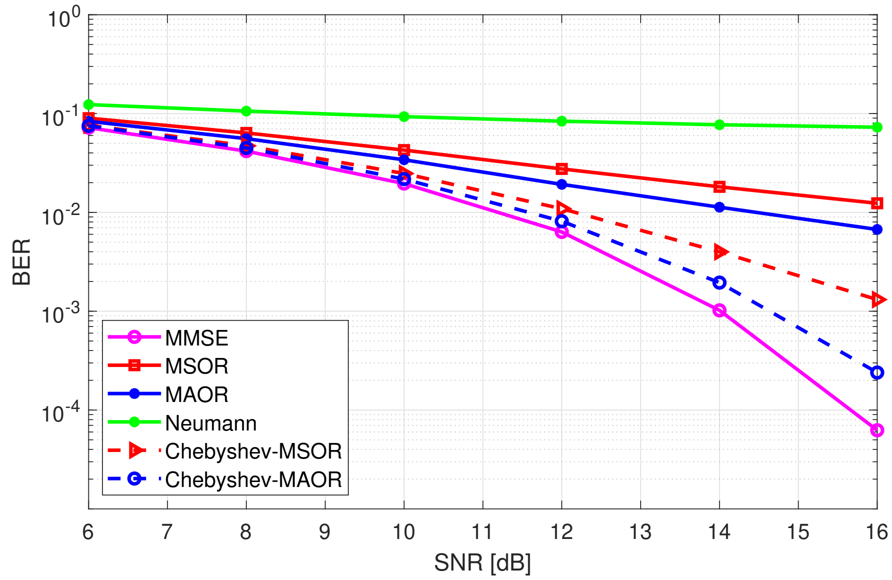


Figure 4.4: BER performance for 64-QAM in a $R \times T = 128 \times 16$ scheme and i=2 iterations.

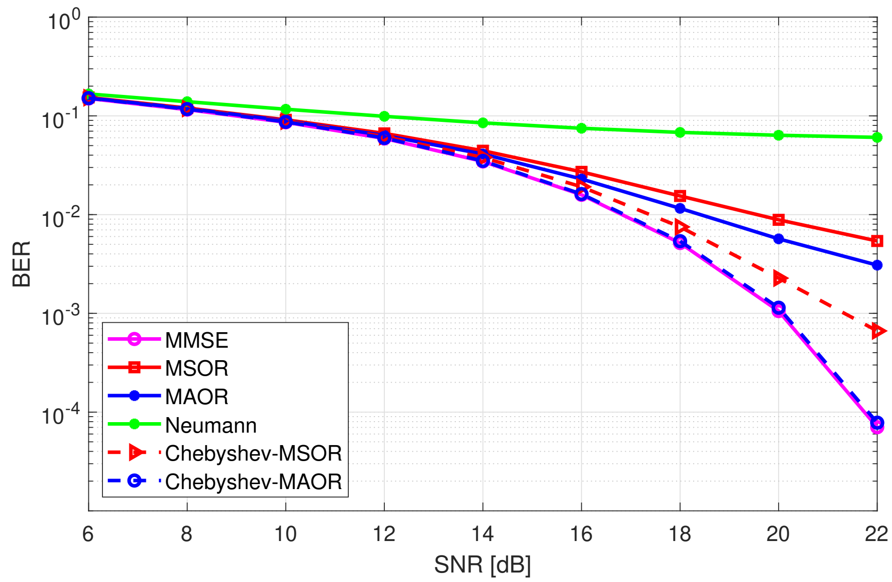


Figure 4.5: BER performance for 256-QAM in a $R \times T = 128 \times 16$ scheme and i=3 iterations.

Chebyshev-MAOR algorithm for 128×16 antenna configuration (i.e., $\frac{R}{T} = \frac{128}{16} = 8$) with different constellation size (64-QAM, 256-QAM).

It is evident from the simulations that the proposed method provides substantially better performance than other methods. In fact, the proposed method achieves the ideal linear MMSE performance with just 2 iterations for 64-QAM, while the Chebyshev-MSOR method and other existing methods require a much larger number of iterations. Even when we increase the constellation size to 256-QAM the performance of the proposed method is still very close to the optimum MMSE performance, with requiring 3 iterations, contrarily to the other techniques, where the performance degradation can increase substantially. Additionally, the Neumann series approximation is the worst method to achieve the performance. Fig. 4.6 presents the BER performance for 64×16 antenna configuration

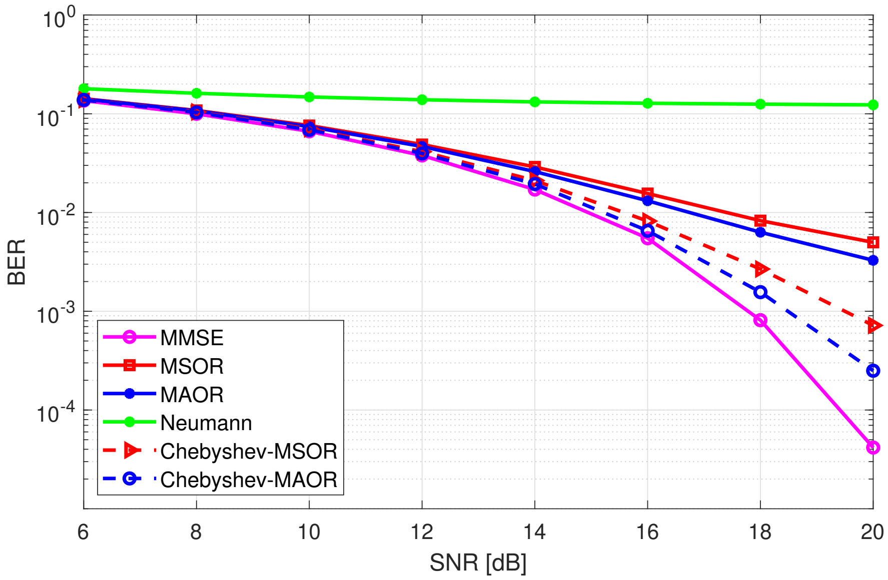


Figure 4.6: BER performance for 64-QAM in a $R \times T = 64 \times 16$ scheme and i=3 iterations.

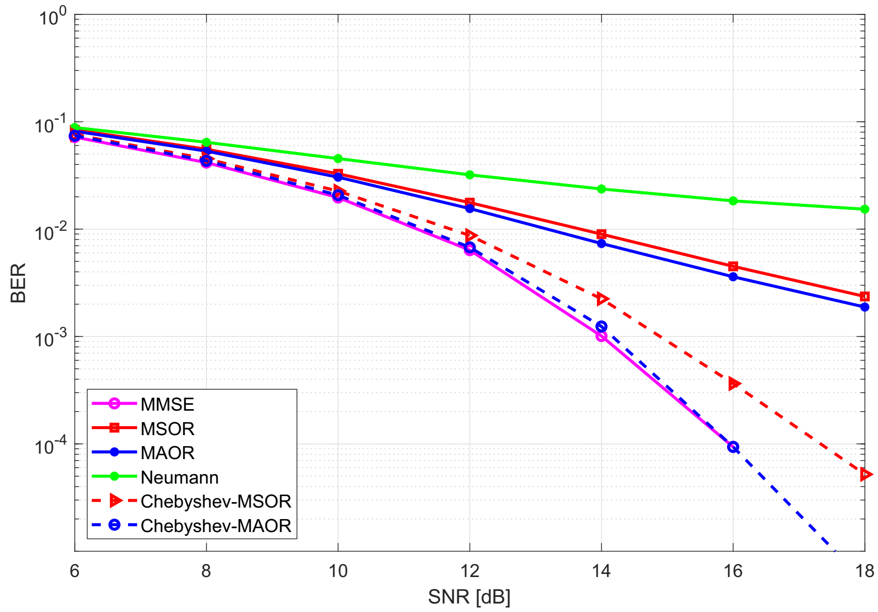


Figure 4.7: BER performance for 64-QAM in a $R \times T = 256 \times 32$ scheme and i=2 iterations.

(i.e., $\frac{R}{T} = \frac{64}{16} = 4$) for a 64-QAM scheme.It is obvious that as the number of iterations increases, the BER efficiency of all methods approaches that of the MMSE performance. However, the Chebyshev-MAOR still significantly outperforms the other methods. As Fig. 4.6, when i=3, the SNR required by the Chebyshev-MAOR method to achieve the BER of 10^{-3} is 18 dB, whereas for the Chebyshev-MSOR method required SNR 19 dB. The proposed method achieved a quasi-MMSE performance in i=3 iterations. Figs. 4.7 and 4.8

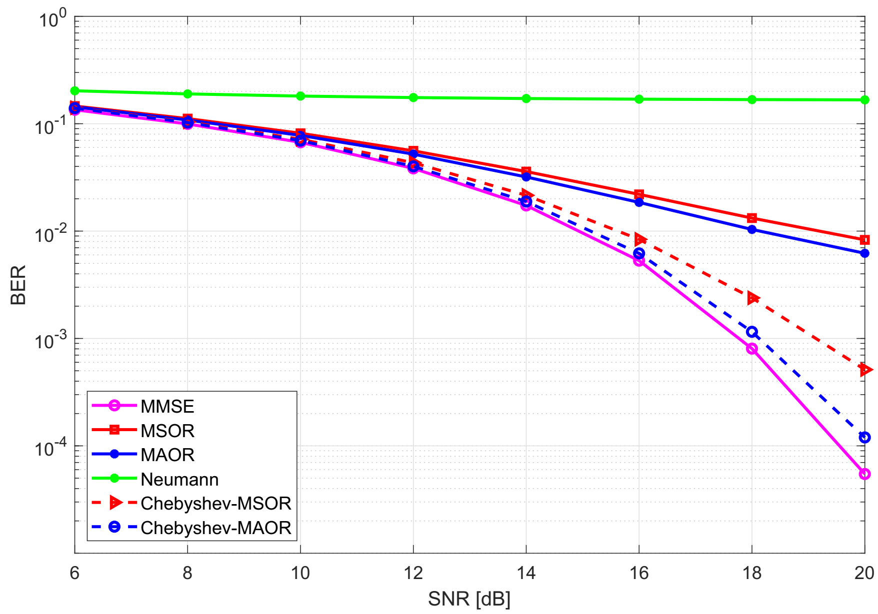
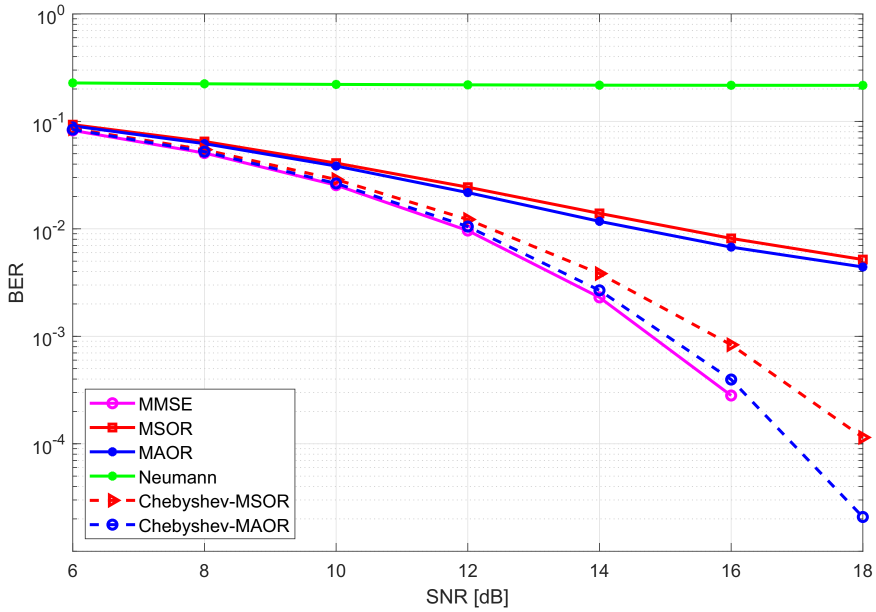


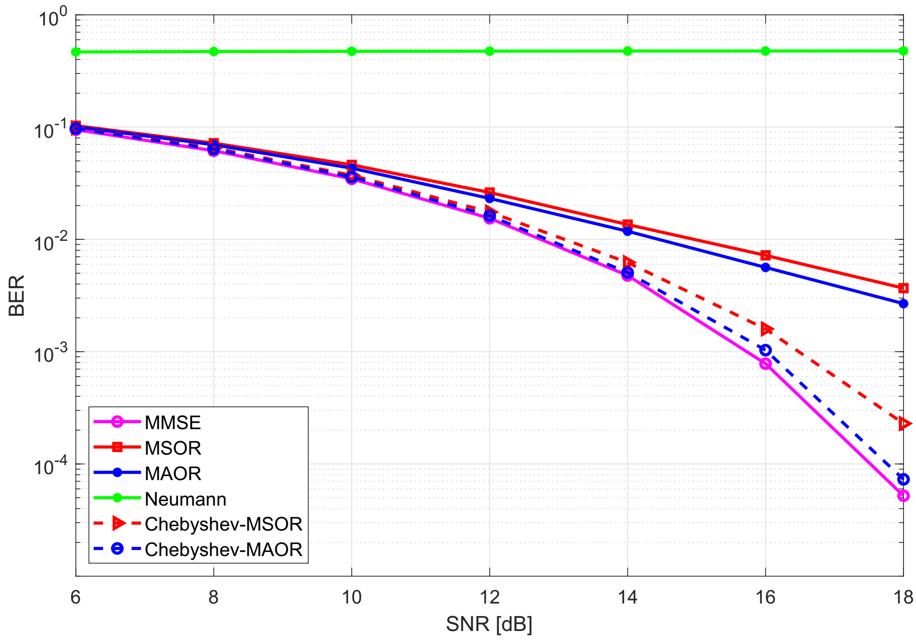
Figure 4.8: BER performance for 64-QAM in a $R \times T = 256 \times 64$ scheme and i=3 iterations.

provide BER comparison of the proposed method and state-of-the-art detection techniques for $\frac{R}{T} = \frac{256}{32} = 8$ and $\frac{256}{64} = 4$ antenna configurations. Once again, Chebyshev-MAOR achieves a quasi-MMSE performance, although we need more iterations when $\frac{N}{K}$ is smaller. In addition, although the performance of the Neumann-based algorithm improves as i increases, it still faces a negligible performance loss.

Next, we consider the impact of antenna correlations. Fig. 4.9 shows the impact of channel correlation on the performance of the proposed Chebyshev-MAOR when $\zeta_{Tx} = 0.4, \zeta_{Rx} = 0$ and $\zeta_{Tx} = 0.6, \zeta_{Rx} = 0$. We consider $\frac{R}{T} = \frac{128}{16} = 8$ for the ratio of a number of BS antennas and the number of users. The modulation scheme is 64-QAM. As expected, the MMSE algorithm degrades when the channel correlation increases, but the proposed method can still converge to the MMSE performance. Contrary to other scenarios, the required number of iterations increases slightly, although the complexity of the proposed method is still much lower than the other methods.



(a)



(b)

Figure 4.9: BER performance for the BS-Correlated Channel with 64-QAM modulation and configuration antenna in a $R \times T = 128 \times 16$. with (a) $\zeta_{Rx} = 0.4$ and $i = 3$, and (b) $\zeta_{Rx} = 0.6$ and $i = 5$

4.6 Conclusions

In this chapter, we proposed an efficient iterative massive MIMO detection algorithm based MAOR method. The speed of convergence for the proposed MAOR-based iterative detection is controlled by two key parameters (known as acceleration and relaxation parameters), and their optimum values were derived. In addition, we applied an eigenvalue-based initial solution instead of the common diagonal matrix approximation, which led to an improved convergence rate. For the acceleration of the MAOR algorithm, we applied the Chebyshev polynomial acceleration method that allows us to reach near-MMSE performance with MAOR. We showed that Chebyshev acceleration has a relatively simple structure with the aid of error analysis. Our performance results showed that the resulting Chebyshev accelerated MAOR provides near MMSE performance with reduced complexity, outperforming state-of-the-art massive MIMO detectors for similar channel conditions.

CHAPTER

5 Fast matrix inversion based on Chebyshev acceleration for Linear detection in massive MIMO systems

Berra, S., Rui Dinis, and Shahriar Shahabuddin. "Fast matrix inversion based on Chebyshev acceleration for linear detection in massive MIMO systems." Electronics Letters 58, no. 11 (2022): 451-453.

5.1 Introduction

In the previous chapter, we presented an MAOR-based iterative method for massive MIMO detection. This chapter presents another alternative low-complexity solution for massive MIMO detection. Linear detection algorithms such as ZF and MMSE schemes were proposed as reduced-complexity alternatives to optimum detection. However, they still involve matrix inversions, whose complexity can be very high, especially in massive MIMO scenarios. Recently, several iterative methods have been proposed to reduce the complexity of the inversion matrix by one magnitude order [84], such as Jacobi (JA), SOR, GS, and RI, but the performance achieved with these methods can still be visibly worse than the performance of the corresponding linear schemes, especially when the number of iterations is not high enough (which corresponds to the scenario of interest, since the complexity increases with the number of iterations).

In this chapter, we propose an acceleration method that is combined with recent iterative methods to improve the convergence. Our method is able to approach the linear MMSE performance with much lower computational complexity. We use a Chebyshev acceleration [95] that rewrites the iterative method with a new vector combination using the spectral radius of the iteration matrix and part of the iterative method. To influence the initialization, the stair matrix [209] has been proposed as the first start of iterative methods. Our performance results show that the Chebyshev acceleration combined with iterative methods are able to approach the linear MMSE performance, outperforming state-of-the-art methods in terms of error rate performance and computational complexity.

5.2 Proposed method

As already pointed out, the convergence of iterative methods like GS, Jacobi (JA), SOR, and RI can be relatively slow. By employing Chebyshev acceleration we can improve the convergence rates. To do this, the matrix to be inverted, $\mathbf{W}$, is decomposed as

$$\mathbf{W} = \mathbf{D} - \mathbf{L} - \mathbf{U}, \tag{5.1}$$

where $\mathbf{D} = diag(\mathbf{W})$, $\mathbf{L}$ and $\mathbf{U}$ are strictly lower and strictly upper triangular matrices, respectively.

The mathematical description of the different iterative methods for the ith iteration is:

- GS-based approach

$$\mathbf{x}^{(i+1)} = (\mathbf{D} - \mathbf{L})^{-1}\mathbf{U}\mathbf{x}^{(i)} + \mathbf{U}\mathbf{y}^{MF} = \mathbf{B}\mathbf{x}^{(i)} + \mathbf{C}. \tag{5.2}$$

- JA approach

$$\mathbf{x}^{(i+1)} = \mathbf{D}^{-1}(\mathbf{D} - \mathbf{W})\mathbf{x}^{(i)} + \mathbf{D}^{-1}\mathbf{y}^{MF} = \mathbf{B}\mathbf{x}^{(i)} + \mathbf{C}. \tag{5.3}$$

- SOR-based approach

$$\mathbf{x}^{(i+1)} = (\mathbf{D} - \omega\mathbf{L})^{-1}((1-\omega)\mathbf{D} + \omega\mathbf{U})\mathbf{x}^{(i)} + (\mathbf{D} - \omega\mathbf{L})^{-1}\mathbf{y}^{MF} = \mathbf{B}\mathbf{x}^{(i)} + \mathbf{C} \tag{5.4}$$

- Richardson-based approach

$$\mathbf{x}^{(i+1)} = (\mathbf{I} - \omega\mathbf{W})\mathbf{x}^{(i)} + \omega\mathbf{y}^{MF} = \mathbf{B}\mathbf{x}^{(i)} + \mathbf{C}. \tag{5.5}$$

In the previous equations, ω of SOR belongs to the interval $[0, 2])$ and ω of Richardson belongs to the interval $[0, 2/\lambda]$, where λ is is the largest eigenvalue [84]. Classically, since no prior knowledge is available about the initial solution $\mathbf{x}^{(0)}$, it is usually selected based on $\mathbf{D}$ [112]. However, it still suffers from the high computational complexity. In this chapter, we propose to utilize the stair matrix-approximate $\mathbf{S}$ as an initial solution [209] of the proposed method to reach a better convergence rate and reduce the computational cost. A matrix $\mathbf{A} = (a_{ij})$, is $m \times m$ matrix. The entries a_{ij} can be $m_i \times m_j$ blocks. A stair matrix is a special case of the tridiagonal matrix [209] and is presented as $\mathbf{S}$=tridiag$(a_{i,i-1}, a_{ii}, a_{i,i+1})$. In order to call a matrix as a stair matrix, it has to satisfy one of the following conditions:

$$a_{i,i-1} = 0, a_{i,i+1} = 0, i = 1, 3, ..., 2\left[\frac{m-1}{2}\right] + 1, \tag{5.6}$$

and

$$a_{i,i-1} = 0, a_{i,i+1} = 0, i = 2, 4, ..., 2\left[\frac{m}{2}\right]. \tag{5.7}$$

For example, a 5×5 stair matrix is of the form

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

In the initial solution $\mathbf{x}^{(0)}$, we apply a stair matrix, which is a special tridiagonal matrix, with partial pivoting [209]. The stair matrix can be expressed as

$$\mathbf{S} = stair(\mathbf{W}(R, R-1); \mathbf{W}(R, R); \mathbf{W}(R, R+1). \tag{5.8}$$

Thus, the initial solution of $\mathbf{x}^{(0)}$ is replaced by a stair matrix, $\mathbf{S}^{(0)}$. Then, the convergence rate of these methods can be further improved by adopting a polynomial acceleration scheme called the Chebyshev acceleration [95]. We generalize the aforementioned iterative detection schemes as (see [95])

$$\mathbf{x}^{(i+1)} = \mathbf{B}\mathbf{x}^{(i)} + \mathbf{c}, \tag{5.9}$$

where $\mathbf{B} = \mathbf{M}^{-1}\mathbf{P}$ and $\mathbf{c} = \mathbf{P}^{-1}\mathbf{y}^{MF}$. Based on (5.9), these iterative methods can generate reasonable approximations of signal vector, $\mathbf{x}$. However, their convergence rate is slow and there are rooms to accelerate the convergence of such iterative detection methods. The Chebyshev acceleration is a good method to accelerate the convergence rate, $(\mathbf{x}^{(i)})_{i=0}^{\infty}$. Based on the signal vector sequence $(\mathbf{x}^{(i)})_{i=0}^{\infty}$, we construct a new linear combination $(\mathbf{y}^{(m)})_{i=1}^{\infty}$, which features faster convergence towards the accurate solution $\mathbf{x}^{(i)}$ of (5.9), known as secondary iteration:

$$\mathbf{y}^{(m)} = \sum_{i=1}^{m} \gamma_{m,i}\mathbf{x}^{(i)}. \tag{5.10}$$

Note that the scalars $\gamma_{m,i}$ must satisfy $\sum_{i=0}^{m} \gamma_{m,i} = 1$. Since if $\mathbf{x}^{(\mathbf{0})} = \mathbf{x}^{(\mathbf{1})} = ... = \mathbf{x}$, we also have $\mathbf{y}^{(m)} = x$. Then, the error between the new combination of polynomial acceleration $\mathbf{y}^{(m)}$ and a signal vector $\mathbf{x}$ can be written as

$$\mathbf{y}^{(m)} - \mathbf{x} = \sum_{i=1}^{m} \gamma_{m,n}(\mathbf{x}^{(i)} - \mathbf{x}) = \sum_{i=1}^{m} \gamma_{m,i}\mathbf{B}^{i}(\mathbf{x}^{(0)} - \mathbf{x}) = p_m(\mathbf{B})(\mathbf{x}_0 - \mathbf{x}), \tag{5.11}$$

where $p_m(\mathbf{B}) = \sum_{i=1}^{m} \gamma_{m,i}\mathbf{B}^i$ is polynomial of degree m with $p_m(1) = 1$, and $p_m(z) = \sum_{i=1}^{m} \gamma_{m,i} z^i$.

The mth Chebyshev polynomial is defined by the three-term recurrence relation

$$T_{m+1}(z) = 2ZT_m(z) - T_{m-1}(z), m \geq 1, \tag{5.12}$$

where initial parameters are $T_0(z) = 1$ and $T_1(z) = z$. A polynomial parameter

$$p_m(z) = \frac{T_m(z/\rho)}{T_m(1/\rho)}, \tag{5.13}$$

can provide minimum error, while ρ is spectral radius of the matrix $\mathbf{B}$ belonging to $[\lambda_{min}, \lambda_{max}]$. If the spectral radius of iterative methods achieves $\rho(\mathbf{B}) < 1$ or $\lim_{t\to\infty} \mathbf{B}_t = 0$, the iterative method will converge for each tth user. According to random matrix theory [201], the spectral radius of $\mathbf{B}$ is given by

$$\rho(\mathbf{B}) = \max |\lambda(\mathbf{B})|, \tag{5.14}$$

where $\lambda(\mathbf{B})$ denotes the eigenvalue of matrix $\mathbf{B}$. The spectral radius of $\mathbf{W}$ [96] and the largest and smallest eigenvalues of $\mathbf{W}$ can be approximated as

$$\rho(\mathbf{W}) = |\lambda_{max}(\mathbf{W})| = R\left(1 + \sqrt{\xi}\right)^2, \tag{5.15}$$

$$\rho(\mathbf{W}) = |\lambda_{min}(\mathbf{W})| = R\left(1 - \sqrt{\xi}\right)^2, \tag{5.16}$$

where $\xi = R/T$ is the ratio between the number of antennas and the number of users. The optimal iteration polynomials $p_m(z)$ can be obtained from the Chebyshev polynomials as

$$p_m(z) = \frac{T_m(1 + 2\frac{z-\epsilon}{\epsilon-\eta})}{T_m(1 + 2\frac{r-\epsilon}{\epsilon-\eta})}, \epsilon = \frac{\lambda_{max} + \lambda_{min}}{\lambda_{max} - \lambda_{min}}, \eta = \frac{\lambda_{max} + \lambda_{min}}{2}. \tag{5.17}$$

The ϵ and η denotes the optimal parameters of polynomials and they are defined by the smallest and the largest eigenvalues of the iteration matrix $\mathbf{B}$, respectively. Here, r is a scalar satisfying the coefficients of polynomials. Let us assume that $\mu_m = 1/T_m(1/\rho)$, so $p_m(\mathbf{B}) = \mu_m/T_m(\mathbf{B})/\rho)$, and $\gamma_{0,0} = 1, \gamma_{1,0} = 0, \gamma_{1,1} = 1$. Then, the properties of μ_m becomes

$$\frac{1}{\mu_{m+1}} = 2\frac{1}{\rho\mu_m} - \frac{1}{\mu_{m-1}}, \tag{5.18}$$

To employ the Chebyshev polynomial method in our proposed technique, we rewrite the equation of $\mathbf{y}^{(m)}$, which also satisfies a three-term recurrence. After some mathematical manipulations, the secondary iteration can be defined as

$$\mathbf{y}^{(m+1)} = 2\frac{\mu_m}{\rho\mu_{m+1}}\mathbf{B}\mathbf{y}^{(m)} - \frac{\mu_{m-1}}{\mu_{m+1}}\mathbf{y}^{(m-1)} + 2\frac{\mu_m}{\rho\mu m + 1}\mathbf{c}, \tag{5.19}$$

where $\mathbf{y}^{(0)} = \mathbf{x}^{(0)}$ and $\mathbf{y}^{(1)} = \mathbf{x}^{(1)}$. Therefore Chebyshev polynomials can be employed to design iterative methods.

5.3 Computational complexity

In this section, we outline the computational complexity of the computation of $\mathbf{W}^{-1}$ in terms of the number of multiplications required. The computational complexities are roughly divided into two parts: 1) The iterative approach; 2) The iterative method combined with the Chebyshev acceleration method. In general, the iterative methods have the complexity of $O(T^2)$, which is lower than the conventional MMSE procedure, which has $O(T^3)$ complexity. Although the current iterative methods can also reduce the complexity to $O(T^2)$, they still require a large number of iterations, which compromises the overall complexity. As a result, the number of iterations necessary to attain a certain level of estimation accuracy reduces. The first step of the stair matrix requires a $T-3$ multiplications. Table 5.1 shows the overall complexities of the proposed methods.

Table 5.1 Computational complexity comparison

Method	Number of multiplications
GS [84]	$i4T^2 + T - 3$
JA [84]	$i(4T^2 - 2T) + T - 3$
RI [84]	$i(4T^2 + 2T) + T - 3$
SOR	$i(16T + 8T^2) + T - 3$
Chebyshev-RI	$i(4T^2 + 4T + 1) + T - 3$
Chebyshev-SOR	$i(16T + 8T^2) + T - 3$
Chebyshev-GS	$i(\frac{5}{2}T^2 + \frac{5}{2}T) + T - 3$
Chebyshev-JA	$i(T^2 + 3K) + T - 3$
MMSE	$T^3 + T^2$

5.4 Performance results

To evaluate the proposed methods based on the Chebyshev acceleration, we provide the BER performance as a function of the signal-to-noise ratio (SNR) simulation, which are compared with conventional iterative methods. We consider a 64-QAM modulation scheme and a massive MIMO configuration of $R = 128$ and $T = 16$. We adopt the Rayleigh fading channel model and a perfect synchronization and channel estimation at the receiver.

Fig. 5.1 shows the BER performance comparison between the iterative methods based on the GS, JA, RI, SOR, where i represents the number of iterations. It is clear that the iterative methods require more iterations to approach the linear MMSE performance, which leads to increased computation complexity. We can also observe the performance of RI method since it has a factor of an overexertion ω [84].

Fig. 5.2 shows the BER performance comparison between the iterative methods combined with the Chebyshev acceleration method. Fig. 5.2 also demonstrates that we get a faster convergence with our acceleration methods compared to conventional iterative methods. For instance, the RI method has almost the MMSE performance with 2 iterations (the same applies to the other methods). It is clear from this result that Chebyshev acceleration method can improve the performance and achieve a fast convergence rate, requiring only 2 iterations to approach the linear MMSE performance.

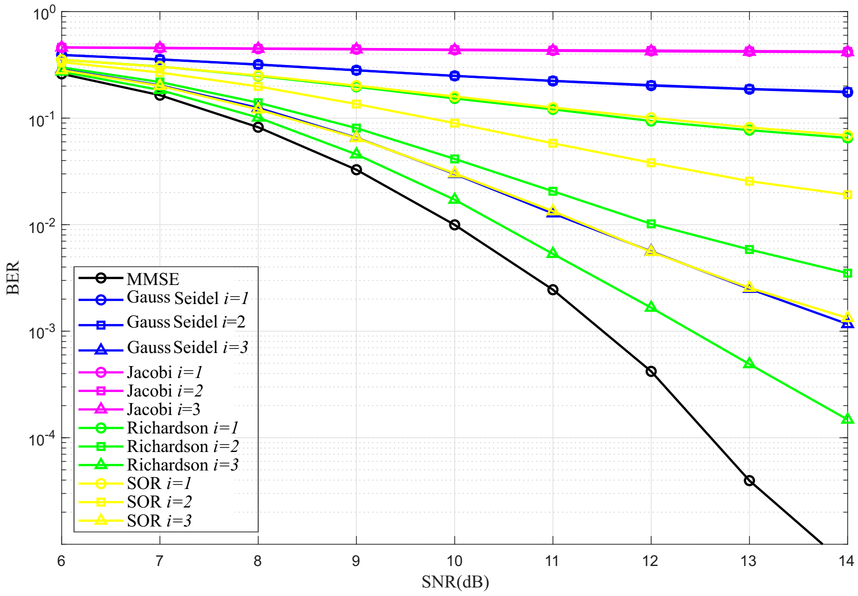


Figure 5.1: BER performance for 64-QAM in a $R \times T = 128 \times 16$ scheme and i iterations.

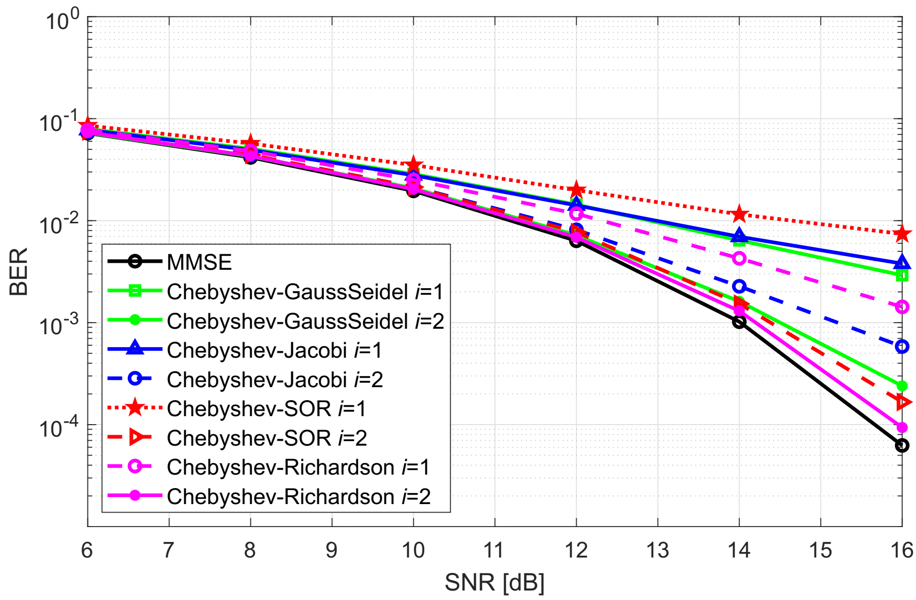


Figure 5.2: BER performance for 64-QAM in a $R \times T = 128 \times 16$ scheme and i iterations.

5.5 Conclusions

In this chapter, we considered low complexity detection for massive MIMO schemes and proposed an acceleration technique based on the Chebyshev method to speed up the existing iterative methods. It was shown that our technique can be employed in different methods, outperforming state-of-the-art methods in terms of error rate performance and computational complexity.

Chapter

6 Deep Unfolding of Chebyshev Accelerated Iterative method for Massive MIMO Detection

Berra, S., Chakraborty, S., Dinis, R., & Shahabuddin, S. (2023). Deep unfolding of the Chebyshev accelerated iterative method for massive MIMO detection. IEEE Access, 11, 52555-52569.

6.1 Introduction

The performance of linear detection methods discussed in previous chapter depends on the parameter values. For example, the successive overrelaxation (SOR) detector [87] can achieve good convergence by utilizing an appropriate matrix splitting pattern and a relaxation factor. In [93], authors proposed an AOR-based detection technique, which requires two parameters: relaxation coefficient and acceleration coefficient. However, selecting the optimum parameters remains challenging since they depend on the spectral radius and eigenvalues of the inversion matrix [94]. While iterative methods can be attractive due to their simplicity, the convergence rate can be further improved through the Chebyshev acceleration [95]. Recent studies have applied the Chebyshev acceleration technique to improve the convergence rates of iterative methods in massive MIMO detection [96, 97]. Recently, there has been a growing interest in using deep unfolding for MIMO detection to balance accuracy and complexity [98, 99]. In [100], the detection network (DetNet) was proposed, which unfolds a projected gradient descent method. Although DetNet has shown significant performance gains for low-order modulations such as BPSK and QPSK, its performance is poor for larger constellations. Moreover, it requires a large number of trainable parameters. In [101], OAMPNet has been proposed, which unfolded the iterations of an OAMP detector. Using data-driven training, OAMPNet has outperformed conventional OAMP in overall detection performance. In addition, the deep unfolding technique was also proposed to improve multiuser interference cancellation schemes. For example, in [102], an iterative sequential detection scheme has been unfolded to enhance its performance. Furthermore, the trainable projected gradient-detector (TPG-detector) has been proposed to optimize the internal parameters using a deep unfolding network [103]. In [104], the modified expectation propagation (MEP) algorithm has been unfolded and named as MEPNet. Two learnable parameters have been optimized through the deep unfolded network. The resulting detector has shown significant improvement in error-rate performance compared to the original MEP. An expectation propagation network (EPNet) has recently been proposed for an orthogonal time frequency space (OTFS) system [105]. The parameters in the resulting detector were optimized using offline training and the detector demonstrated superior performance.

Overall, existing deep unfolded networks in MIMO detection either have many trainable parameters or high complexity. As discussed earlier, iterative methods perform well in massive MIMO detection with lower complexity. However, the performance of some promising iterative methods, such as SOR and AOR, depends on the choice of algorithm parameters. Therefore, optimizing those parameters can improve detection performance. This motivates this work to deep unfold the conventional iterative method to optimize its parameters.

In this chapter, We first propose a Chebyshev acceleration for the AOR and SOR algorithms. The proposed methods include several parameters. Therefore, a deep unfolded network structure has been introduced in this work. In addition, offline training has been employed to optimize these parameters. The significant contributions of this work are summarized as follows:

- We introduce two novel methods for massive MIMO detection, namely AC-AOR, and AC-SOR, which improve the detection performance of conventional AOR and SOR methods, respectively. We demonstrate that they outperform the conventional AOR and SOR methods with only a minor increase in complexity.

- We present the AC-AORNet and AC-SORNet networks, which are the deep unfolding of AC-AOR and AC-SOR algorithms. The network consists of several cascaded layers with the same architecture but different trainable parameters. Additionally, we present a training scheme to optimize the learnable parameters.

- We tested the proposed methods under different channel scenarios. The proposed AC-SORNet and AC-AORNet algorithms exhibit superior performance and robustness compared to their conventional counterparts and other current state-of-the-art methods.

6.2 System model

The uplink massive MIMO system consists of a base station with R antennas serving T user terminals. The receiver signal $\tilde{\mathbf{y}} \in \mathbb{C}^{R\times 1}$ can be modeled as:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{6.1}$$

where $\mathbf{H} \in \mathbb{C}^{R\times T}$ represents the channel between the base station and users, $\mathbf{n} \in \mathbb{C}^{R\times 1}$ is the additive white Gaussian noise (AWGN) that follows a $\mathcal{CR}(0, \tilde{\sigma}^2\mathbf{I}_R)$ distribution, and $\mathbf{x} \in \mathbb{A}^{T\times 1}$ is the transmitted symbol vector from all users, where $\mathbb{A}$ denotes the constellation alphabet. Further, we can transform equation (6.1) into the real domain as follows:

$$\tilde{\mathbf{y}} = \tilde{\mathbf{H}}\tilde{\mathbf{x}} + \tilde{\mathbf{n}}, \tag{6.2}$$

where

$$\tilde{\mathbf{H}} \triangleq \begin{bmatrix} \Re(\mathbf{H}) & -\Im(\mathbf{H}) \\ \Im(\mathbf{H}) & \Re(\mathbf{H}) \end{bmatrix}, \tilde{\mathbf{y}} \triangleq \begin{bmatrix} \Re(\mathbf{y}) \\ \Im(\mathbf{y}) \end{bmatrix}, \tilde{\mathbf{x}} \triangleq \begin{bmatrix} \Re(\mathbf{x}) \\ \Im(\mathbf{x}) \end{bmatrix},$$
$$\tilde{\mathbf{n}} \triangleq \begin{bmatrix} \Re(\mathbf{n}) \\ \Im(\mathbf{n}) \end{bmatrix},$$

and $\tilde{\mathbf{H}} \in \mathbb{R}^{2R\times 2T}$, $\tilde{\mathbf{y}} \in \mathbb{R}^{2R}$, $\tilde{\mathbf{x}} \in \mathbb{A}^{2T}$ and $\tilde{\mathbf{n}} \in \mathbb{R}^{2R}$. Here, $\tilde{\mathbb{A}}$ is the real equivalent constellation set of $\mathbb{A}$. The variance of the real equivalent noise $\tilde{\mathbf{n}}$ is σ^2, where $\sigma^2 = \tilde{\sigma}^2/2$. Linear detectors such as zero-forcing ZF have performed excellently in massive MIMO systems. The detected signal vector for the ZF detector [32] is given by

$$\hat{\tilde{\mathbf{x}}} = (\tilde{\mathbf{H}}^T\tilde{\mathbf{H}})^{-1}\tilde{\mathbf{H}}^T\tilde{\mathbf{y}} = \mathbf{W}^{-1}\tilde{\mathbf{H}}^T\tilde{\mathbf{y}}, \tag{6.3}$$

where $\mathbf{W} = \tilde{\mathbf{H}}^T\tilde{\mathbf{H}}$ for the ZF detector, and $\tilde{\mathbf{y}}^{MF} = \tilde{\mathbf{H}}^T\tilde{\mathbf{H}}$ is the matched filter response. However, inverting the matrix $\mathbf{W}$ has a complexity of $O(T^3)$, which can be prohibitively high even for moderate values of the number of users T.

6.3 Enhancing iterative detectors with Chebyshev acceleration method

In this section, we introduce iterative methods based on SOR and AOR and provide an analysis of their algorithms and parameters. We also explore the characteristics of the Chebyshev acceleration method and present an updated version called the Accelerated Chebyshev (AC) method. Additionally, we demonstrate how the AC method can enhance the convergence rates of AOR and SOR.

6.3.1 The AOR and SOR detector

The matrix to be inverted in the ZF method is Hermitian positive semi-definite [32], and highly likely to be positive define in massive MIMO systems with more antennas at the receiver side than the transmitter side (i.e., more antennas at BS than the users, in the uplink case that we are considering). This unique property has inspired the use of iterative methods such as SOR and AOR to solve (6.3) without requiring the computation of the inversion matrix. The lth iteration of the SOR method can be expressed as [87]

$$\hat{\mathbf{x}}^{(l+1)} = (\mathbf{D} - \omega\mathbf{L})^{-1}((1-\omega)\mathbf{D} + \omega\mathbf{U})\hat{\mathbf{x}}^{(l)} + (\mathbf{D} - \omega\mathbf{L})^{-1}\mathbf{y}^{MF}, \tag{6.4}$$

where $l = 0, 1, \ldots, L-1$, L is the total number of iterations, $\mathbf{x}^{(0)}$ is the initial iteration, ω is the relaxation parameter, where $0 < \omega < 2$. The iteration matrix of SOR is expressed as

$$\mathbf{B}_\omega = (\mathbf{D} - \omega\mathbf{L})^{-1}((1-\omega)\mathbf{D} + \omega\mathbf{U}). \tag{6.5}$$

Similar to SOR, the AOR method can also iteratively solve the (6.3) as follows [93]:

$$\hat{\mathbf{x}}^{(l+1)} = (\mathbf{D} - \beta\mathbf{L})^{-1}((1-\alpha)\mathbf{D} + (\alpha-\beta)\mathbf{L} + \alpha\mathbf{U})\hat{\mathbf{x}}^{(l)} + (\mathbf{D} - \beta\mathbf{L})^{-1}\alpha\mathbf{y}^{MF}, \tag{6.6}$$

where $l = 0, 1, \ldots, L-1$, L is the total number of iterations, $\mathbf{x}^{(0)}$ is the initial iteration, $0 < \alpha < 2$ is the relaxation parameter, and β is the acceleration parameter. The iteration matrix of AOR is expressed as

$$\mathbf{B}_{\alpha,\beta} = (\mathbf{D} - \beta\mathbf{L})^{-1}((1-\alpha)\mathbf{D} + (\alpha-\beta)\mathbf{L} + \alpha\mathbf{U}). \tag{6.7}$$

It is important to understand that the SOR method is a special case of the AOR method when $\alpha = \beta$ [210]. The spectral radius of a matrix $\mathbf{B}$ (such as $\mathbf{B}_\omega$ for the SOR case and $\mathbf{B}_{\alpha,\beta}$ for the AOR case) is defined as $\rho(\mathbf{B}) = \max_{1\le i\le T} |\lambda_i|$, where λ_i denotes the ith eigenvalue of $\mathbf{B}$. To ensure the convergence of (6.4) and (6.6), the spectral radius of the iteration matrix should satisfy $\rho(\mathbf{B}) < 1$ [95]. In [87], it was shown that the spectral radius of the SOR method can be approximated by

$$\rho(\mathbf{B}_\omega) \approx \omega - 1. \tag{6.8}$$

However, achieving a closed-form expression of $\rho(\mathbf{B}_{\alpha,\beta})$ in AOR is difficult and is still an open research topic. Therefore, we can approximate the $\lambda_{max}(\mathbf{B})$ and $\lambda_{min}(\mathbf{B})$ of AOR through a numerical simulation for a given α and β. It was shown in [93] that the optimal parameters of SOR and AOR methods are

$$\begin{aligned}\alpha^{opt} = \omega^{opt} &\approx \frac{2}{1+\sqrt{2(1-\rho(\mathbf{B}_{Jacobi})^2)}},\\ &\approx \frac{2}{1+\sqrt{2(1-|\lambda_{max}(\mathbf{B}_{Jacobi})|^2)}},\end{aligned} \tag{6.9}$$

and

$$\begin{aligned}\beta^{opt} &\approx \frac{\rho(\mathbf{B}_{Jacobi})^2 - \rho(\mathbf{B}_{Jacobi})^4}{\rho(\mathbf{B}_{Jacobi})^2\left(1-\rho(\mathbf{B}_{Jacobi})^2\right)} \\ &\approx \frac{|\lambda_{max}(\mathbf{B}_{Jacobi})|^2 - |\lambda_{min}(\mathbf{B}_{Jacobi})|^4}{|\lambda_{max}(\mathbf{B}_{Jacobi})|^2\left(1-|\lambda_{min}(\mathbf{B}_{Jacobi})|^2\right)},\end{aligned} \tag{6.10}$$

where $\lambda_{max}(\mathbf{B}_{Jacobi})$ and $\lambda_{min}(\mathbf{B}_{Jacobi})$ are the largest and smallest eigenvalues of the Jacobi iteration matrix $\mathbf{B}_{Jacobi}$, where [85]

$$\mathbf{B}_{Jacobi} = \mathbf{D}^{-1}\mathbf{W} - \mathbf{I}. \tag{6.11}$$

Since the number of antennas R and users T is large, and $R >> T$, the matrix $\mathbf{W}$ is diagonally dominant, and that allows us to approximate the inverse of the diagonal matrix $\mathbf{D}$ as

$$\mathbf{D}^{-1} \approx \frac{1}{R}\mathbf{I}. \tag{6.12}$$

According to [93], the largest and smallest eigenvalues of $\mathbf{W}$ can be approximated by

$$\lambda_{max}(\mathbf{W}) \approx R\left(1+\sqrt{\xi}\right)^2, \tag{6.13}$$

$$\lambda_{min}(\mathbf{W}) \approx R\left(1-\sqrt{\xi}\right)^2, \tag{6.14}$$

where $\xi = R/T$ is the ratio between the number of antennas and users.Based on the analysis above, the optimal value of α^{opt} , β^{opt}, and ω^{opt} can be expressed as [88]

$$\alpha^{opt} = \omega^{opt} = \frac{2}{1+\sqrt{1-\overline{\eta}^2}}, \tag{6.15}$$

$$\beta^{opt} = \frac{\overline{\eta}^2 - \underline{\eta}^4}{\overline{\eta}^2\left(1-\underline{\eta}^2\right)}, \tag{6.16}$$

where η can be presented as

$$\begin{aligned}\overline{\eta} &= \lambda_{max}(\mathbf{B}_{Jacobi}) = \left(1+\sqrt{\xi}\right)^2 - 1, \\ \underline{\eta} &= \lambda_{min}(\mathbf{B}_{Jacobi}) = \left(1-\sqrt{\xi}\right)^2 - 1.\end{aligned} \tag{6.17}$$

For SOR to converge, the relaxation parameter ω needs to be between 0 and 2. In massive MIMO, the optimal values ω^{opt} and β^{opt} in (6.15) are determined by the system configuration parameters R and T. When the massive MIMO system configuration is fixed, ω^{opt} and β^{opt} can be approximated as a constant value. Another parameter of the AOR method is the acceleration parameter, which is also related to the configuration of massive MIMO and should satisfy $0 < \beta < \alpha$ [93]. It has been demonstrated that the optimal AOR approach yields faster convergence rates than the optimal SOR approach [203].

Despite all the efforts mentioned above, achieving the desired convergence rate and performance by optimizing the parameters of AOR and SOR is challenging.

6.3.2 The Chebyshev acceleration method

In this subsection, we analyze the Accelerated Chebyshev method [95, 211]. This method involves forming a new vector sequence by linearly combining the iterates obtained from the basic method, and it includes two parameters: under-relaxation and under-extrapolation. Let us examine the following splitting equation:

$$\hat{\mathbf{x}}^{(l+1)} = \mathbf{B}\hat{\mathbf{x}}^{(l)} + \mathbf{c}, \tag{6.18}$$

where $\mathbf{B}$ is the iteration matrix (for the AOR case $\mathbf{B} = \mathbf{B}_{\alpha,\beta}$, and for the SOR case $\mathbf{B} = \mathbf{B}_{\omega}$) and $\mathbf{c}$ is the other parameter ($\mathbf{c}_{\alpha,\beta}$ =$(\mathbf{D} - \beta\mathbf{L})^{-1}\alpha\mathbf{y}^{MF}$ for AOR and $\mathbf{c}_{\omega}$=$(\mathbf{D} - \omega\mathbf{L})^{-1}\mathbf{y}^{MF}$ for SOR, respectively). To generate the polynomial based on the coefficients, we construct a new linear combination $(\bar{\mathbf{x}}^{(l)})_{i=0}^{\infty}$, which features faster convergence towards the accurate solution $\hat{\mathbf{x}}^{(l)}$ of (6.18) and is known as secondary iteration. The secondary iteration can be expressed as

$$\bar{\mathbf{x}}^{(l)} = \sum_{i=0}^{l} \nu_{l,i}\hat{\mathbf{x}}^{(i)}, \tag{6.19}$$

Note that the scalars $\nu_{l,i}$ must satisfy $\sum_{i=0}^{l} \nu_{l,i} = 1$. If $\hat{\mathbf{x}}^{(\mathbf{0})} = \hat{\mathbf{x}}^{(\mathbf{1})} = ... = \mathbf{x}$, then $\bar{\mathbf{x}}^{(l)} = \mathbf{x}$. So, the error vector $\mathbf{e}^{(l)}$ at the lth iteration for the approximation of $\mathbf{x}$ can be expressed as

$$\begin{aligned}
\mathbf{e}^{(l)} = \bar{\mathbf{x}}^{(l)} - \mathbf{x} &= \sum_{i=0}^{l} \nu_{l,i}(\hat{\mathbf{x}}^{(i)} - \mathbf{x}) \\
&= \sum_{i=0}^{l} \nu_{l,i}\mathbf{B}^{i}(\mathbf{x}^{(0)} - \mathbf{x}) \\
&= P_l(\mathbf{B})(\mathbf{x}^{(0)} - \mathbf{x}) \\
&= P_l(\mathbf{B})\mathbf{e}^{(0)},
\end{aligned} \tag{6.20}$$

where $P_l(\mathbf{B}) = \sum_{i=0}^{l} \nu_{l,i}\mathbf{B}^{i}$ is polynomial of degree l with $P_l(1) = 1$, and $P_l(x) = \sum_{i=0}^{l} \nu_{l,i}x^{i}$. The norm error of $\mathbf{e}^{(l)}$ can be presented as

$$||\mathbf{e}^{(l)}|| = ||P_l(\mathbf{B})\mathbf{e}^{(0)}|| \leq ||P_l(\mathbf{B})||.||\mathbf{e}^{(0)}||, \quad l \geq 0 \tag{6.21}$$

and we have (see equation (5.16), of [207])

$$||P_l(\mathbf{B})|| = \rho(P_l(\mathbf{B})) = \max_{1\leq i\leq T} |P_l(\lambda_i(\mathbf{B}))|, \quad l \geq 0 \tag{6.22}$$

where $\lambda_i(\mathbf{B})$ denotes the eigenvalues of $\mathbf{B}$. Computing the eigenvalues of a large linear system is often impractical. However, we know that the eigenvalue of $\lambda_i(\mathbf{B})$ will lie in some interval $-1 < \lambda_{min}(\mathbf{B}) \leq \lambda_i(\mathbf{B}) \leq \lambda_{max}(\mathbf{B}) < 1$ [207]. Then, we define the algebraic polynomial $P_l(x)$ using the associated polynomial $P_l(\mathbf{B})$, as follows (see equation (5.17), of [207]):

$$||P_l(\mathbf{B})|| = \rho(|P_l(\lambda_i(\mathbf{B}))|) = \max_{\lambda_{min}(\mathbf{B})\leq x\leq\lambda_{max}(\mathbf{B})} |P_l(x)|, \tag{6.23}$$

where $P_l(x)$ is a real polynomial. A particular polynomial based on the Chebyshev polynomial can be used as a solution for $P_l(x)$ in (6.23). It is given by [212]

$$P_l(x) = \frac{T_l(x/\rho(\mathbf{B}))}{T_l(1/\rho(\mathbf{B}))}, \tag{6.24}$$

where $\rho(\mathbf{B})$ is the spectral radius of the iteration matrix $\mathbf{B}$, and T_l is the Chebyshev polynomial. The three-term recurrence relation defines the lth expression of Chebyshev polynomials as [95]

$$T_{l+1}(x) = 2xT_l(x) - T_{l-1}(x), \quad l \geq 0 \tag{6.25}$$

where initial parameters indicates to $T_0(x) = 1$ and $T_1(x) = x$. Using (6.24), we can modify (6.25) as

$$\begin{aligned} P_{l+1}(x)T_{l+1}(1/\rho(\mathbf{B})) &= \frac{2x}{\rho(\mathbf{B})}T_l(x/\rho(\mathbf{B})) - T_{l-1}(x/\rho(\mathbf{B})) \\ &= \frac{2x}{\rho(\mathbf{B})}P_l(x)T_l(1/\rho(\mathbf{B})) \\ &- P_{l-1}(x)(\frac{2}{\rho(\mathbf{B})}T_l(1/\rho(\mathbf{B})) \\ &- T_{l+1}(1/\rho(\mathbf{B}))). \end{aligned} \tag{6.26}$$

Equation (6.26) can be further rewritten as

$$T_{l+1}(1/\rho(\mathbf{B}))(P_{l+1}(x) - P_{l-1}(x)) = \frac{2T_l(1/\rho(\mathbf{B}))}{\rho(\mathbf{B})}(xP_l(x) - P_{l-1}(x)). \tag{6.27}$$

By replacing x with $\mathbf{B}$ and multiplying by $\mathbf{e}^{(0)}$ on both sides of (6.27), and using (6.20) , we find

$$T_{l+1}(1/\rho(\mathbf{B}))(\bar{\mathbf{x}}^{(l+1)} - \bar{\mathbf{x}}^{(l-1)}) = \frac{2T_l(1/\rho(\mathbf{B}))}{\rho(\mathbf{B})} \times (\mathbf{B}(\bar{\mathbf{x}}^{(l)} - \mathbf{x}) - \bar{\mathbf{x}}^{(l-1)} + \mathbf{x}), \tag{6.28}$$

where $\bar{\mathbf{x}}^{(l)}$ is polynomial. Hence, (6.28) can be rewritten as

$$\bar{\mathbf{x}}^{(l+1)} = p_{(l+1)}(\mathbf{B}(\bar{\mathbf{x}}^{(l)} - \mathbf{x}) - \bar{\mathbf{x}}^{(l-1)} + \mathbf{x}) + \bar{\mathbf{x}}^{(l-1)}, \tag{6.29}$$

with the under-relaxation parameter

$$p_{l+1} = \frac{2T_l(1/\rho(\mathbf{B}))}{\rho(\mathbf{B})T_{l+1}(1/\rho(\mathbf{B}))}. \tag{6.30}$$

By the definition of the Chebyshev polynomial, we can further modify p_{l+1} as follows [213]

$$p_{l+1} = \frac{2T_l}{\rho(\mathbf{B})(\frac{2}{\rho}T_l - T_{l-1})} = \frac{2T_l}{\rho(\mathbf{B})(\frac{2}{\rho(\mathbf{B})}T_l - \frac{\rho(\mathbf{B})p_lT_l}{2})} = \frac{4}{4 - \rho(\mathbf{B})^2 p_l}, \tag{6.31}$$

The initial solutions for (6.31) are $p_1 = 1, p_2 = \frac{2}{2-\rho(\mathbf{B})^2}$. Using (6.18), the sum $-\mathbf{B}\mathbf{x} + \mathbf{x}$ is equal to $\mathbf{B}\mathbf{b}$. Hence, (6.29) can be rewritten as

$$\bar{\mathbf{x}}^{(l+1)} = p_{l+1}\left(\mathbf{B}(\bar{\mathbf{x}}^{(l)} + \mathbf{c}) - \bar{\mathbf{x}}^{(l-1)}\right) + \bar{\mathbf{x}}^{(l-1)}. \tag{6.32}$$

To improve the convergence rate and performance of (6.32), we include another parameter called under-extrapolation γ, as defined in [95, 211]. Then (6.32) can be rewritten as

$$\begin{aligned}\bar{\mathbf{x}}^{(l+1)} &= p_{l+1}\left(\gamma(\mathbf{B}\bar{\mathbf{x}}^{(l)} + \mathbf{c}) + (1-\gamma)\bar{\mathbf{x}}^{(l)}\right) + (1-p_{l+1})\bar{\mathbf{x}}^{(l-1)} \\ &= p_{l+1}\left(\gamma\hat{\mathbf{x}}^{(l+1)} + (1-\gamma)\bar{\mathbf{x}}^{(l)}\right) + (1-p_{l+1})\bar{\mathbf{x}}^{(l-1)},\end{aligned} \tag{6.33}$$

where $\hat{\mathbf{x}}^{(l+1)} = \mathbf{B}\bar{\mathbf{x}}^{(l)} + \mathbf{c}$ is based on the some form of (6.18). Rearranging (6.33) results , which can be as

$$\bar{\mathbf{x}}^{(l+1)} = p_{l+1}\left(\gamma(\hat{\mathbf{x}}^{(l+1)} - \bar{\mathbf{x}}^{(l)}) + (\bar{\mathbf{x}}^{(l)} - \bar{\mathbf{x}}^{(l-1)})\right) + \bar{\mathbf{x}}^{(l-1)}. \tag{6.34}$$

The final equation (6.34) is called the Accelerated Chebyshev method (AC), which contains two additional parameters: p_{l+1} and γ, along with the AOR/SOR parameters. The parameter γ has been estimated in [211] as

$$\gamma = \frac{2}{2 - \lambda_{max}(\mathbf{B}) - \lambda_{min}(\mathbf{B})}. \tag{6.35}$$

On the other hand, reference [212] suggested that choosing γ value between 0.6 to 0.9 can result in an improved convergence rate. The Accelerated Chebyshev method is highly advantageous due to its ease of use. Unlike Krylov subspace iterative methods [214], such as generalized minimum residual and conjugate gradient, the Chebyshev acceleration method has a concise recurrence form and does not rely on inner products, making it suitable for parallel computing. Furthermore, according to [95], the use of the Accelerated Chebyshev method has notably accelerated the convergence of an iterative method (6.18). In addition, fine-tuning the parameters γ and p_{l+1} can significantly improve its overall performance.

Based on the previous discussion, we present two algorithms to improve the convergence rate of AOR and SOR. In Algorithm 2, we present the AC-AOR algorithm. The inputs of the algorithm are $\tilde{\mathbf{y}}$, $\tilde{\mathbf{H}}$, $\mathbf{D}$, $\mathbf{L}$, $\mathbf{U}$, β, α, γ, and ρ (here we abbreviated $\rho(\mathbf{B})$ as ρ). The parameters α and β can be determined using (6.15) and (6.16), respectively. The parameter γ can be determined using (6.35) or can be approximated from the range $[0.6, 0.9]$. The spectral radius ρ of the iteration matrix can be estimated using numerical simulation. For $R >> T$ with a given α and β, ρ can be approximated as a constant for all channel matrices. Hence, to avoid the complexity burden for exact ρ computation, in this work, we have used an approximated value of ρ. The initialization is presented between lines 4-7. On line 4

Algorithm 2 **Accelerated Chebyshev-based AOR**

1: **Input:**$\tilde{\mathbf{y}}, \tilde{\mathbf{H}}, \mathbf{D}, \mathbf{L}, \mathbf{U}, \beta, \alpha, \gamma, \rho$
2: **Output:** Estimated signal $\bar{\mathbf{x}}^{(L)}$
3: **Initialization phase:**
4: $\tilde{\mathbf{y}}^{MF} = \tilde{\mathbf{H}}^T \tilde{\mathbf{y}}$
5: $\hat{\mathbf{x}}^{(0)} = \bar{\mathbf{x}}^{(0)} = \frac{2}{\lambda_{max}(\mathbf{W})+\lambda_{min}(\mathbf{W})} \tilde{\mathbf{y}}^{MF}$.
6: $\bar{\mathbf{x}}^{(1)} = \gamma(\mathbf{B}_{\alpha,\beta}\bar{\mathbf{x}}^{(0)} + \mathbf{c}_{\alpha,\beta}) + (1-\gamma)\bar{\mathbf{x}}^{(0)}$
7: $p_2 = \frac{2}{2-\rho^2}$
8: **Iteration phase:**
9: **for** $l = 1 : L-1$ **do**
10: $\quad \hat{\mathbf{x}}^{(l+1)} = (\mathbf{D} - \beta\mathbf{L})^{-1}((1-\alpha)\mathbf{D} + (\alpha-\beta)\mathbf{L} + \alpha\mathbf{U})\hat{\mathbf{x}}^{(l)} + (\mathbf{D} - \beta\mathbf{L})^{-1}\alpha\tilde{\mathbf{y}}^{MF}$.
11: $\quad \bar{\mathbf{x}}^{(l+1)} = p_{l+1}\Big(\gamma(\hat{\mathbf{x}}^{(l+1)} - \bar{\mathbf{x}}^{(l)}) + (\bar{\mathbf{x}}^{(l)} - \bar{\mathbf{x}}^{(l-1)})\Big) + \bar{\mathbf{x}}^{(l-1)}$.
12: $\quad \bar{\mathbf{x}}^{(l-1)} = \bar{\mathbf{x}}^{(l)}, \bar{\mathbf{x}}^{(l)} = \bar{\mathbf{x}}^{(l+1)}$.
13: $\quad p_{l+2} = \frac{4}{4-\rho^2 p_{l+1}}$.
14: **end for**
15: **Return** $\bar{\mathbf{x}}^{(L)}$.

Algorithm 3 **Accelerated Chebyshev-based SOR**

1: **Input:**$\tilde{\mathbf{y}}, \tilde{\mathbf{H}}, \mathbf{D}, \mathbf{L}, \mathbf{U}, \omega, \gamma, \rho$
2: **Output:** Estimated signal $\bar{\mathbf{x}}^{(L)}$
3: **Initialization phase:**
4: $\tilde{\mathbf{y}}^{MF} = \tilde{\mathbf{H}}^T \tilde{\mathbf{y}}$
5: $\hat{\mathbf{x}}^{(0)} = \bar{\mathbf{x}}^{(0)} = \frac{2}{\lambda_{max}(\mathbf{W})+\lambda_{min}(\mathbf{W})} \tilde{\mathbf{y}}^{MF}$.
6: $\bar{\mathbf{x}}^{(1)} = \gamma(\mathbf{B}_{\omega}\bar{\mathbf{x}}^{(0)} + \mathbf{c}_{\omega}) + (1-\gamma)\bar{\mathbf{x}}^{(0)}$
7: $p_2 = \frac{2}{2-\rho^2}$
8: **Iteration phase:**
9: **for** $l = 1 : L-1$ **do**
10: $\quad \hat{\mathbf{x}}^{(l+1)} = (\mathbf{D} - \omega\mathbf{L})^{-1}\left((1-\omega)\mathbf{D} + \omega\mathbf{U}\right)\hat{\mathbf{x}}^{(l)} + (\mathbf{D} - \omega\mathbf{L})^{-1}\tilde{\mathbf{y}}^{MF}$
11: $\quad \bar{\mathbf{x}}^{(l+1)} = p_{l+1}\Big(\gamma(\hat{\mathbf{x}}^{(l+1)} - \bar{\mathbf{x}}^{(l)}) + (\bar{\mathbf{x}}^{(l)} - \bar{\mathbf{x}}^{(l-1)})\Big) + \bar{\mathbf{x}}^{(l-1)}$.
12: $\quad \bar{\mathbf{x}}^{(l-1)} = \bar{\mathbf{x}}^{(l)}, \bar{\mathbf{x}}^{(l)} = \bar{\mathbf{x}}^{(l+1)}$
13: $\quad p_{l+2} = \frac{4}{4-\rho^2 p_{l+1}}$.
14: **end for**
15: **Return** $\bar{\mathbf{x}}^{(L)}$.

we calculate the matched filter output $\tilde{\mathbf{y}}^{MF} = \tilde{\mathbf{H}}^T\tilde{\mathbf{y}}$. The initial value presented in line 5 is based on the channel hardening property of massive MIMO systems. Since $R \gg T$, the channel matrix $\mathbf{H}$ is asymptotically orthogonal, and the matrix $\mathbf{W}$ is diagonally dominant, which leads us to $\frac{2}{\lambda_{max}+\lambda_{min}}\tilde{\mathbf{y}}^{MF}$ (more details can be found in [96]). On lines 6 and 7 we have the initial value of the Accelerated Chebychev method, and the initial value of under-relaxation. Afterward, we present the iterative part of AC-AOR, which includes two main parts: the AOR calculation at line 10, which is adopted from (6.6), and the Accelerated Chebyshev method in line 11 based on equation (6.34). Finally, the AC-SOR algorithm is presented in Algorithm 3, which is very similar to AC-AOR except that we have the parameter ω as input and the SOR equation at line 10, which is adopted from (6.4).

6.4 Deep unfolded network

In this section, we first introduce SORNet, AORNet, AC-AORNet, and AC-SORNet based on the deep unfolding of SOR, AOR, AC-SOR, and AC-AOR methods, respectively. Each deep unfolding network consists of several cascaded layers with the same architecture but different trainable parameters. We present a training scheme for optimizing the learnable parameters. Moreover, a complexity analysis of the proposed methods has been presented, which applies to the testing phase.

6.4.1 Structure of the deep unfolded network

The structures of the proposed AC-SORNet and AC-AORNet are illustrated in Figures 6.1 and 6.2, respectively. AC-SORNet and AC-AORNet are obtained by unfolding the iterations of AC-SOR and AC-AOR. As a result, the learnable parameters at the lth layer are ω_l, γ_l, and p_l for AC-SORNet, and for AC-AORNet the parameters are β_l, α_l, p_l, and γ_l .

The input of the AC-SORNet and AC-AORNet algorithms consists of the matched filter output $\tilde{\mathbf{y}}^{MF}$, and three matrices $\mathbf{D}$, $\mathbf{L}$, and $\mathbf{U}$, which are obtained by decomposing $\mathbf{W}$. The final output estimates the transmitted symbol vector, denoted by $\hat{\mathbf{x}}$.

Similarly, the parameters of conventional SOR and AOR can be optimized through the DUN scheme, resulting in detectors known as SORNet and AORNet. The networks presented in Figures 6.1 and 6.2 consist of L layers, with three variables learned at each layer for AC-SORNet and four variables learned at each layer for each layer AC-AORNet. Therefore, the lth layer of the AORNet can be presented as

$$\hat{\mathbf{x}}^{(l+1)} = (\mathbf{D} - \beta_l \mathbf{L})^{-1}((1 - \alpha_l)\mathbf{D} + (\alpha_l - \beta_l)\mathbf{L} + \alpha_l \mathbf{U})\hat{\mathbf{x}}^{(l)} + (\mathbf{D} - \beta_l \mathbf{L})^{-1}\alpha_l \tilde{\mathbf{y}}^{MF}. \quad (6.36)$$

On the other hand, the lth layer of the SORNet is represented by

$$\hat{\mathbf{x}}^{(l+1)} = (\mathbf{D} - \omega_l \mathbf{L})^{-1}((1 - \omega_l)\mathbf{D} + \omega_l \mathbf{U})\hat{\mathbf{x}}^{(l)} + (\mathbf{D} - \omega_l \mathbf{L})^{-1}\tilde{\mathbf{y}}^{MF}. \quad (6.37)$$

In the case of AC-AORNet and AC-SORNet, the Accelerated Chebyshev method is combined with iterative methods such as AORNet or SORNet. The general expression for the lth layer of the Accelerated Chebyshev method can be expressed as

$$\bar{\mathbf{x}}^{(l+1)} = p_{l+1}\Big(\gamma_l(\hat{\mathbf{x}}^{(l+1)} - \bar{\mathbf{x}}^{(l)}) + (\bar{\mathbf{x}}^{(l)} - \bar{\mathbf{x}}^{(l-1)})\Big) + \bar{\mathbf{x}}^{(l-1)}, \quad (6.38)$$

where $\hat{\mathbf{x}}^{(l+1)}$ is obtained from (6.36) for AC-AORNet and in case of AC-SORNet, $\hat{\mathbf{x}}^{(l+1)}$ is obtained from (6.37). Also, after computing (6.38), we need to perform the following operations for the next layer: $\bar{\mathbf{x}}^{(l-1)} = \bar{\mathbf{x}}^{(l)}$, $\hat{\mathbf{x}}^{(l)} = \bar{\mathbf{x}}^{(l+1)}$. From (6.38), we notice that the trainable parameters for the AC-AORNet are $(\{\alpha_l\}_{l=1}^{L}, \{\beta_l\}_{l=1}^{L}, \{\gamma_l\}_{l=1}^{L}, \{p_{l+1}\}_{l=1}^{L})$, and for AC-SORNet they are $(\{\omega_l\}_{l=1}^{L}, \{\gamma_l\}_{l=1}^{L}, \{p_{l+1}\}_{l=1}^{L})$.

6.4.2 Training of deep unfolded network

The parameters of the AC-AORNet and AC-SORNet methods are trained using the DUN method, as illustrated in Figures 6.1 and 6.2. As the AC-AORNet and AC-SORNet have a small number of trainable parameters, the training requires fewer iterations to optimize those parameters. Additionally, the number of trainable parameters in these detectors is independent of the number of antennas R and users T and is solely determined by the number of layers L.

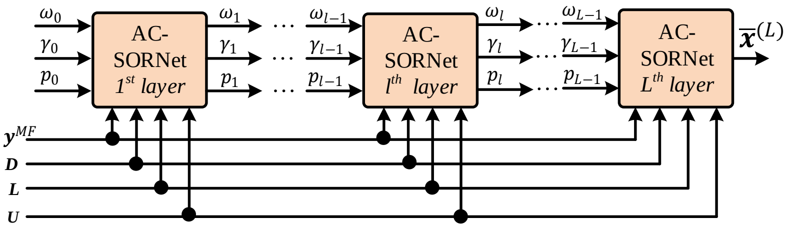


Figure 6.1: AC-SORNet architecture. Each of the l cascading layers has the same structure.

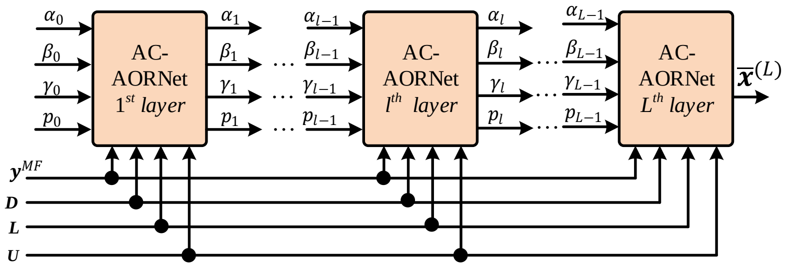


Figure 6.2: AC-AORNet architecture. Each of the l cascading layers has the same structure.

In figure 6.3, we present a generic functional architecture for training the proposed DUNs (i.e., SORNet, AORNet, AC-SORNet, and AC-AORNet). The transmitted symbols are selected from the constellation alphabet $\tilde{\mathbb{A}}$ (without loss of generality, we assumed QAM constellations). The channel matrix $\tilde{\mathbf{H}}$ is generated using either of IID Rayleigh channel model or correlated channel model, or realistic channel model. The noise vector $\tilde{\mathbf{n}}$ has zero mean and variance $\tilde{\sigma}^2$. Then, we train the AC-SORNet or AC-AORNet until all learnable variables converge. The limiting performance of the iterative methods, such as SOR, AOR, AC-AOR, AC-SOR, etc., is the ZF performance. Hence, the response

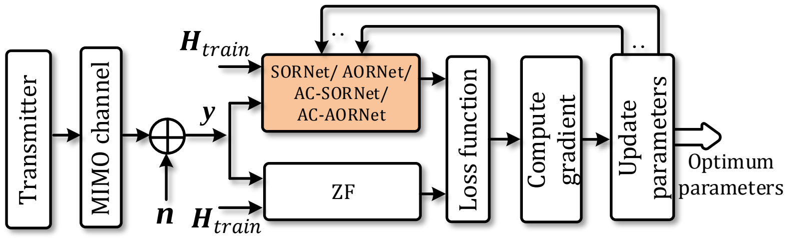


Figure 6.3: Block diagram of training scheme of AC-SORNet, AC-AORNet, SORNet, and AORNet.

of the ZF detector has been considered as the reference for the training. In other words, the objective of the training is to optimize the parameters so that the gap between the ZF and the proposed algorithm output is minimized and possibly reduced to zero. The cost function used in training the DUN is the mean-squared error (MSE) between the estimated symbol vector obtained through ZF detection $\mathbf{x}_{ZF}$ and the output of the DUN $\bar{\mathbf{x}}^{(L)}$, which can be expressed as:

$$\mathcal{L}(\mathbf{x}_{ZF}, \bar{\mathbf{x}}^{(L)}) = \frac{1}{M}\sum_{i=1}^{M} ||\mathbf{x}_{ZF} - \bar{\mathbf{x}}^{(L)}||^2, \tag{6.39}$$

where M is the batch size. In backpropagation, the adaptive moment estimation (ADAM) optimizer [215] has been employed to optimize the learnable parameters. Once the training process is over, the optimized parameters are stored and used in respective detection algorithms. Consequently, the AC-SOR, AC-AOR, SOR, and AOR can retrieve the relevant parameters automatically based on the MIMO system configuration. The proposed methods used offline training to optimize the iterative methods' parameters. During the training phase, the model is trained using multiple batches of random channel matrices at a range of SNR values. After training convergence, the optimized parameters are stored. It should be noted that we do not need to retrain AC-SORNet and AC-AORNet methods for each channel matrix (say, The SNR is fixed for the channel during the training case). In addition, a feasible implementation plan is illustrated in Figure 6.4.

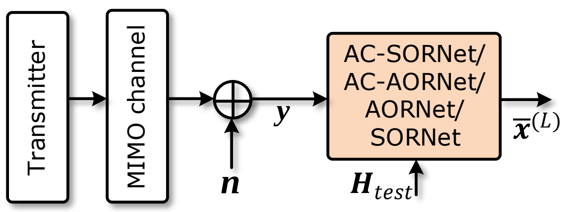


Figure 6.4: Block diagram of the testing phase of AC-SORNet, AC-AORNet, SORNet, and AORNet.

6.5 Computational complexity analysis

In this section, first, the complexity of SOR and AOR methods has been analyzed. Then, the complexity of the proposed AC-SORNet and AC-AORNet is analyzed based on the complexity derived for SOR and AOR methods, respectively. From the AOR iteration defined in (6.6), we can express the mth element $\hat{\mathbf{x}}^{(l+1)}$ as

$$\hat{x}_m^{(l+1)} = \frac{1}{W_{m,m}}\left[(1-\alpha)W_{m,m}\hat{x}_m^{(l)}+(\alpha-\beta)\sum_{k=1}^{m-1} W_{m,t}\hat{x}_m^{(l)}+\alpha\sum_{t=m+1}^{2T} W_{m,t}\hat{x}_m^{(l)}+\beta\sum_{t=m+1}^{2T} W_{m,t}\hat{x}_m^{(l+1)}+\alpha\tilde{y}_m^{MF}\right]. \tag{6.40}$$

Similarly, the mth element of the SOR iteration as defined in (6.4) ,can be expressed as

$$\begin{aligned}\hat{x}_m^{(l+1)} &= \frac{1}{W_{m,m}}\left[(1-\omega)W_{m,m}\hat{x}_m^{(l)} + \omega\sum_{t=1}^{m-1} W_{m,t}\hat{x}_m^{(l)}\right.\\ &\left. + \omega\sum_{t=m+1}^{2K} W_{m,t}\hat{x}_m^{(l+1)} + \tilde{y}_m^{MF}\right].\end{aligned} \tag{6.41}$$

where $\hat{x}_m^{(l+1)}$, $\hat{x}_m^{(l)}$, and y_m^{MF} denote the mth element of $\hat{\mathbf{x}}^{(l+1)}$, $\hat{\mathbf{x}}^{(l)}$, and $\mathbf{y}^{MF}$, respectively, and $W_{m,t}$ is the (m,t)th element of $\mathbf{W}$. To calculate the number of multiplications required for $\hat{x}_m^{(l+1)}$ in the AOR method, we can use the following formula: for computing $\sum_{t=1}^{m-1} W_{m,k}\hat{x}_m^{(l)}$, we require $m-1$ multiplications; similarly, for calculating $\sum_{t=m+1}^{2T} W_{m,k}\hat{x}_m^{(l)}$, we need $2T-m$ multiplications. For the diagonal matrix $W_{m,m}\hat{x}_m^{(l)}$, only one multiplication is necessary.

Using the previous discussions, in the AOR method, the number of real multiplications required for L iterations is $2L(3T^2+7T)$, and for the SOR method, it is $4L(2T^2+4T)$.

To calculate the complexity of the AC-SOR and AC-AOR methods, we must consider the additional multiplications required in (6.34), where two scalar parameters p_{l+1} and γ are multiplied with $2T$ dimensional vectors. Thus, an additional $4T$ real multiplications are required, bringing the total number of real multiplications for the AC-SOR to $4L(4T^2+8T)$ and for the AC-AOR to $2L(3T^2+15T)$. The complexity of the training and testing phases of a deep learning network is different. Generally, the complexity of the training phase is higher than that of the testing phase, as the training phase requires MSE computation and back-propagation to optimize the learnable parameters. From the training scheme presented in figure 6.3, the proposed AC-SORNet and AC-AORNet need the ZF Output. Therefore, the training phase complexity of the proposed AC-SORNet (or AC-AORNet) will be approximately equal to the sum of ZF Output complexity and AC-SOR (or AC-AOR) complexity. It should be noted that the computations required for back-propagation are not considered here, as the number of layers and the number of trainable parameters are small enough. On the other hand, the testing phase of DUN does not require any ZF operation, as shown in Figure 6.4. Therefore, the evaluation phase of AC-SORNet and AC-AORNet is identical to AC-SOR and AC-AOR, respectively. In the evaluation phase, the only difference between AC-SORNet (AC-AORNet) and AC-SOR (AC-AOR) is the iteration parameters optimized by the DUN method. As a result, the computational complexity of AC-SORNet and AC-AORNet is identical to AC-AOR and AC-SOR in the evaluation phase.

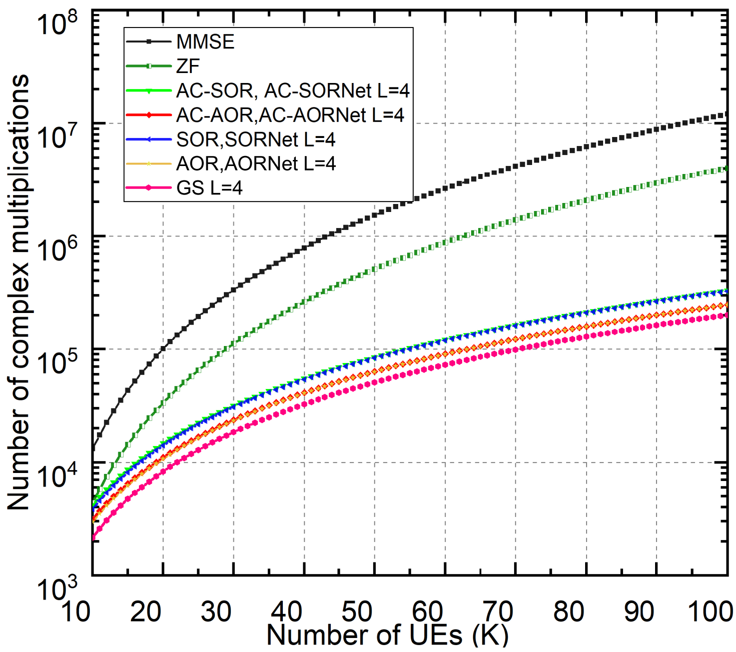


Figure 6.5: Complexity comparison against the number of users.

The number of complex multiplications performed by the ZF and MMSE detector and the state-of-the-art iterative techniques for various numbers of users T is depicted in Figure 6.5. The curve of complex multiplication for the MMSE and ZF-based detectors dramatically increases with different users, while the proposed methods AC-SORNet and AC-AORNet increase slowly. The computational complexity of the ZF and MMSE detection techniques is substantially larger than that of the AC-SORNet and AC-AORNet iterative algorithms when the number of transmit antennas $T \gg 10$. The reason is that the ZF and MMSE detection algorithms use complex multiplications in the order of $O(T^3)$. In contrast, the AC-SORNet and AC-AORNet iterative algorithms have an order of $O(T^2)$. As the increase in computational complexity due to the number of users T is relatively small for both AC-SORNet and AC-AORNet, it is possible to minimize the error rate with an appropriate increment in the number of iterations.

6.6 Performance analysis

In this section, we evaluate the performance of the proposed AC-SOR, AC-AOR, and their deep unfolded versions in a massive MIMO system and compare them with state-of-the-art

Table 6.1 Computational complexity comparison

Methods	Number of Multiplications
ZF/MMSE	$4(T^3 + 3T)$
AC-SOR/AC-SORNet	$4L(2T^2 + 8T)$
AC-AOR/AC-AORNet	$2L(3T^2 + 15T)$
SOR/SORNet	$4L(2T^2 + 4T)$
AOR/AORNet	$4L(3T^2 + 7T)/2$
Gauss-Seidel [112]	$4(L-1)T^2 + 16T$

iterative methods. We first briefly detail the training setup for the proposed methods and then show the MSE plot for different DUNs. Next, we present the simulation results for the proposed and other existing detectors under i.i.d. MIMO channels, correlated Rayleigh MIMO channels. Additionally, we investigate the proposed detection schemes under a realistic channel scenario. For simplicity, the following sections consider a 16×64 MIMO antenna configuration if it is not mentioned.

6.6.1 Implementation details

The Google Colab online platform has been used to train and validate the proposed methods. Several machine-learning platforms are available to train and test the models. In this work, PyTorch has been used for training and testing the detectors. Additionally, GPU acceleration has been employed to accelerate the training phase of the proposed deep unfolded networks.

Table 6.2 Simulation setup

Type	Value/Name
Programming language	Python 2.7
Machine learning framework	PyTorch
N_t	16
N_r	64
Channel	Rayleigh fading and Correlated Channel
Optimizer	Adam
Learning rate	10^{-5}
Epoch	5
Batch size	500
Training iterations	10^4
Modulation	256-QAM and 64-QAM

Table 6.3 List of trainable parameters for different algorithms

DUN Algorithm	Initialization of trainable parameters
SORNet	$\omega_l \sim \mathcal{N}(1.05, 0.01)$
AORNet	$\alpha_l \sim \mathcal{N}(1.09, 0.01)$, $\beta_l \sim \mathcal{N}(1.14, 0.01)$
AC-SORNet	$\omega_l \sim \mathcal{N}(1.05, 0.01)$ $\gamma_l \sim \mathcal{N}(0.7, 0.01)$, $p_l \sim \mathcal{N}(0.9435, 0.01)$
AC-AORNet	$\alpha_l \sim \mathcal{N}(1.09, 0.01)$, $\beta_l \sim \mathcal{N}(1.14, 0.01)$, $\gamma_l \sim \mathcal{N}(0.7, 0.01)$, $p_l \sim \mathcal{N}(1.0909, 0.01)$

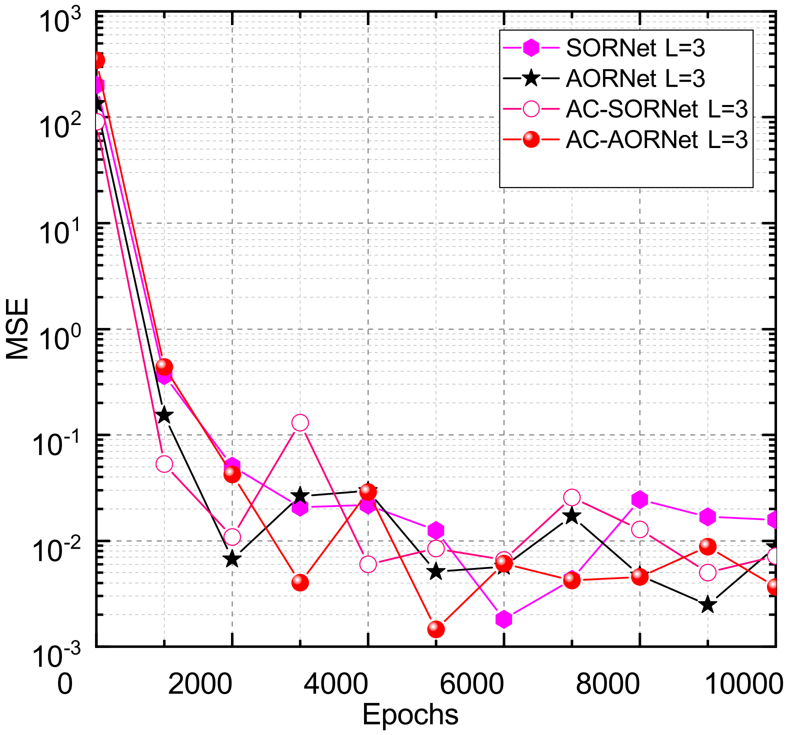


Figure 6.6: MSE vs epochs iterations for different deep unfolded networks with 64-QAM.

The simulation setup for the DUN-based detector is given in Table 6.2. The training dataset consists of a batch of transmit vectors $\tilde{\mathbf{x}}$, channel matrices $\tilde{\mathbf{H}}$, and received vectors $\tilde{\mathbf{y}}$. The dataset is generated randomly for each training iteration according to their distribution. We set the number of epochs to five for training each batch of the dataset. We use 10^4 training iterations where, at each iteration, a batch of the dataset is generated. In this work, we consider the Adam optimizer during backpropagation in DUN. The learning rate is set to 10^{-5}, and the batch size is set to 500. We consider a small range of SNR for training purposes. The SNR values for training are generated using a uniform distribution defined by $\mathcal{U}(SNR_{min}, SNR_{max})$. For example, we choose the SNR range of 10 dB to 15 dB for 256-QAM constellations and 7 dB to 12 dB for 64-QAM constellations. Training a network using different SNR values can make the network more robust and able to detect signals over a wider range of SNR values. It also helps to prevent overfitting of the learnable parameters.

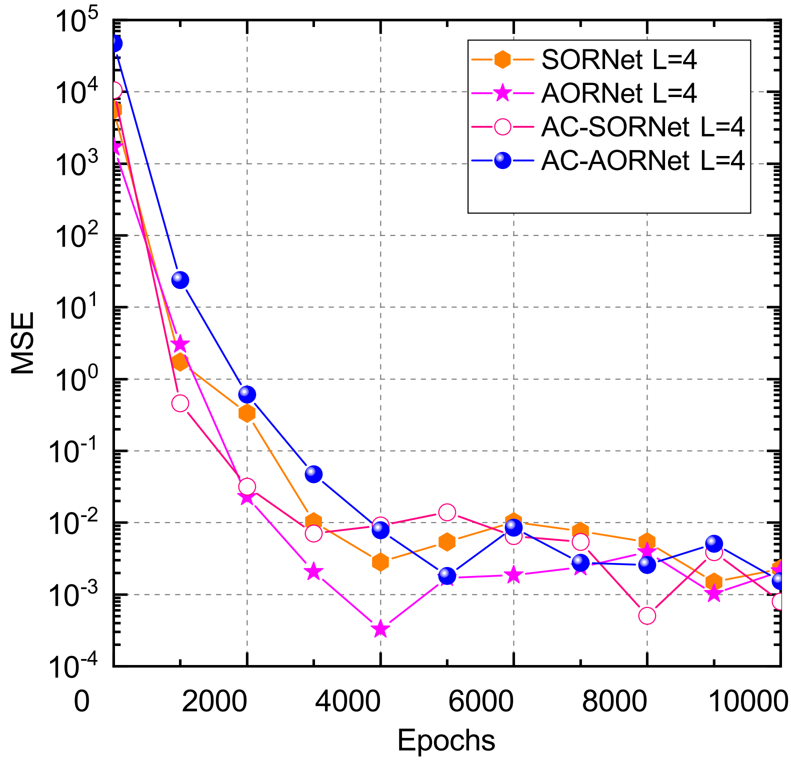


Figure 6.7: MSE vs. epochs iterations for different deep unfolded networks with 256-QAM.

Properly initializing the trainable parameters is crucial in achieving good performance with a small training data set. Our proposed method initializes the trainable parameters with randomly generated Gaussian variables. The mean and variance of the Gaussian distribution are predefined and can significantly affect the convergence and accuracy of the training process. Furthermore, appropriate initialization of the parameters can also prevent the model from getting stuck in local minima and help it converge faster to the global minimum. Therefore, we carefully choose the initial values of the parameters to ensure the best performance of the proposed method. According to [212], the under-extrapolation coefficient γ_l is suitable for sufficient convergence in the $[0.6, 0.9]$. Therefore, in our training process, we set the mean initial value of γ_l to 0.7. Next, we set assign the mean value of the under-relaxation parameter p_{l+1} to 1. Based on the numerical analysis presented in [216], SOR achieves near-optimal detection performance when $\omega_l \approx 1.05$. The AOR method has two parameters, namely α and β. Based on the numerical analysis presented in [96], we set the mean value of α_l and β_l to 1.09 and 1.14, respectively. However, the initial parameter values we defined earlier do not result in optimal performance and hence require optimization through training. In all cases, we set the variance to a minimal value, such as 0.01, to avoid considerable variation from the mean value. It is important to note that only the under-relaxation parameter p_{l+1} is defined layer-wise. The other parameters, such as ω_l in SOR, α_l and β_l in AOR, and γ_l, have been considered the same for all layers. Table 6.3 lists the distributions that generate the initial random values for all trainable parameters. Figures 6.6 and 6.7 show the learning curves of the proposed detector with 4 and 5 layers, respectively, for 64-QAM and 256-QAM constellations. The MSE, calculated using (6.39), is plotted against the number of training iterations. As the number of iterations increases, the MSE decreases, indicating a near ZF detection performance. During the training process, the MSE stabilizes at around 10^{-2} and 10^{-3}, indicating that the parameters have been optimized and further training is unnecessary.

6.6.2 IID MIMO channel performance

In this section, we evaluate the performance of the ZF, SOR, AOR, AC-SOR, AC-AOR, SORNet, AORNet, AC-SORNet, and AC-AORNet detectors under an IID MIMO channel with 64-QAM and 256-QAM constellations. Figure 6.8 shows the SER versus SNR

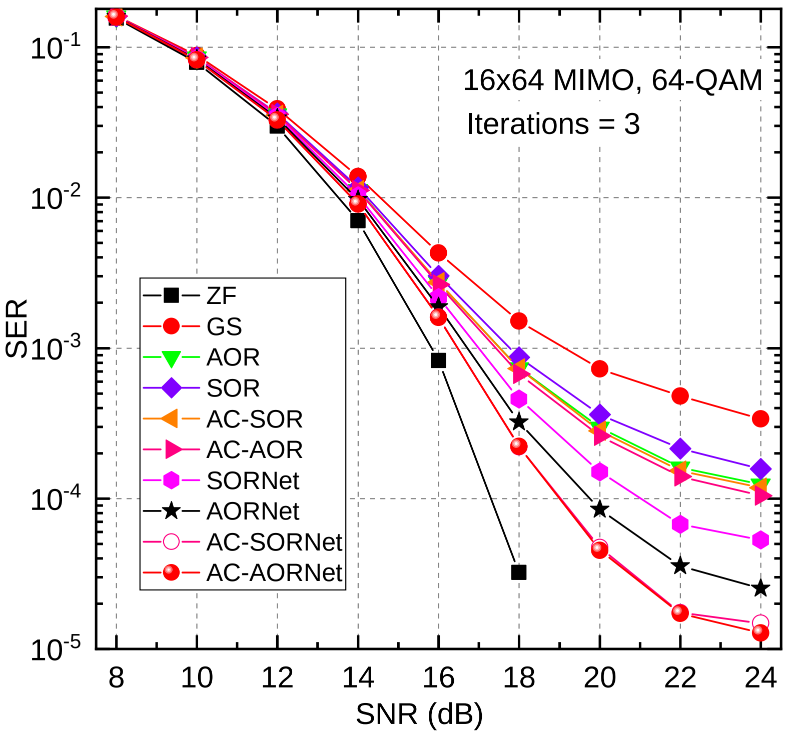


Figure 6.8: SER performance of the proposed method and conventional method versus SNRs is evaluated under 16×64 Rayleigh MIMO channel with 64-QAM modulation.

variation plot for the 64-QAM modulation scheme. The results indicate that SOR and SORNet have vastly different performances. For instance, at a target BER of 10^{-3}, the SOR method requires an SNR of approximately 18 dB, whereas the SORNet requires around 17 dB SNR. Therefore, the SORNet significantly outperforms conventional SOR due to its optimized parameters. Similarly, AORNet outperforms conventional AOR by a large margin. Additionally, the detection schemes combined with the Accelerated Chebyshev method exhibit better performance than the conventional iterative detection schemes. Finally, the AC-SOR and AC-AOR, jointly with training parameters, have further improved performance. For example, AC-SORNet and AC-AORNet achieved near-zero-forcing detection performance for only three iterations.

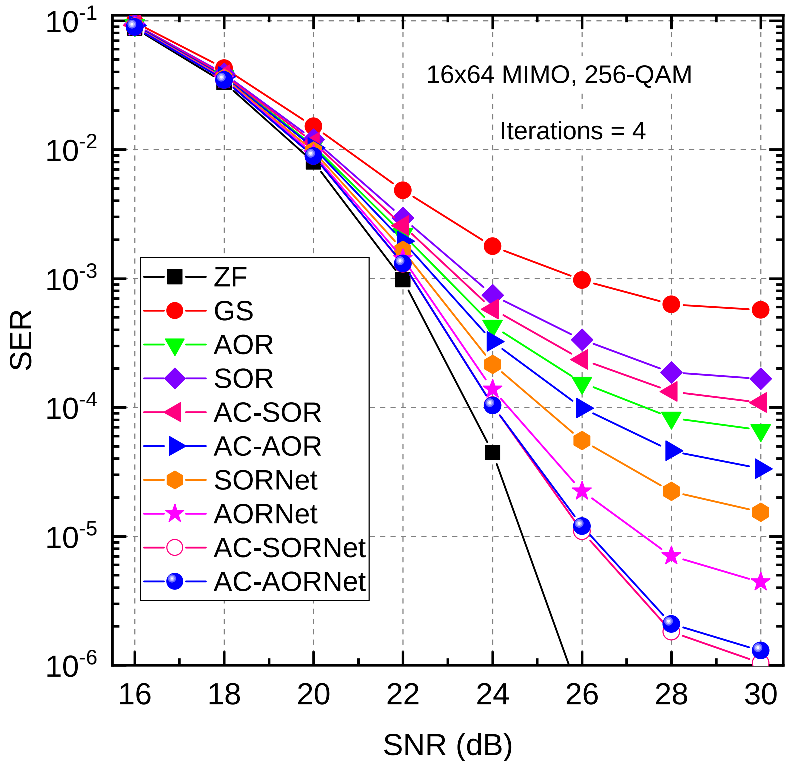


Figure 6.9: SER performance of the proposed method and conventional method versus SNRs is evaluated under 16×64 Rayleigh MIMO channel with 256-QAM modulation.

Figure 6.9 illustrates the SER performance for 256-QAM constellations, where the proposed methods and their DUN-versions are compared against conventional SOR, AOR, and ZF detectors. In higher-order modulation, we need a higher number of iterations to achieve satisfactory results. For example, a number of iterations have been set to four for 256-QAM modulation. Similar to 64-QAM modulation, both AC-SORNet and AC-AORNet methods exhibit a significant performance improvement in 256-QAM modulation.

The impact of different numbers of antenna at the base station on the SER performance is shown in figure 6.10, where the number of users and the value of SNR has been set to 32 and 16dB, respectively. We can see when R increases, the performance improves significantly. Moreover, the proposed methods outperform other methods except ZF in this case.

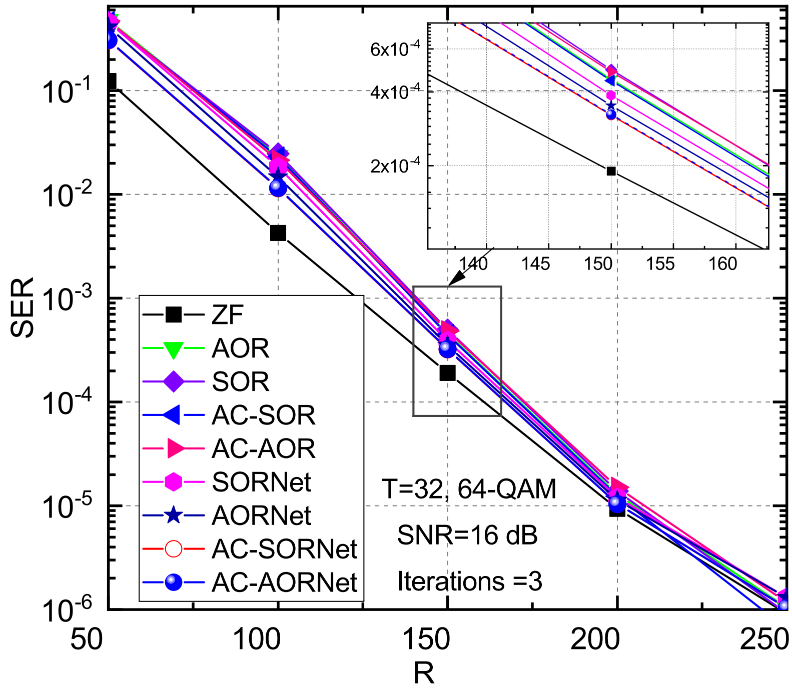


Figure 6.10: SER performance of the proposed method and conventional method versus the number of antenna at the base station is evaluated under $T = 32$ and SNR=16 dB with Rayleigh MIMO channels and 64-QAM modulation.

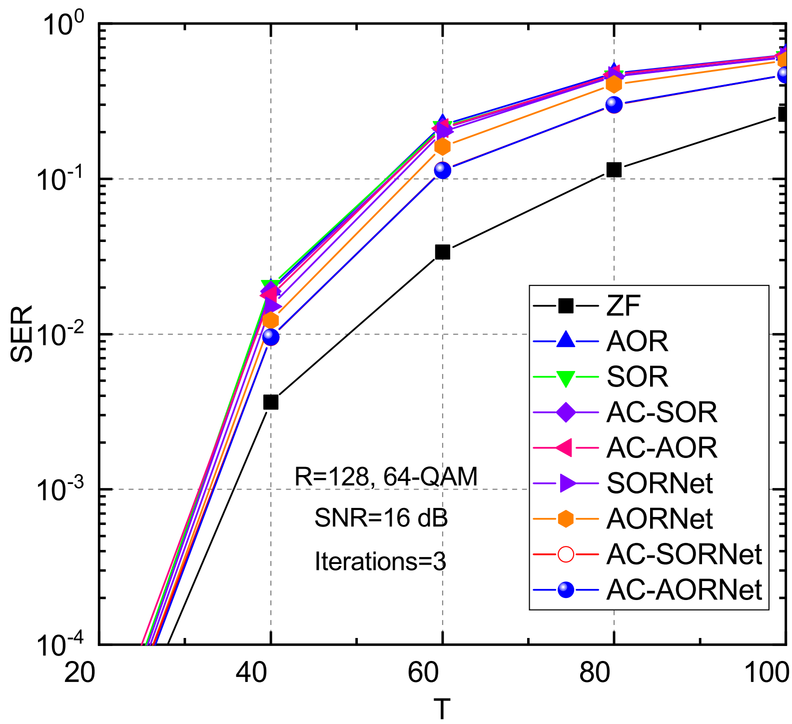


Figure 6.11: SER performance of the proposed method and conventional method versus a number of users is evaluated under $R = 128$ and SNR=16 dB with Rayleigh MIMO channels and 64-QAM modulation.

The impact of varying the number of users on the SER performance for $R = 128$ at an SNR of 16 dB is illustrated in Figure 6.11. When the number of users is relatively low, the difference in performance between the proposed DUNs (i.e. AC-SORNet, and AC-AORNet) and exact ZF detection is negligible. In contrast, conventional techniques perform inferior to the ZF solution. Among the other detection algorithms, the proposed method shows a performance close to exact ZF detection, particularly for the values of users up to 40. However, there is a noticeable performance gap between ZF and other methods for a higher number of users.

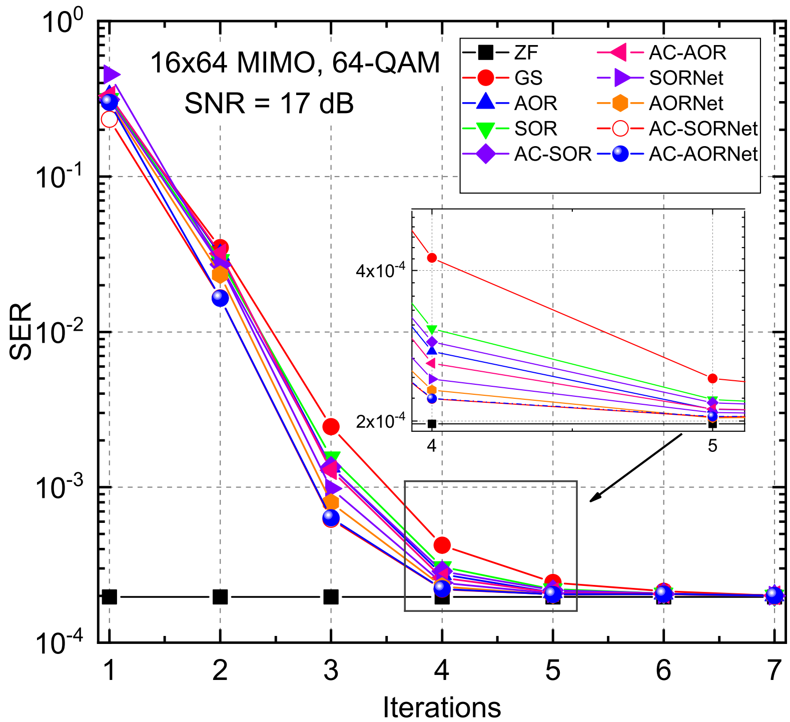


Figure 6.12: SER performance of the proposed method and conventional method versus a number of iterations is evaluated under 16×64 Rayleigh MIMO channels with 64-QAM modulation.

Figures 6.12 and 6.13 show the SER performance against the number of iterations or layers for 64-QAM and 256-QAM constellations, respectively. The results show that for 64-QAM, the proposed method can achieve the same performance as ZF with only 3 or 4 iterations, while for 256-QAM, it requires around 4 or 5 iterations to achieve the same performance as ZF. Moreover, the proposed DUN-based detectors outperform the conventional iterative detectors with the same number of iterations.

In Section 6.5 the complexity analysis has been presented for all methods. It is noticed from Table 6.1 that all algorithms have the same computational complexity of $O(T^2)$, except for the Zero Forcing detector, which has $O(T^3)$ complexity. Since the state-of-the-art algorithms have similar computational complexity order, the focus is on their error-rate performance and the number of iterations required to achieve a result close to ZF. As demonstrated in the following section, the proposed AC-SORNet and AC-AORNet

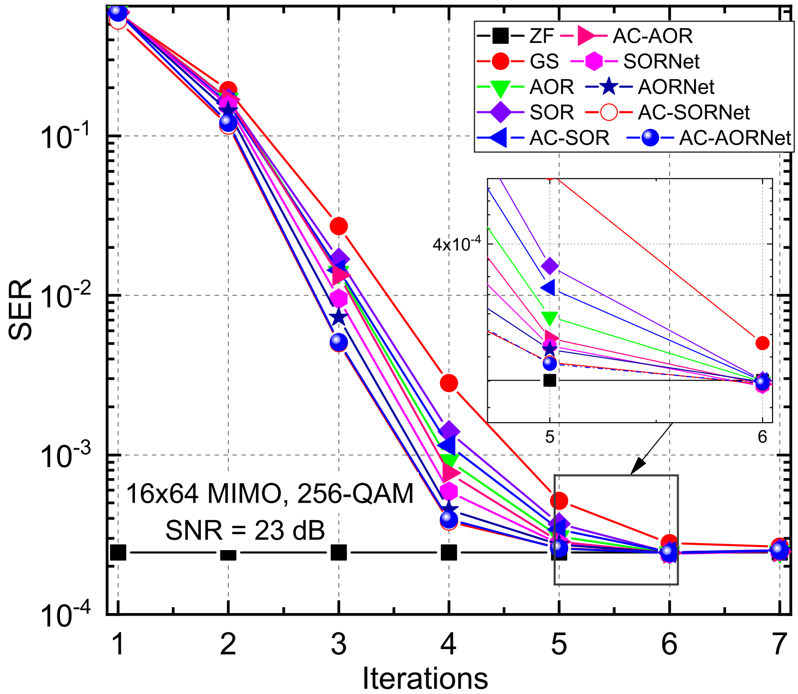


Figure 6.13: SER performance of the proposed method and conventional method versus a number of iterations is evaluated under 16×64 Rayleigh MIMO channels with 256-QAM modulation.

algorithms outperform the state-of-the-art algorithms with fewer iterations, resulting in lower implementation complexity. The proposed method achieves the ZF performance with only a few iterations.

Tables 6.4 and 6.5 present the SNR required to achieve an SER of 10^{-3}, as well as the number of real multiplications need for different detection techniques. Table 6.4 considers 64-QAM modulation, while Table 6.5 considers the same system but with 256-QAM modulation.

The tables show that the AC-SORNet and AC-AORNet require the same number of multiplications as the original AC-SOR and AC-AOR methods, respectively. However, the proposed methods require lower SNR to achieve SER=10^{-3} compared to the original AC-SOR and AC-AOR methods. Although the degradation is larger for 256-QAM, but the proposed methods still require a lower SNR than the original methods. These results demonstrate that the proposed methods can perform better with fewer computational resources.

Table 6.4 Number of multiplications and SNR required to achieve an SER of 10^{-3} for 16×64 and 64-QAM

Method	Multiplications ($\times 10^5$)	SNR(dB)
ZF	31.9488	15.9443
AC-SORNet	1.02912	16.8744
AC-AORNet	0.19296	16.8709
AC-SOR	0.77184	17.7321
AC-AOR	0.77184	17.6671
SORNet	1.01376	17.3677
AORNet	0.76416	17.1279
SOR	1.01376	17.8776
AOR	0.76416	17.7346
Gauss-Seidel [112]	0.6656	19.3041

Table 6.5 Number of multiplications and SNR required to achieve an SER of 10^{-3} for 16×64 and 256-QAM

Method	Multiplications ($\times 10^5$)	SNR(dB)
ZF	31.9488	21.9930
AC-SOR	1.37216	23.5764
AC-AOR	1.02912	23.1712
AC-SORNet	1.37216	22.5008
AC-AORNet	1.02912	22.5024
SORNet	1.35168	22.9051
AORNet	1.01888	22.6575
SOR [87]	1.35168	23.7659
AOR [93]	1.01888	23.3609
Gauss-Seidel [112]	0.82944	25.9243

6.6.3 Performance in correlated MIMO channel

To study the impact of spatial correlation on the performance of the proposed methods, the exponential correlation model has been used in this section [68]. The correlation factor ζ $(0 \leq \zeta \leq 1)$ represents the correlation between two adjacent antennas. We consider the number of iterations or layers in a correlated channel as $L = 4$ and $L = 5$ for 64-QAM and 256-QAM constellations, respectively. It is observed that the detection performance is affected significantly by spatial correlation variation, as shown in Figures 6.14 and 6.15. The proposed AC-SORNet and AC-AORNet methods converge to the ZF performance at the low correlation ($\zeta = 0.3$). Although, for higher correlations ($\zeta = 0.7$) the performance of the proposed methods is worse than the ZF method, it still performs better than the state-of-the-art methods.

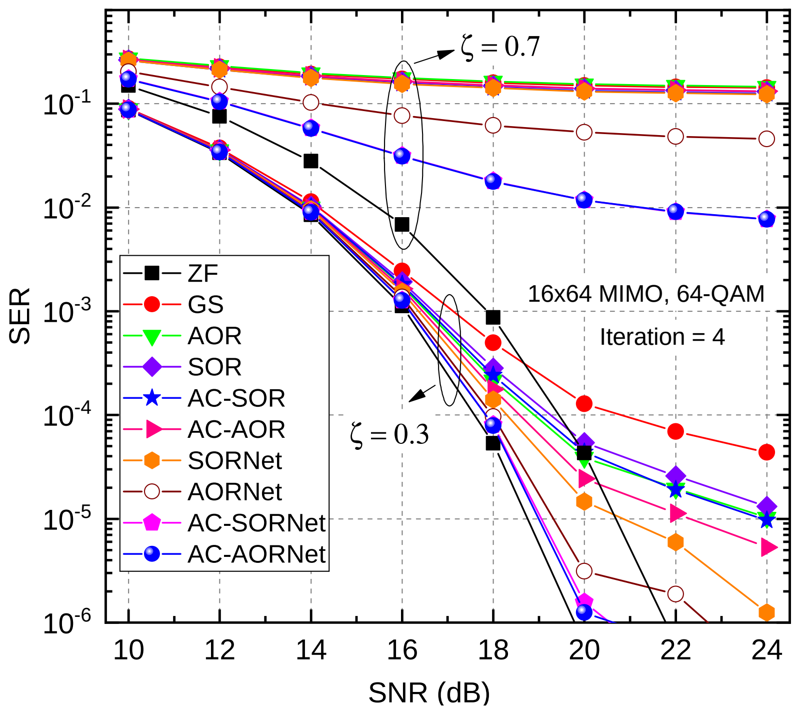


Figure 6.14: SER performance of the proposed method and conventional method versus SNRs is evaluated under 16×64 correlated MIMO channels with 64-QAM modulation.

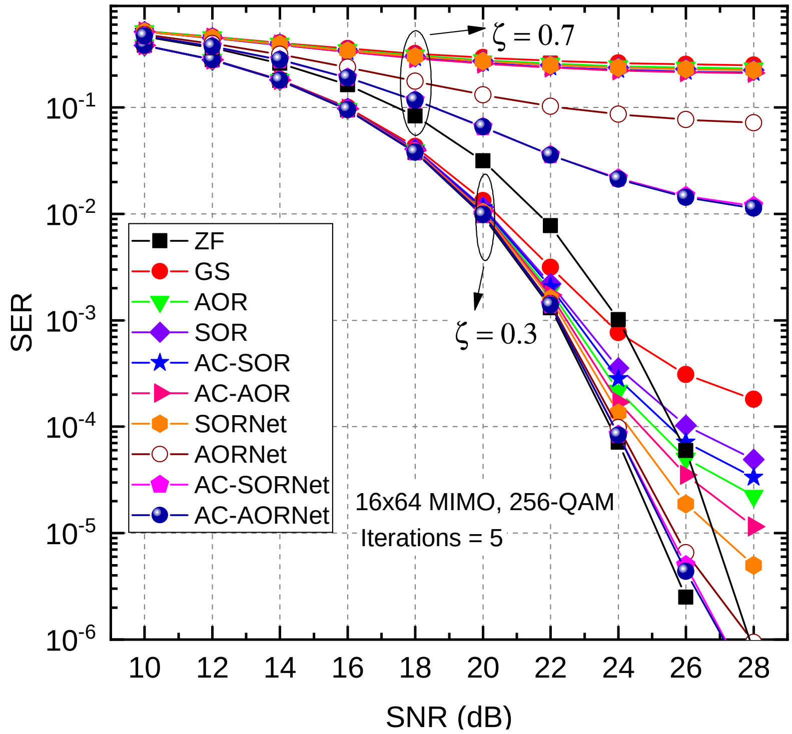


Figure 6.15: SER performance of the proposed method and conventional method versus SNRs is evaluated under 16×64 correlated MIMO channels with 256-QAM modulation.

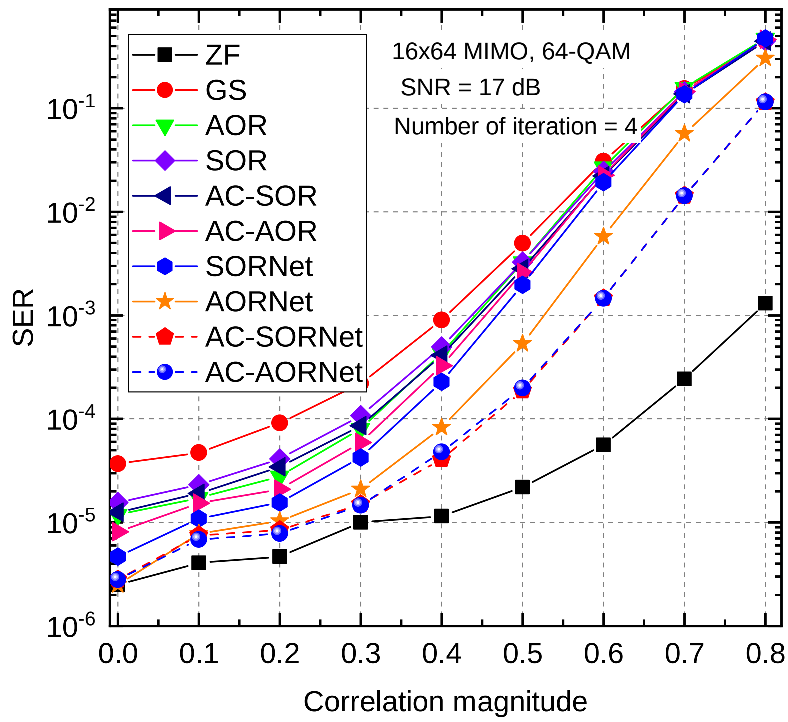


Figure 6.16: Impact of the correlated magnitude on the SER performance for 16×64 correlated MIMO channels with 64-QAM modulation.

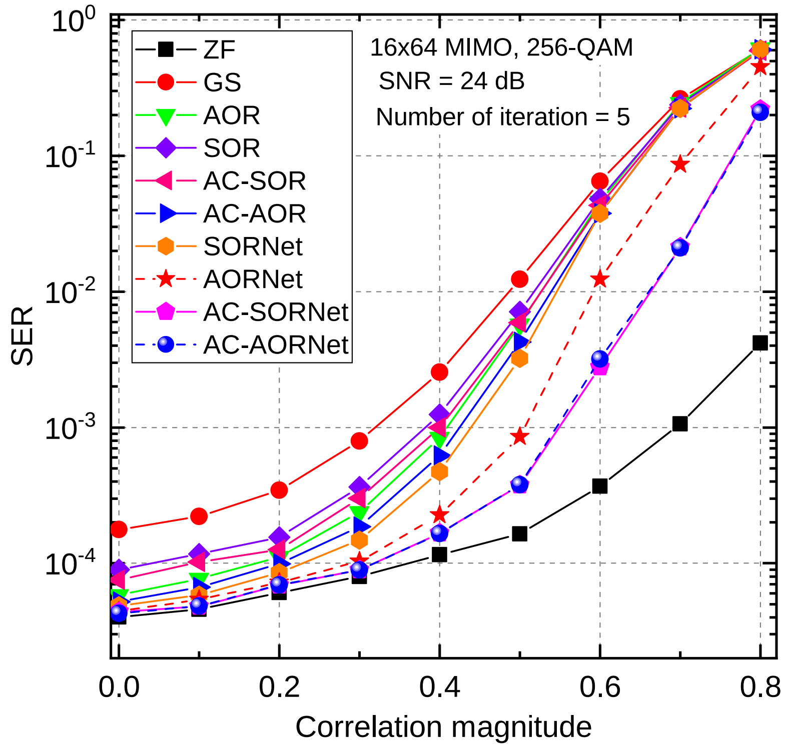


Figure 6.17: Impact of the correlated magnitude on the SER performance for 16×64 correlated MIMO channels with 256-QAM modulation.

Figures 6.16 and 6.17 present the SER versus the ζ variation plot for 64-QAM and 256-QAM constellations, respectively. Increasing the correlation leads to significant performance degradation for all methods. However, the proposed methods perform better, outperforming the other algorithms. For instance, Figure 6.16 demonstrates that the proposed methods achieve near-ZF performance when the correlation magnitude is between $\zeta = 0.1$ and 0.4. Similarly, Figure 6.17 shows that for 256-QAM modulation, the proposed methods provide near-ZF performance for correlation values of $\zeta = 0.1$ to 0.3.

6.6.4 Performance in a Typical Channel

This section uses WINNER II typical urban micro channel model has been used [217] to evaluate the proposed methods under a realistic propagation environment. The carrier frequency and the number of active sub-carriers have been set to 2 GHz and 1200, respectively. The corresponding BER performance for a (16×64) massive MIMO scenario with 64-QAM and 256-QAM constellations are depicted in figures 6.18 and 6.19, respectively. It is evident that the proposed methods, AC-AORNet and AC-SORNet, show superior performance compared to other existing methods, making them suitable detectors for practical channels. Although performance degradation is observed with high-order modulation, such as 256-QAM, the proposed methods still outperform other methods.

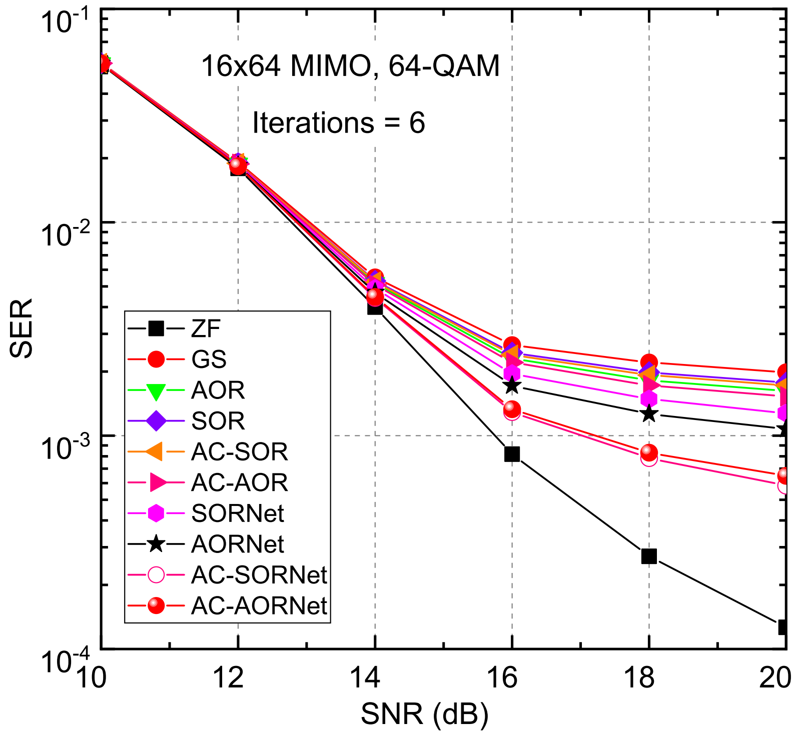


Figure 6.18: SER performance of the proposed method and conventional method versus SNRs is evaluated under 16×64 realistic channel with 64-QAM modulation.

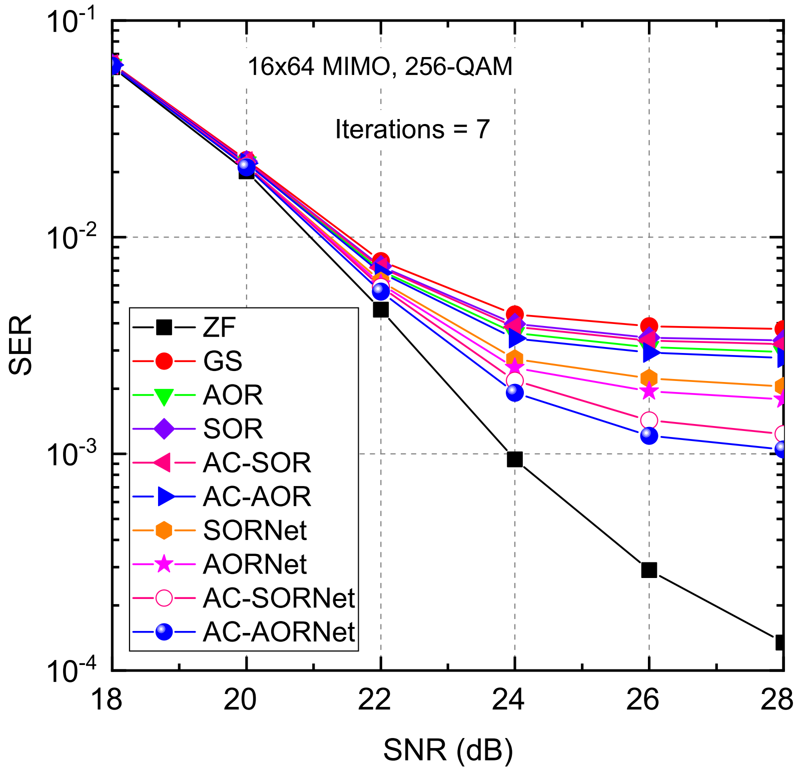


Figure 6.19: SER performance of the proposed method and conventional method versus SNRs is evaluated under 16×64 realistic channel with 256-QAM modulation.

6.7 Conclusions

In this chapter, we presented two iterative massive MIMO detectors, namely AC-AOR and AC-SOR, formulated using the accelerated Chebyshev method to enhance the performance of conventional AOR and SOR methods, respectively. Moreover, we proposed DUNs for the AC-SOR and AC-AOR methods to optimize the algorithm's parameters through training. The proposed DUNs are simple and fast to train and require only a few adjustable parameters to be optimized. Due to the offline training of the proposed methods, the number of trainable variables of the AC-SORNet and AC-AORNet is independent of the number of antennas. It is only determined by the number of layers, which leads to a speedy and stable training process and reasonable scalability for larger systems. The computational complexity of AC-AORNet and AC-SORNet is the same as that of AC-AOR and AC-SOR, respectively. Our performance results showed that the AC-SORNet and AC-AORNet detectors outperform the AC-SOR and AC-AOR detectors in all settings, indicating that DUN can enhance the AC-SOR and AC-AOR detectors by learning the optimal parameters. Furthermore, we conducted experiments in various channel scenarios, such as non-ideal channel conditions, including spatially correlated channel scenarios and realistic propagation environment models. The results demonstrate that the proposed method outperforms existing detectors.

Chapter

7 A Low Complexity Iterative SC-FDE Receiver for Massive MIMO Systems

Berra, S., Chakraborty, S., & Dinis, R. (2024). A Low Complexity Iterative SC-FDE Receiver for Massive MIMO Systems. IEEE Transactions on Vehicular Technology.

7.1 Introduction

The conventional approach for massive MIMO systems with frequency-selective channels is to employ an OFDM scheme and perform MIMO operations at the subcarrier level [157]. However, OFDM signals are known to have large envelope fluctuations and a PAPR, leading to amplification difficulties [218]. On the other hand, the SC-FDE scheme has a much lower PAPR for wireless transmission [219]. Also, SC-FDE receiver can employ the IB-DFE method to remove residual interference iteratively [175]. The IB-DFE receivers, optimized under the MMSE criterion, show excellent detection performance, which is close to the matched filter bound (MFB) [172]. Still, the complexity of IB-DFE receivers is significantly high for massive MIMO systems due to the large dimensional matrix inversion for each iteration and subcarrier [177]. Although few IB-DFE receivers like MRC [179] and EGC [180] do not require matrix inversions, these methods exhibit poor detection performance for higher-order constellations. They are only suitable for massive MIMO systems where the constellation size is small and the number of receiving antennas is much larger than the number of users [181].

Therefore, to implement the linear IB-DFE receiver in a massive MIMO system, replacing the large dimensional inversion matrix with a low-complexity solution is required. However, the conventional IB-DFE receiver strictly requires the exact matrix inversion operation. This motivates this work to develop an alternate IB-DFE receiver that can avoid the matrix inversion operation and replace the inversion operations with low-complexity methods based on iterative techniques.

In the context of large dimensional matrix inversion, used in linear detection techniques such as ZF and MMSE methods, various low complexity iterative techniques have already been proposed in the literature [84]. As a result, these iterative detectors can provide an effective balance between performance and complexity. These algorithms have two categories: approximate matrix inversion methods and iterative techniques to solve linear equations [84]. The first category is based on approximate matrix inversion methods such as the NI method [81, 124], the NS method [119], and the TPE method [220], etc. These methods provide a way to simplify matrix inversion by expressing it as a series of low-complexity matrix operations. For example, in [81] NI has been used to achieve the MMSE performance at a much lower complexity. The convergence rate of NI is further improved in [125] using the Tchebychev polynomial to determine the initial solution. This approach reaches optimum performance with only a few iterations.

For the second category, iterative techniques such as the JA method [113], RI method [89], SOR method [87, 221], AOR method [88], GS method [112], CG method [90], etc. have been reported by different authors to simplify the MMSE detection. Each algorithm estimates the transmitted symbol without calculating the inversion matrix, and these approaches can present compelling trade-offs between the required computational complexity and achievable performance. Those techniques can also be applied in precoding for massive MIMO systems to reduce the complexity of Regularized ZF or ZF precoding [82, 93, 222]. Additionally, a similar work has been reported for precoding in [223, 224].

However, conventional iterative methods sometimes exhibit poor convergence rates and may require several iterations. Therefore, several techniques have been reported to address the slow convergence issue and improve the performance of iterative methods. One of these techniques is the Chebyshev polynomial acceleration approach. The Chebyshev polynomials construct a new sequence of iterations that can converge to the exact solution more quickly than the original solution obtained from a basic iterative method [95]. Recent works on Chebyshev polynomials shows that it enhances the overall detection performance with negligible complexity overhead [96, 97].

Apart from the above-mentioned methods, algorithms such as Gaussian elimination, Singular Value Decomposition (SVD) , and Cholesky decomposition [225] can be used instead of the exact matrix inversion required in ZF/MMSE. However, their time complexity and computational complexity are much higher than the previously mentioned iterative methods, and hence these techniques are impractical for massive MIMO systems. For example, SVD can offer good performance, but its complexity can be even higher than the exact matrix inversion [214, 226]. Moreover, SVD requires channel knowledge at both the transmitter and receiver side, and joint processing of the signals of all transmit and receive antennas.

Iterative detectors for massive MIMO receivers can work with flat fading channels or operate at the subcarrier level of OFDM-like systems [67, 227]. However, designing a low-complexity massive MIMO receiver for SC-FDE schemes, employing IB-DFE techniques, is still an open issue. The massive MIMO receivers proposed for the SC-FDE method do not consider IB-DFE receivers [228] or use IB-DFE receivers to cope with residual interference levels arising from a sub-optimum feedforward part [179, 180, 229]. In fact, to the best of our knowledge, no low-complexity massive MIMO detection techniques take advantage of the characteristics of IB-DFE receivers.

This chapter focuses on massive MIMO systems using SC-FDE modulations and an IB-DFE receiver. To reduce receiver complexity, we design an alternative structure that incorporates the Richardson iterative method to avoid the calculation of the inversion matrix. Furthermore, we apply the Chebyshev polynomial to accelerate the Richardson method.

The contributions of this chapter are as follows.

- A novel IB-DFE receiver structure is proposed in this work, which allows us to replace the exact matrix inversions within IB-DFE iterations with low-complexity iterative methods, used to solve systems of linear equations.

- In addition, we also propose the Chebyshev-accelerated Richardson method for the IB-DFE receiver to achieve conventional IB-DFE performance with much lower computational complexity. We also explore alternative versions of the IB-DFE receivers, utilizing existing iterative approaches like GS, CG, AOR, RI-NS, etc.

- We thoroughly analyze the computational complexity associated with the proposed IB-DFE method. Furthermore, we compare the complexities of various potential IB-DFE implementations with those of existing IB-DFE receivers. As a result, the proposed receiver can significantly reduce the overall complexity of each IB-DFE iteration, dropping from $O(T^3)$ to $O(T^2)$.

7.2 The proposed IB-DFE receiver for the massive MIMO system

7.2.1 Alternative IB-DFE structure

In the previous section, we have determined the optimal expression for $\mathbf{F}_k^{(i)}$ and $\mathbf{B}_k^{(i)}$. From (3.17), we notice that an exact matrix inversion is required for $\mathbf{F}_k^{(i)}$. For small-scale MIMO, the complexity of $\mathbf{F}_k^{(i)}$ is low. But, in the massive-MIMO case, the complexity of (3.17) is too high due to the large matrix size. We observe that $\mathbf{B}_k^{(i)}$ can only be determined after calculating $\mathbf{F}_k^{(i)}$. We aim to have a reduced complexity receiver that avoids the direct implementation of (3.17). To achieve this, we use (3.18) in (3.11) to have

$$\tilde{\mathbf{X}}_k^{(i)} = \mathbf{F}_k^{(i)}\mathbf{Y}_k - \left(\mathbf{F}_k^{(i)}\mathbf{H}_k - \mathbf{I}_T\right)\hat{\mathbf{X}}_k^{(i-1)} = \mathbf{F}_k^{(i)}\left(\mathbf{Y}_k - \mathbf{H}_k\hat{\mathbf{X}}_k^{(i-1)}\right) + \hat{\mathbf{X}}_k^{(i-1)}. \tag{7.1}$$

Now, using (3.17), we can express (7.1) as

$$\tilde{\mathbf{X}}_k^{(i)} = \mathbf{\Gamma}_k^{(i)}\left(\alpha\mathbf{I}_T + \mathbf{H}_k^H\mathbf{H}_k(\mathbf{I}_T - (\mathbf{\Upsilon}_k^{(i-1)2}))\right)^{-1}\mathbf{H}_k^H \times \left(\mathbf{Y}_k - \mathbf{H}_k\hat{\mathbf{X}}_k^{(i-1)}\right) + \hat{\mathbf{X}}_k^{(i-1)}, \tag{7.2}$$

which leads to

$$\tilde{\mathbf{X}}_k^{(i)} = \mathbf{\Gamma}_k^{(i)}\left(\alpha\mathbf{I}_T + \mathbf{H}_k^H\mathbf{H}_k(\mathbf{I}_T - (\mathbf{\Upsilon}_k^{(i-1)2}))\right)^{-1} \times \left(\mathbf{Y}_k^{MF} - \mathbf{H}_k^H\mathbf{H}_k\hat{\mathbf{X}}_k^{(i-1)}\right) + \hat{\mathbf{X}}_k^{(i-1)}. \tag{7.3}$$

We can rewrite $\tilde{\mathbf{X}}_k^{(i)}$ as

$$\tilde{\mathbf{X}}_k^{(i)} = \mathbf{\Gamma}_k^{(i)}(\mathbf{A}_k^{(i)})^{-1}\mathbf{b}_k^{(i)} + \hat{\mathbf{X}}_k^{(i-1)}, \tag{7.4}$$

where $\mathbf{b}_k^{(i)} = \mathbf{Y}_k^{MF} - \mathbf{H}_k^H\mathbf{H}_k\hat{\mathbf{X}}_k^{(i-1)}$, and $\mathbf{A}_k^{(i)} = \alpha\mathbf{I}_T + \mathbf{H}_k^H\mathbf{H}_k(\mathbf{I}_T - \mathbf{\Upsilon}_k^{(i-1)2})$, which entails high-dimensional matrix inversion required for massive MIMO systems, with computation complexity $O(T^3)$. We can further write (7.4) as

$$\tilde{\mathbf{X}}_k^{(i)} = \mathbf{\Gamma}_k^{(i)}\breve{\mathbf{X}}_k^{(i)} + \hat{\mathbf{X}}_k^{(i-1)}, \tag{7.5}$$

where

$$\breve{\mathbf{X}}_k^{(i)} = (\mathbf{A}_k^{(i)})^{-1}\mathbf{b}_k^{(i)}. \tag{7.6}$$

Fig. 7.1 presents the conventional IB-DFE and the proposed structure. Although the proposed IB-DFE structure still involves the inversion matrix $\mathbf{A}_k^{(i)}$, but its structure is very simple and attractive. In the following, we propose iterative techniques to avoid matrix inversions and reduce receiver complexity.

Let us start by considering the characteristics of the matrix to be inverted $\mathbf{A}_k^{(i)} = \alpha\mathbf{I}_T + \mathbf{H}_k^H\mathbf{H}_k(\mathbf{I}_T - \mathbf{\Upsilon}_k^{(i-1)2})$. It is well known that the Gram matrix $\mathbf{G}_k = \mathbf{H}_k^H\mathbf{H}_k$ is semi-definite positive, i.e., its eigenvalues are non-negative ($\lambda(\mathbf{G}_k) \geq 0$). We also have $\lambda(\mathbf{I}_T - (\mathbf{\Upsilon}_k^{(i-1)2})) \geq 0$, since $\mathbf{\Upsilon}_k^{(i-1)2}$ is diagonal, and its diagonal elements are the absolute values of correlation factors that, by definition, are between 0 and 1. This means the eigenvalues of $\mathbf{H}_k^H\mathbf{H}_k(\mathbf{I}_T - (\mathbf{\Upsilon}_k^{(i-1)2}))$ are non-negative. By adding the term $\alpha\mathbf{I}_T$ (where $\alpha > 0$), we end up with a matrix whose eigenvalues are non-negative (i.e., $\lambda(\mathbf{A}_k) > 0$), which means that $\mathbf{A}_k$ is non-singular.

It is based on the condition that $R >> T$, which leads to the near-orthogonality of the channel matrix $\mathbf{H}_k$. In general, the diagonal elements of the Gram matrix $\mathbf{G}_k$ are larger than the other line (and column) elements, i.e., $|[\mathbf{G}_k]_{t,t}| \geq |[\mathbf{G}_k]_{t,t'}|$ for $t' \neq t$.

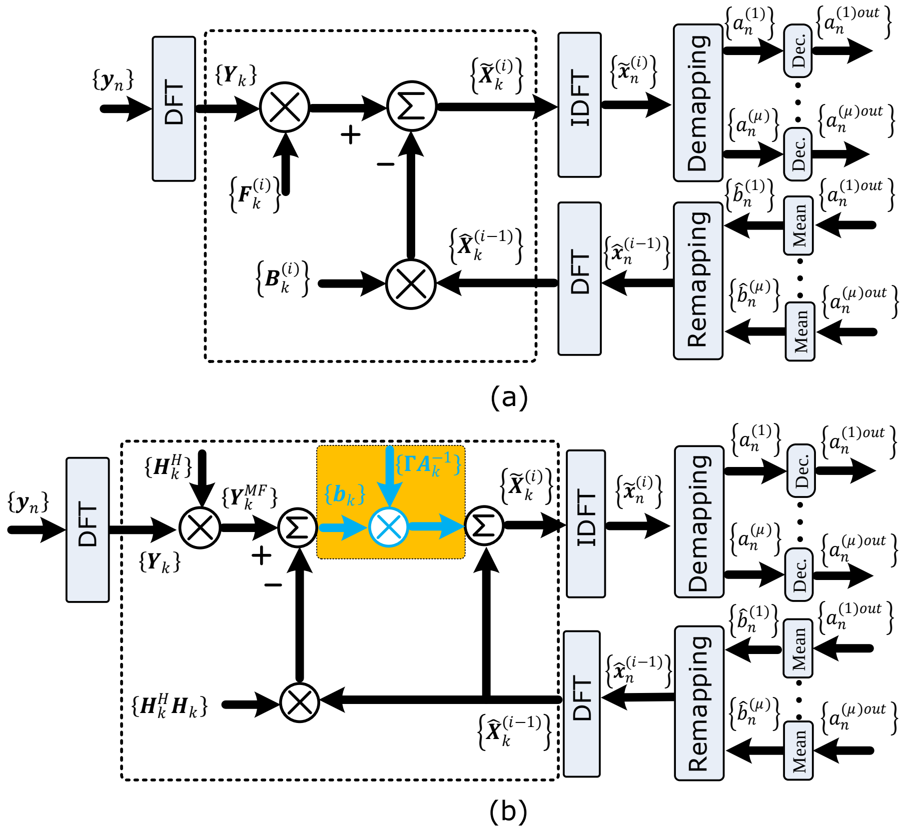


Figure 7.1: The block diagram of IB-DFE receivers; (a) the conventional method, (b) the proposed method.

However, for massive MIMO systems where $R >> T$, the Gram matrix becomes diagonally dominant [30]. That is, $|[\mathbf{G}_k]_{t,t}| >> |[\mathbf{G}_k]_{t,t'}|$ for $t \neq t'$, and in fact, the Gram matrix is strictly diagonally dominant for $|[\mathbf{G}_k]_{t,t}| > \sum_{t' \neq t} |[\mathbf{G}_k]_{t,t'}|$. Under reasonable propagation conditions (namely, low-to-moderate correlations between antennas), when $R >> T$, the channel matrix $\mathbf{H}_k$ is asymptotically orthogonal for massive MIMO system[67, 230], $\mathbf{G}_k$ is a suitable diagonal matrix.[1] Since the proposed receiver does not compute $\mathbf{F}_k^{(i)}\mathbf{H}_k$ as shown in Fig. 7.1, we have simplified the computation of $\mathbf{\Gamma}_k^{(i)}$ by assuming the diagonal approximation of $\mathbf{G}_k$.

[1]It should be noted that, although $\mathbf{G}_k$ is diagonally dominant for massive MIMO systems, the same is not necessarily true for $\mathbf{H}_k^H\mathbf{H}_k(\mathbf{I}_T - (\mathbf{\Upsilon}_k^{(i-1)2}))$ when one diagonal element of $\mathbf{I}_T - (\mathbf{\Upsilon}_k^{(i-1)2})$ is much closer to 0 than the other ones (i.e., $[\mathbf{\Upsilon}_k^{(i-1)2}]_{t,t} >> [\mathbf{\Upsilon}_k^{(i-1)2}]_{t',t'}, t \neq t'$). In that case, we still have the diagonal larger than the other column elements, but it is not necessarily larger than the other line elements. However, that situation occurs when one of the estimates associated with the t-th data block is much better than the estimates for the other blocks, which means that we cancel almost perfectly the interference from the t block, reducing the $T \times R$ massive MIMO system to a $(T-1) \times R$ system after that iteration (the generalization to the case where more than one diagonal element of $\mathbf{I}_T - (\mathbf{\Upsilon}_k^{(i-1)2})$ is much closer to 0 than the other ones is straightforward).

7.2.2 Chebyshev accelerated Richardson iteration

For simplicity of notation, in this section, we shall omit the outer iteration index $'i'$ and write $\mathbf{A}_k^{(i)}$, $\mathbf{b}_k^{(i)}$, $\check{\mathbf{X}}_k^{(i)}$, and $\mathbf{E}_k^{(i)}$ as $\mathbf{A}_k$, $\mathbf{b}_k$, $\check{\mathbf{X}}_k$, and $\mathbf{E}_k$, respectively [2]. The matrix inversion $\mathbf{A}_k^{-1}$ presented in (7.6) affects the computational complexity. Since it was shown earlier that $\mathbf{A}_k$ is a non-singular matrix with larger diagonal elements, we can take advantage of this to reduce the complexity of the matrix inversion. The large diagonal of matrix $\mathbf{A}_k$ inspired us to use the Richardson iterative method [231] to approximate the vector $\check{\mathbf{X}}_k$ in (7.6) without matrix inversion, as this method can solve linear equations with reduced complexity. To further accelerate the Richardson iterative convergence, we incorporate the Chebyshev acceleration approach [95]. This technique has enhanced the convergence rate of the Richardson iterative method [97]. Then, the Richardson iteration can be described as

$$\check{\mathbf{X}}_k^{(\ell)} = \check{\mathbf{X}}_k^{(\ell-1)} + \omega\left(\mathbf{b}_k - \mathbf{A}_k\check{\mathbf{X}}_k^{(\ell-1)}\right), \tag{7.7}$$

where ℓ is the number of inner iterations of the Richardson iteration and $\mathbf{b}_k$ is defined in (7.6). For the first iteration, where there is no prior information about the initial solution, $\check{\mathbf{X}}_k^{(0)}$ can be assumed to be zero. In (7.7), ω denotes the relaxation parameter, which plays a significant role in the convergence of the Richardson method. The performance and complexity of the method are contingent on the value of ω [199]. Without loss of generality, we can rewrite (7.7) as

$$\check{\mathbf{X}}_k^{(\ell)} = \mathbf{E}_k\check{\mathbf{X}}_k^{(\ell-1)} + \omega\mathbf{b}_k, \tag{7.8}$$

where $\mathbf{E}_k$ is the iteration matrix of Richardson's method, given by

$$\mathbf{E}_k = \mathbf{I}_T - \omega\mathbf{A}_k. \tag{7.9}$$

Thus, the necessary and sufficient conditions for the convergence of (7.7) are that the spectral radius should satisfy [111], $\rho(\mathbf{E}_k) < 1$, which leads to $\lim_{\ell\to\infty}\mathbf{E}_k^{(\ell)} = \mathbf{0}_{T\times T}$. In [199], it was shown that this means that the relaxation parameter ω should be in the range.

$$0 < \omega < 2/\lambda_{max}(\mathbf{A}_k). \tag{7.10}$$

However, the actual choice of the relaxation parameter affects the convergence rate, making it an essential issue of the Richardson technique [94]. Since the matrix $\mathbf{A}_k$ is non-singular, when the number of R and T is large enough while T / R is kept fixed, the smallest and largest eigenvalues of $\mathbf{A}_k$ can be well approximated by [30]

$$\lambda_{max}(\mathbf{A}_k) = R\left(1 + \sqrt{\xi}\right)^2, \tag{7.11}$$

[2] It should be noted that $\mathbf{A}_k$, $\mathbf{b}_k$, and $\mathbf{E}_k$ change with the IB-DFE iterations. However, since this section focuses on the number of iterations of the RI and RI-Chebyshev methods (ℓ), we will not consider the number of iterations in the IB-DFE method (denoted by "i").

and

$$\lambda_{min}(\mathbf{A}_k) = R\left(1-\sqrt{\xi}\right)^2, \tag{7.12}$$

respectively, where $\xi = T/R$ is the ratio between the users and the number of antennas. In addition, [94] gives us the optimum value for the relaxation parameter.

$$\omega_{opt} = \frac{2}{\lambda_{max}(\mathbf{A}_k) + \lambda_{min}(\mathbf{A}_k)}. \tag{7.13}$$

By replacing (7.11) and (7.12) in (7.13), the optimum ω_{opt} can be expressed as

$$\omega_{opt} = \frac{1}{R+T}. \tag{7.14}$$

The spectral radius of $\mathbf{E}_k$ is denoted by $\rho(\mathbf{E}_k) = \max_{1\le i\le T} |\lambda_i|$, where λ_i denotes the i-th eigenvalue of $\mathbf{E}_k$. The value of $\rho(\mathbf{E}_k)$ can be approximated as [94]

$$\rho(\mathbf{E}_k) = 1 - \omega_{opt}\lambda_{max}(\mathbf{A}_k) = 1 - \frac{R}{R+T}\left(1-\sqrt{\xi}\right)^2. \tag{7.15}$$

Therefore, the spectral radius can be estimated from the number of antennas and users [30].

Another crucial aspect concerning the Richardson method is the initial solution. This work uses an initial solution to enhance convergence rates, as demonstrated in [125]. It starts by presenting the initial matrix inversion solution as

$$\hat{\mathbf{A}}_k = \psi\mathbf{I}_T + \upsilon\mathbf{A}_k, \tag{7.16}$$

where ψ and υ can be approximated as [125]

$$\psi = \frac{2\xi(1+\xi)}{R(1+\xi^2)}, \upsilon = -\frac{\xi^2}{R^2(1+\xi^2)}. \tag{7.17}$$

Then, the initial solution can be presented as

$$\begin{aligned}\check{\mathbf{X}}_k^{(0)} &= \hat{\mathbf{A}}_k\mathbf{b}_k,\\ &= (\psi\mathbf{I}_T + \upsilon\mathbf{A}_k)\mathbf{b}_k.\end{aligned} \tag{7.18}$$

It has been shown in [92], that this initial solution has shown that it can significantly improve the convergence rate. After the proposed initial solution, the proposed iterative method still required the advantage of speeding up the convergence rate; for that, we utilized the Chebyshev acceleration method with Richardson iterations to attain faster convergence rates. Chebyshev acceleration employs a polynomial to transform the iterative method into a more straightforward structure, making it more adept at solving linear equations. The main benefit of the Chebyshev acceleration method is that it can offer faster convergence than conventional iterative methods, yet maintain the same computing complexity [95]. In general, a Chebyshev acceleration method is designed to accelerate the convergence of $\check{\mathbf{X}}_k^{(\ell)}$. To define the Chebyshev acceleration method, we start by generating the approximation solution of $\check{\mathbf{X}}_k^{(\ell)}$ in (7.8) as a combination of previous estimates.

$$\check{\mathbf{X}}_k^{(\ell)} = \sum_{\iota=0}^{\ell} a_{\ell,\iota} \check{\mathbf{X}}_k^{(\iota)}, \tag{7.19}$$

where the parameters $a_{\ell,\iota}$ can be referred to within the polynomials Q_ℓ as

$$Q_\ell(\mathbf{E}_k) = \sum_{\iota=0}^{\ell} a_{\ell,\iota} \mathbf{E}_k^{\iota}, \tag{7.20}$$

with $a_{\ell,\iota}$ donating scalars that satisfy $\sum_{\iota=0}^{\ell} a_{\ell,\iota} = 1$ (this condition comes form the assumption that when $\check{\mathbf{X}}_k^{(0)} = \check{\mathbf{X}}_k^{(1)} = \ldots = \check{\mathbf{X}}_k$, it is reasonable to have $\check{\mathbf{X}}_k^{\ell} = \check{\mathbf{X}}_k$). Then, we calculate the error in $\check{\mathbf{X}}_k^{(\ell)}$, which can be written as

$$\begin{aligned}
\check{\mathbf{X}}_k^{(\ell)} - \check{\mathbf{X}}_k &= \sum_{\iota=0}^{\ell} a_{\ell,\iota} \check{\mathbf{X}}_k^{(\iota)} - \check{\mathbf{X}}_k, \\
&= \sum_{\iota=0}^{\ell} a_{\ell,\iota} \mathbf{E}_k^{\iota} (\check{\mathbf{X}}_k^{(0)} - \check{\mathbf{X}}_k), \\
&= Q_\ell(\mathbf{E}_k)(\check{\mathbf{X}}_k^{(0)} - \check{\mathbf{X}}_k).
\end{aligned} \tag{7.21}$$

Consequently, the norm of error can be defined as

$$||\check{\mathbf{X}}_k^{(\ell)} - \check{\mathbf{X}}_k|| = ||Q_\ell(\mathbf{E}_k)|| \cdot ||(\check{\mathbf{X}}_k^{(0)} - \check{\mathbf{X}}_k)||. \tag{7.22}$$

Therefore, a polynomial of a matrix $\mathbf{E}_k$ can be minimized by minimizing

$$||Q_\ell(\mathbf{E}_k)|| = \max_{\lambda \in \lambda(\mathbf{E}_k)} |Q_\ell(\lambda)| \leq \max_{-\rho \leq \lambda \leq \rho} |Q_\ell(\lambda)|. \tag{7.23}$$

We need to create a polynomial $Q_\ell(x)$ which is small on $[-\rho, \rho]$ subject to the constraint $Q_\ell(1) = 1$ so that the norm of $Q_\ell(\mathbf{E}_k)$ as small as possible. Since Chebyshev polynomials minimize $Q_\ell(x)$ with $x \in [-\rho; \rho]$ (i.e., they are minimum ∞-norm), they are suitable to meet (7.23) [206]. The Chebyshev polynomials can be defined through the recurrence relation [206]

$$T_{\ell+1}(x) = 2xT_\ell(x) - T_{\ell-1}(x), \text{where } \ell \geq 1, \tag{7.24}$$

with $T_0(x) = 1$ and $T_1(x) = x$. Since we want to minimize $|Q_\ell(x)|$ with $x \in [-\rho; \rho]$, we can define Q_ℓ as in [232]

$$Q_\ell(x) = \frac{T_\ell(x/\rho)}{T_\ell(1/\rho)}, \tag{7.25}$$

where $T_\ell(x)$ is the ℓ-th Chebyshev polynomial and ρ is the spectral radius of the matrix $\mathbf{E}_k$. In [233], it was shown that the optimum $Q_\ell(x)$ can be defined as

$$\begin{aligned} Q_\ell(x) &= \frac{T_\ell(1+2\frac{x-\epsilon}{\epsilon-\eta})}{T_\ell(1+2\frac{r-\epsilon}{\epsilon-\eta})}, \\ \epsilon &= \frac{\lambda_{max}(\mathbf{E}_k)+\lambda_{min}(\mathbf{E}_k)}{\lambda_{max}(\mathbf{E}_k)-\lambda_{min}(\mathbf{E}_k)}, \\ \eta &= \frac{\lambda_{max}(\mathbf{E}_k)+\lambda_{min}(\mathbf{E}_k)}{2}. \end{aligned} \tag{7.26}$$

The ϵ and η denote the optimal parameters of polynomials, and they are defined by the smallest and the largest eigenvalues of the iteration matrix $\mathbf{E}_k$, respectively. In addition, r is any scalar that satisfies $r \notin [\epsilon, \eta]$. In [206], it is shown that $c_\ell = 1/T_\ell(1/\rho)$, which means $Q_\ell(\mathbf{E}_k) = c_\ell T_\ell(\mathbf{E}_k/\rho)$. Then, it can be shown that

$$\frac{1}{c_{\ell+1}} = 2\frac{1}{\rho c_\ell} - \frac{1}{c_{\ell-1}}. \tag{7.27}$$

From (7.21), (7.24) and (7.27), we rewrite $\check{\mathbf{X}}_k^{(\ell)}$ as

$$\begin{aligned} \check{\mathbf{X}}_k^{(\ell)} - \check{\mathbf{X}}_k, &= Q_\ell(\mathbf{E}_k)(\check{\mathbf{X}}_k^{(0)} - \check{\mathbf{X}}_k), \\ &= c_\ell T_\ell\Big(\frac{\mathbf{E}_k}{\rho}\Big)(\check{\mathbf{X}}_k^{(0)} - \check{\mathbf{X}}_k), \\ &= c_\ell\Big[2.\frac{\mathbf{E}_k}{\rho}.T_\ell\Big(\frac{\mathbf{E}_k}{\rho}\Big)(\check{\mathbf{X}}_k^{(0)} - \check{\mathbf{X}}_k) \\ &\quad - T_{\ell-1}\Big(\frac{\mathbf{E}_k}{\rho}\Big)(\check{\mathbf{X}}_k^{(0)} - \check{\mathbf{X}}_k)\Big], \\ &= c_\ell\Big[2.\frac{\mathbf{E}_k}{\rho}.\frac{Q_\ell(\frac{\mathbf{E}_k}{\rho})(\check{\mathbf{X}}_k^{(0)} - \check{\mathbf{X}}_k)}{c_\ell} \\ &\quad - \frac{Q_{\ell-1}(\frac{\mathbf{E}_k}{\rho})(\check{\mathbf{X}}_k^{(0)} - \check{\mathbf{X}}_k)}{c_{\ell-1}}\Big], \\ &= c_\ell\Big[2.\frac{\mathbf{E}_k}{\rho}.\frac{\check{\mathbf{X}}_k^{(\ell)} - \check{\mathbf{X}}_k}{c_\ell} - \frac{\check{\mathbf{X}}_k^{(\ell-1)} - \check{\mathbf{X}}_k}{c_{\ell-1}}\Big], \\ &= 2\frac{c_\ell}{\rho c_{\ell+1}}\mathbf{E}_k\check{\mathbf{X}}_k^{(\ell)} - \frac{c_{\ell-1}}{c_{\ell+1}}\check{\mathbf{X}}_k^{(\ell-1)} + \mathbf{d}_\ell, \end{aligned} \tag{7.28}$$

where

$$\mathbf{d}_\ell = 2\frac{c_\ell}{\rho c_{\ell+1}}\omega\mathbf{b}_k. \tag{7.29}$$

(see the details in [96]).

After some mathematical manipulations, the final equation for Chebyshev acceleration of the iterative method can be written as

$$\check{\mathbf{X}}_k^{(\ell+1)} = 2\frac{c_\ell}{\rho c_{\ell+1}}\mathbf{E}_k\check{\mathbf{X}}_k^{(\ell)} - \frac{c_{\ell-1}}{c_{\ell+1}}\check{\mathbf{X}}_k^{(\ell-1)} + 2\frac{c_\ell}{\rho c_{\ell+1}}\omega\mathbf{b}_k, \tag{7.30}$$

where $\rho = \rho(\mathbf{E}_k)$, $\check{\mathbf{X}}_k^{(0)} = \check{\mathbf{X}}_k^{(0)}$ and $\check{\mathbf{X}}_k^{(1)} = \mathbf{E}_k\check{\mathbf{X}}_k^{(0)} + \omega\mathbf{b}_k$.

We present Algorithm 4, which combines the IB-DFE method with Chebyshev-accelerated RI and is referred to as the iterative block decision feedback equalizer. Additionally, we present Algorithm 5, which employs the Richardson-Chebyshev method.

Algorithm 4 **Proposed iterative block decision-feedback equalizer (IB-DFE-RI-Cheby)**

1: **Input:** $\mathbf{Y}_k, \mathbf{H}_k, \sigma^2_{W_k}, \sigma^2_{X_k}$.
2: **Output:** Estimated signal $\tilde{\mathbf{X}}_k^{(I)}$.
3: **Initialization phase:**
4: $\mathbf{Y}_k^{MF} = \mathbf{H}_k^H\mathbf{Y}_k$.
5: $\mathbf{G}_k = \mathbf{H}_k^H\mathbf{H}_k$.
6: $\hat{\mathbf{X}}_k^{(0)} = \mathbf{0}_{T\times 1}$; $\boldsymbol{\Upsilon}_k^{(0)} = \mathbf{0}_{T\times T}$.
7: **Iteration phase:**
8: **for** $i = 1 : I$ **do**
9: $\quad \boldsymbol{\Gamma}_k^{(i)} = diag\left(\gamma_{k,1}^{(i)}, \gamma_{k,2}^{(i)}, \ldots\gamma_{k,T}^{(i)}\right)$.
10: $\quad \mathbf{b}_k^{(i)} = \mathbf{Y}_k^{MF} - \mathbf{G}_k\hat{\mathbf{X}}_k^{(i-1)}$.
11: $\quad \mathbf{A}_k^{(i)} = \frac{\sigma^2_{W_k}}{\sigma^2_{X_k}}\mathbf{I}_T + \mathbf{G}_k\left(\mathbf{I}_T - \boldsymbol{\Upsilon}_k^{(i-1)2}\right)$.
12: $\quad \check{\mathbf{X}}_k \leftarrow \textit{RI-Cheby}\left(\mathbf{A}_k^{(i)}, \mathbf{b}_k^{(i)}\right)$. $\qquad \triangleright\ \check{\mathbf{X}}_k = (\mathbf{A}_k^{(i)})^{-1}\mathbf{b}_k^{(i)}$.
13: $\quad \tilde{\mathbf{X}}_k^{(i)} = \boldsymbol{\Gamma}_k^{(i)}\check{\mathbf{X}}_k + \hat{\mathbf{X}}_k^{(i-1)}$.
14: $\quad \boldsymbol{\Upsilon}_k^{(i-1)} = \frac{E[\hat{\mathbf{x}}_k^{(i-1)}\mathbf{x}_k^H]}{E[|\mathbf{x}_k|^2]} = \frac{\sum_{\nu=0}^{M-1}|g_\nu|^2\prod_{l=0}^{\mu-1}|\hat{b}_{n,t}^{(l)}|^{\Omega_{l,\nu}}}{\sum_{\nu=0}^{M-1}|g_\nu|^2}$.
15: **end for**
16: **Return** $\tilde{\mathbf{X}}_k^{(I)}$.

The inputs of Algorithm 1 are the received vector $\mathbf{Y}_k$, the channel matrix $\mathbf{H}_k$, the noise variance $\sigma^2_{W_k}$, and the signal variance $\sigma^2_{X_k}$. The initialization phase defines the one-time computations for Algorithm 1. Line 4 computes the matched filter response $\mathbf{Y}_k^{MF}$. The Gram matrix is computed at line 5. The vector $\hat{\mathbf{X}}_k^{(0)}$, and the matrix $\boldsymbol{\Upsilon}_k^{(0)}$ are initialized at line 6. Next, the iteration phase of the iterative block decision-feedback equalization starts. This phase contains two parts: the first part calculates $\boldsymbol{\Gamma}_k^{(i)}$, $\mathbf{b}_k^{(i)}$, and $\mathbf{A}_k^{(i)}$, in lines 9, 10, and 11, respectively. Line 12 uses Algorithm 2, which takes the matrix $\mathbf{A}_k^{(i)}$ and the vector $\mathbf{b}_k^{(i)}$ as input. The output of the algorithm is the estimated solution $\check{\mathbf{X}}_k$. Line 13 calculates $\tilde{\mathbf{X}}_k^{(i)}$, as presented in (7.5). Line 14 presents the calculation of the average symbol values and correlation coefficients $\boldsymbol{\Upsilon}_k^{(i)}$, as in (3.28). Note that when the value of diagonal elements of $\boldsymbol{\Upsilon}_k^{(i)}$ reaches one, no further performance improvement is possible. Therefore, Algorithm 1 terminates when $\boldsymbol{\Upsilon}_k^{(i)}$ becomes equal to $\mathbf{I}_T$.

Algorithm 5, on the other hand, is used to solve a linear system of equations $\mathbf{A}_k\check{\mathbf{X}}_k = \mathbf{b}_k$ using the RI-Cheby algorithm. It should be noted that the outer iteration index $'i'$ is dropped for simplicity. The initialization of ω_{opt}, the iteration matrix $\mathbf{E}_k$, and the approximation of

Algorithm 5 **Richardson-Chebyshev (RI-Cheby) method**

1: **Input:**$\mathbf{A}_k, \mathbf{b}_k$. ▷ The index i is dropped for simplicity
2: **Output:** Estimated signal $\check{\mathbf{X}}_k$.
3: **Initialization phase:**
4: $\omega_{opt} = \frac{1}{R+T}$.
5: $\mathbf{E}_k = \mathbf{I}_T - \omega_{opt}\mathbf{A}_k$.
6: $\xi = T/R$.
7: $\rho(\mathbf{E}_k) = 1 - \frac{R}{R+T}\left(1 - \sqrt{\xi}\right)^2$.
8: $\psi = 2\xi(1+\xi)/(R(1+\xi^2))$.
9: $\upsilon = -\xi^2/(R^2(1+\xi^2))$.
10: **Phase1:** Initial iteration
11: $\check{\mathbf{X}}_k^{(0)} = (\psi\mathbf{I}_T + \upsilon\mathbf{A}_k)\mathbf{b}_k$.
12: $\check{\mathbf{X}}_k^{(1)} = \mathbf{E}_k\check{\mathbf{X}}_k^{(0)} + \omega_{opt}\mathbf{b}_k$.
13: $c_0 = 1; c_1 = \frac{1}{\rho(\mathbf{E}_k)}$.
14: **Phase2:** Chebyshev accelerated iteration
15: **for** $\ell = 1 : L-1$ **do**
16: $\quad \frac{1}{c_{\ell+1}} = 2\frac{1}{\rho c_\ell} - \frac{1}{c_{\ell-1}}$. ▷ use recursion for c_ℓ instead of
17: $\quad$ the definition $c_\ell = 1/T_\ell(1/\rho)$.
18: $\quad \check{\mathbf{X}}_k^{(\ell+1)} = 2\frac{c_\ell}{\rho c_{\ell+1}}\mathbf{E}_k\check{\mathbf{X}}_k^{(\ell)} - \frac{c_{\ell-1}}{c_{\ell+1}}\check{\mathbf{X}}_k^{(\ell-1)} + \frac{2c_\ell}{\rho c_{\ell+1}}\omega_{opt}\mathbf{b}_k$.
19: **end for**
20: **Return** $\check{\mathbf{X}}_k = \check{\mathbf{X}}_k^{(L)}$.

the spectral radius (defined by $\rho(\mathbf{E}_k)$) are presented in lines 4, 5, and 7. The algorithm is executed in two phases. In the first phase, the initial solution is generated using Newton's method, as shown in line 11 (see (7.18)). Next, the solution for the first iteration is computed using the conventional Richardson method in line 12. In line 13, the initial parameters for the Chebyshev acceleration method are computed. The remaining $L-1$ iterations of Algorithm 2 are performed in Phase 2. The coefficient of Chebyshev recursion is computed in line 16, using (7.27). After computing the Chebyshev polynomial coefficients, line 18 computes the Richardson-Chebyshev iteration (defined in (7.30)).

7.3 Complexity analysis

This section presents the complexity analysis of the proposed receiver. Additionally, we compare the complexity of various detectors based on the number of floating-point operations, defined as FLOPs. In the case of number of complex multiplications, four real multiplications and two real additions are required, corresponding to the equivalent of six FLOPs. Similarly, for complex addition and square-root operation, the number of required FLOPs are two and nine, respectively [234]. In Algorithm 1, we need to calculate the Gram matrix $\boldsymbol{G}_k = \boldsymbol{H}_k^H\boldsymbol{H}_k$ and the matched filter output $\boldsymbol{Y}_k^{MF} = \boldsymbol{H}_k^H\boldsymbol{Y}_k$ in the initialization step. In this context, the number of FLOPs required for the Gram matrix and the matched filter is $8T^2R - 2T^2$ and $8TR - 2T$, respectively. Next, in the iteration phase, the computation of the matrix $\boldsymbol{A}_k = \frac{\sigma^2_{W_k}}{\sigma^2_{X_k}}\boldsymbol{I}_T + \boldsymbol{G}_k(\boldsymbol{I}_T - \boldsymbol{\Upsilon}_k^{(i-1)2})$ requires T^2 complex-to-real multiplication and $2T$ real additions. Thus, the total required FLOPs for $\boldsymbol{A}_k$ matrix computation is $2T^2 + 2T$. The term $\boldsymbol{b}_k = \boldsymbol{Y}_k^{MF} - \boldsymbol{G}_k\boldsymbol{X}_k^{(i-1)}$ requires one matrix-vector multiplication and one vector addition. Thus, $8T^2$ FLOPs is required for the computation of $\boldsymbol{b}_k$.

Next, we need to determine the complexity of the proposed Chebyshev accelerated Richardson iteration, presented in Algorithm 2. In the initialization phase, the calculation of $\hat{\boldsymbol{A}}_k = \psi\boldsymbol{I}_T + \upsilon\boldsymbol{A}_k$ requires T^2 number of complex-to-real multiplications and T number of complex additions. Thus, a total of $2T^2 + 2T$ FLOPs are required to compute $\hat{\boldsymbol{A}}_k$. After generation of $\hat{\boldsymbol{A}}_k$, computation of the initial solution vector $\breve{\boldsymbol{X}}_k^{(0)} = \hat{\boldsymbol{A}}_k\boldsymbol{b}_k$ involves matrix-vector multiplication, which requires $8T^2 - 2T$ FLOPs. The first iteration solution vector $\breve{\boldsymbol{X}}_k^{(1)} = (\boldsymbol{I}_T - \omega_{opt}\boldsymbol{A}_k)\breve{\boldsymbol{X}}_k^{(0)} + \omega_{opt}\boldsymbol{b}_k$ requires total $10T^2$ FLOPs. The rest of the $(L-1)$ iterations are the Chebyshev accelerated Richardson iteration. The ℓ-th iteration (defined in Algorithm 2, line 18) requires $10T^2 + 4T$ FLOPs.

Table 7.1 presents the number of FLOPs required and the corresponding order of complexity for different IB-DFE, where L is the number of inner iterations required for the iterative solution and i indicates the number of outer iterations for IB-DFE. Through this paper, the conventional IB-DFE detector is abbreviated as conv.IB-DFE. On the other hand, the IB-DFE detector with the proposed RI-Cheby iterative method is named IB-DFE-RI-Cheby. Similarly, the other IB-DFE detectors, such as IB-DFE-RI-NS, IB-DFE-AOR, IB-DFE-GS, and IB-DFE-CG, are the alternative version of Algorithm 1 where line 12 is computed using the methods RI-NS [92], AOR [88], GS [112], CG [90], respectively.

Table 7.1 Computational complexity comparison

Method	FLOPs ($i > 1$)	The complexity Order
Conv.IB-DFE [176]	$8T^2R - 2T^2 + 8TR - 2T + (4T^3 + 32T^2 + 3T)i$	$O(8T^2R + i4T^3)$
IB-DFE-RI-NS	$8T^2R - 2T^2 + 8TR - 2T + (20T^2 + 4T + 2T(5T-1)L)i$	$O(8T^2R + (20T^2 + 10T^2L)i)$
IB-DFE-RI-Cheby	$8T^2R - 2T^2 + 8TR - 2T + (30T^2 - 2T + 4T(2T+1)(L-1))i$	$O(8T^2R + (30T^2 + 8T^2L)i)$
IB-DFE-AOR	$8T^2R - 2T^2 + 8TR - 2T + (10T^2 + 5T + 2T(14T-1)L)i$	$O(8T^2R + (10T^2 + 28T^2L)i)$
IB-DFE-GS	$8T^2R - 2T^2 + 8TR - 2T + (10T^2 + 5T + 2T(8T+2)L)i$	$O(8T^2R + (10T^2 + 16T^2L)i)$
IB-DFE-CG	$8T^2R - 2T^2 + 8TR - 2T + (10T^2 + 5T + 2T(7T-3)L)i$	$O(8T^2R + (10T^2 + 14T^2L)i)$
IB-DFE-MRC [229]	$8T^2R + 2T^2(4i-5) + 12TR + 2T$	$O(8T^2R + i8T^2)$
IB-DFE-ECG [229]	$8T^2R + 2T^2(4i-5) + 19TR + 2T$	$O(8T^2R + i8T^2)$

The complexities presented in Table 7.1 can be separated into the complexities of the initialization and iteration phases. The complexity of initialization (expressed by $O(T^2R)$) is dominated primarily by the calculation of the Gram matrix. As the Gram matrix is computed only once per channel matrix update and is the same for all methods, the complexity of the iteration phase is the key factor in assessing the effective receiver complexity.

In general, the complexity of the iteration phase is proportional to the number of inner and outer iterations. However, it is important to understand that the values of i and L are much lower than the value of T for the detectors presented in this work and, as a result, it is valid to assume $iL << T$. Therefore, we can say that the impact of i and L on receiver complexity is negligible compared to the impact of T. In Table 7.1 we can notice that all methods, such as IB-DFE-RI-Cheby, IB-DFE-RI-NS, IB-DFE-AOR, IB-DFE-GS, IB-DFE-CG, IB-DFE-MRC, and IB-DFE-EGC, reduce the complexity of each IB-DFE iteration by a factor of T compared to the conventional IB-DFE receiver (that is, from $\mathcal{O}(T^3)$ to $\mathcal{O}(T^2)$).

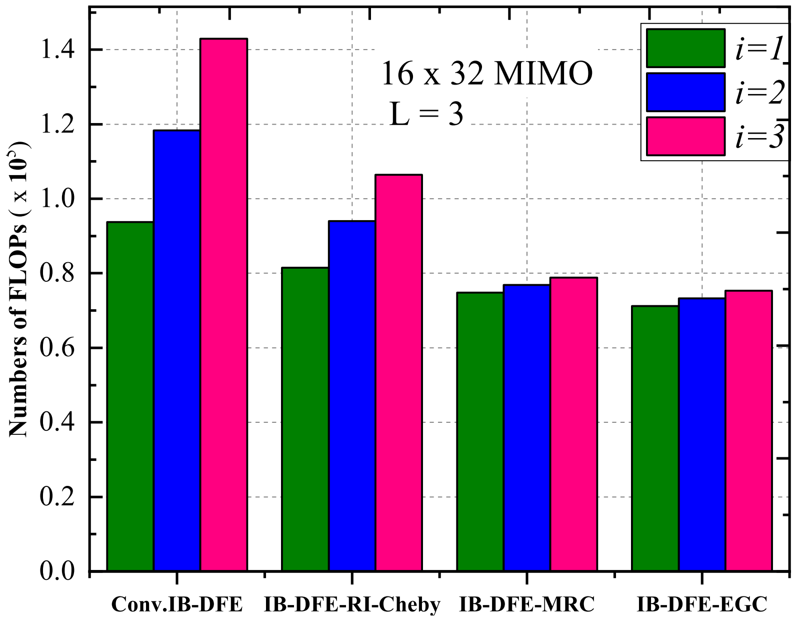


Figure 7.2: Number of FLOPs for 16×32 MIMO system and $L = 3$.

To better understand, Figs. 7.2 and 7.3 illustrate the number of FLOPs necessary for 16×32 and 64×256 MIMO systems, respectively. In both cases, the value of inner iterations for the proposed method is set to three. Clearly, the proposed method reduces the complexity of the IB-DFE receiver in all system configurations compared to the conventional method. However, the number of FLOPs required for IB-DFE-MRC and IB-DFE-EGC is still lower in all cases.

7.4 Performance analysis

This section evaluates the performances of the proposed IB-DFE receivers. Although this work is focused on the proposed RI-Cheby method, other iterative methods, such as RI-NS [92], GS [112], CG [90], AOR [88], etc., are also adapted to the IB-DFE receiver. We also compare them with the conventional IB-DFE receiver that requires matrix inversions (denoted as 'Conv. IB-DFE'), as well as with EGC and MRC techniques from [180] and

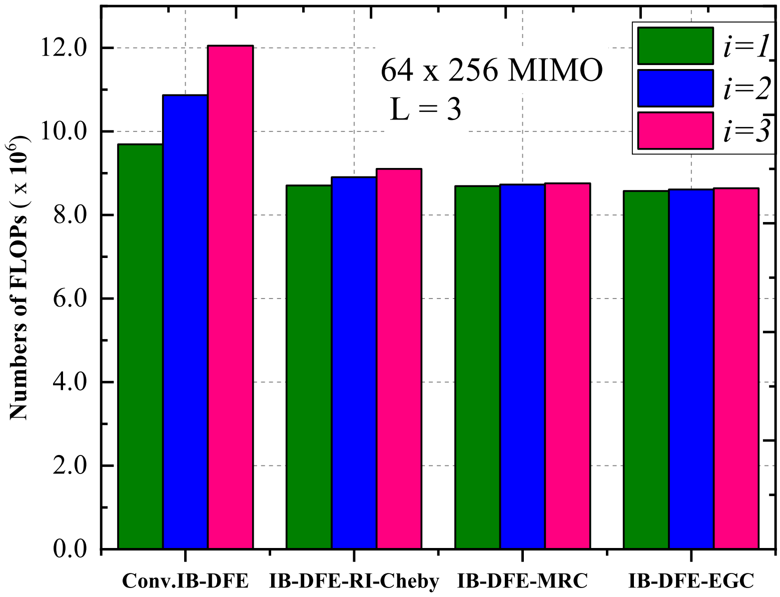


Figure 7.3: Number of FLOPs for 64×256 MIMO system and $L = 3$.

[179] respectively. In all cases, 'i' denotes the number of outer IB-DFE iterations, and 'L' denotes the number of inner iterations for iterative methods such as RI-Cheby, GS, CG, AOR, etc. We consider data symbol blocks of size $N = 256$ and a suitable cyclic prefix has been added. The channel between each transmitting and receiving antenna has 16 equal power, symbol-spaced multipath components, where each component is uncorrelated and follows Rayleigh distribution. At the receiver, we employ different IB-DFE receivers with soft decisions within the feedback loop. The message bits are mapped to M-QAM symbols using (3.6) and (3.7). Furthermore, the bitwise LLR values are generated using (3.24). Throughout this section, the antenna configuration is denoted by $T \times R$.

7.4.1 Performance under perfect channel state information

First, we consider the perfect channel state information available at the receiver to evaluate the bit error rate (BER) performance against the variation of E_b/N_0, where N_0 is the one-sided power spectral density of the noise and E_b is the energy of the transmitted bits. We consider various massive MIMO antenna configurations for simulation, including 16×32, 16×128, 32×128, and 64×256, with different modulation schemes.

Figs. 7.4, 7.5, and 7.6 compare the BER of the different IB-DFE receivers for the 16×32 MIMO system with the QPSK, 16-QAM, and 64-QAM modulation scheme, respectively. From Fig. 7.4 we can observe that although the proposed iterative method exhibits a poor detection performance compared to conventional IB-DFE in the first iteration, the difference decreases as the outer IB-DFE iterations increased. The performance of conventional and proposed receivers becomes similar and close to that of the MFB after three IB-DFE iterations. However, the performance of receivers based on the MRC or EGC is much inferior than that of other receivers.

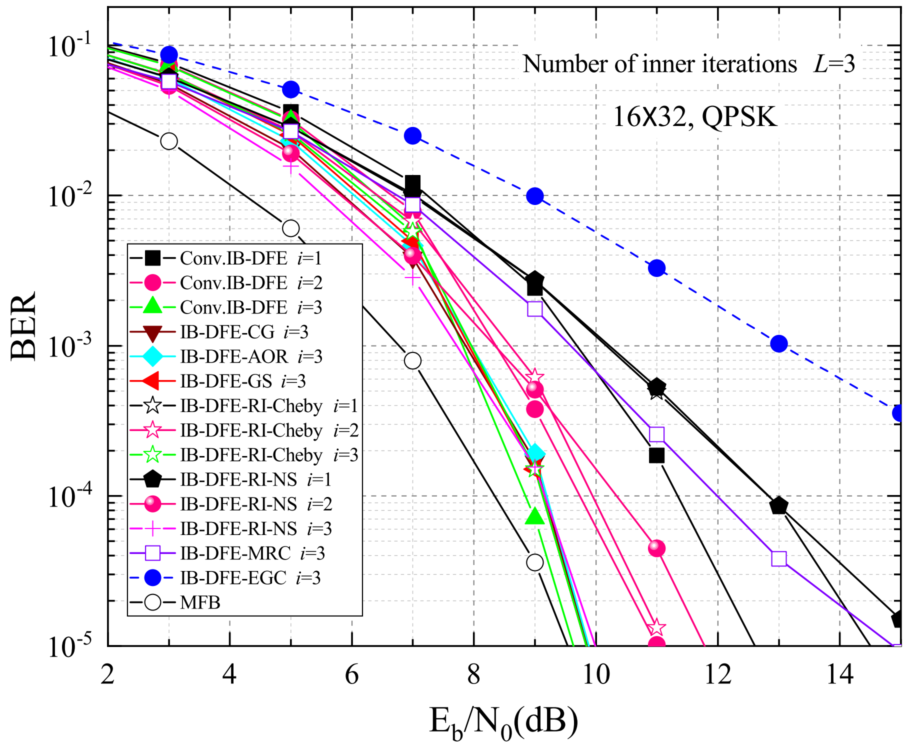


Figure 7.4: BER performance for 16×32 MIMO systems with QPSK modulations.

Next, we evaluated the BER performances of the proposed and conventional receivers for larger constellations, such as 16-QAM and 64-QAM. Figs. 7.5 and 7.6 show the corresponding plots for 16-QAM and 64-QAM constellations, respectively, under the same conditions as Fig. 7.4. The receivers based on EGC or MRC exhibit flat BER characteristics for all signal-to-noise ratios, and hence they are not suitable for larger constellations. Furthermore, the receiver based on the Richardson method also shows a substantial performance degradation at higher levels of E_b/N_0. However, when the Richardson method is combined with Chebyshev acceleration, the resulting IB-DFE receiver shows a significant improvement in detection performance compared to its conventional counterpart. This means that the proposed Chebyshev-accelerated Richardson method, presented in Algorithm 5, has a faster convergence rate compared to other proposed methods, such as GS, CG, and AOR. In summary, the proposed receiver and the conventional IB-DFE receiver have substantial BER improvements as we increase the number of IB-DFE iterations. Furthermore, for the same number of inner iterations, the proposed IB-DFE-RI-Cheby performs better than IB-DFE-RI-NS in all cases.

Figs. 7.7, 7.8, and 7.9 show the BER performance for antenna configurations: 16×128, 32×128, and 64×256, respectively, with the 64-QAM modulation scheme. For simplicity, IB-DFE-MRC and IB-DFE-EGC receivers are not presented in these plots, since their

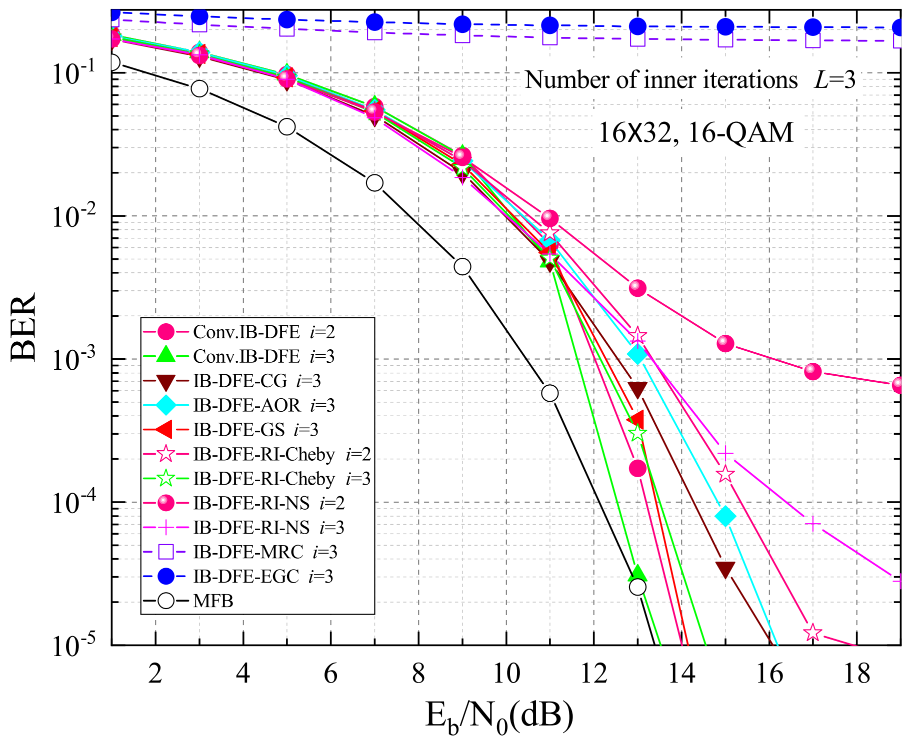


Figure 7.5: BER performance for 16×32 MIMO systems with 16-QAM modulations.

performance is very poor, with high irreducible error floors. In Fig. 7.7, it is observed that the performance of the IB-DFE-RI-Cheby, IB-DFE-RI-NS, and IB-DFE-GS receivers almost matches that of the conventional IB-DFE receiver in just two inner iterations. However, IB-DFE-AOR and IB-DFE-CG exhibit poor performance. One way to improve the performance of IB-DFE-AOR and IB-DFE-CG is to increase the number of inner iterations, but it increases the complexity. It is worth mentioning that the value of R/T plays a critical role in massive MIMO detection. We have already examined the cases of $R/T = 2$ and $R/T = 8$ in previous plots. Therefore, in Figs. 7.8 and 7.9 we consider the value of R/T as 4 and both plots consider the 64-QAM modulation scheme, four inner iterations, and three outer iterations.

Clearly, the proposed IB-DFE-RI-Cheby and IB-DFE-GS receivers can achieve the performance of the conventional IB-DFE receiver. However, the performance of the IB-DFE-RI-NS, IB-DFE-AOR, and IB-DFE-CG receivers is poorer than that of the conventional IB-DFE receiver. From the previous BER plots, we can say that the performance of IB-DFE receivers depends on several factors, such as: R/T ratio, modulation order, number of inner iterations, number of outer iterations, etc.

As discussed earlier, the proposed IB-DFE structure can reduce the outer iteration complexity from $O(T^3)$ to $O(T^2)$. However, the individual advantages of these methods vary in terms of simplicity and the number of iterations required to achieve the target performance. Although the performance of the GS method is good, its execution is

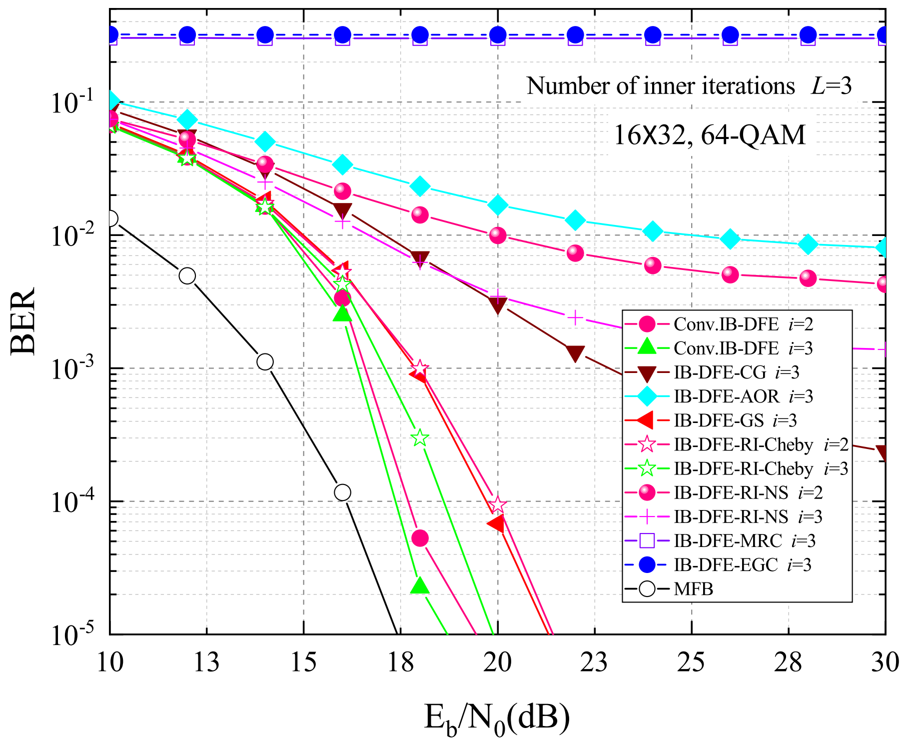


Figure 7.6: BER performance for 16×32 MIMO systems with 64-QAM modulations.

sequential in nature. This may limit the speed of detection of GS, making it less suitable for practical implementation than the proposed RI-Cheby method.

Fig. 7.10 shows the impact of the number of IB-DFE iterations for different constellations when $L = 3$ (the E_b/N_0 is selected to have a target BER around 10^{-5}). As expected, when the number of outer iterations increases, the performance of all IB-DFE receivers improves. The proposed IB-DFE-RI-Cheby receiver converges faster, requiring only three iterations for smaller constellations (such as QPSK, and 16-QAM) and four for 64-QAM. In contrast, the other techniques can require a much larger number of iterations (it should be noted that the exact requirement of the number of outer and inner iterations also depends on the target BER and E_b/N_0).

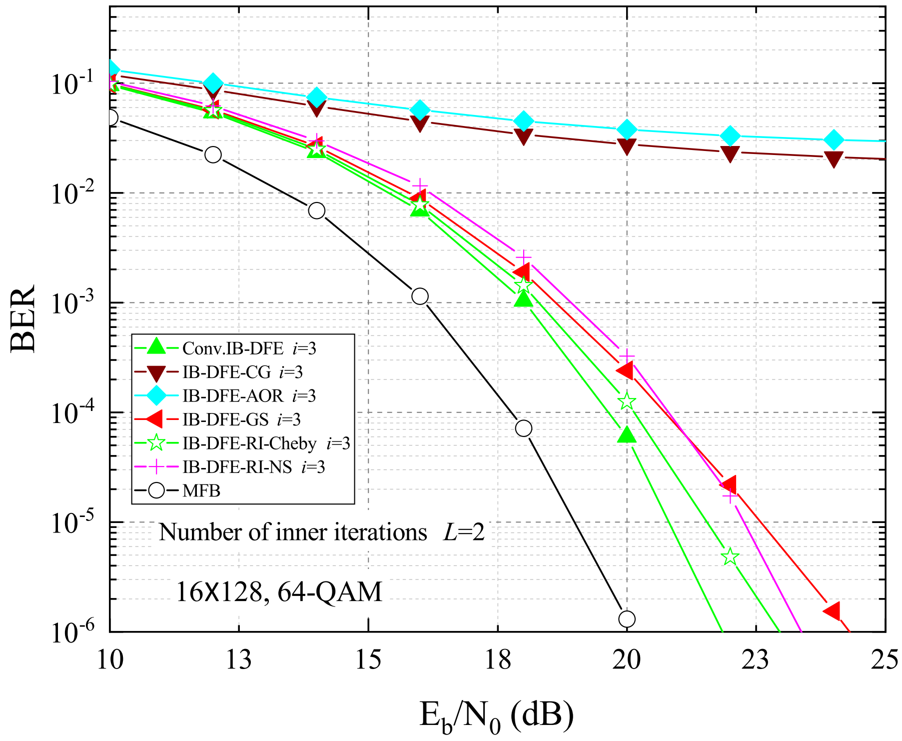


Figure 7.7: BER performance for 16×128 MIMO systems with 64-QAM modulations.

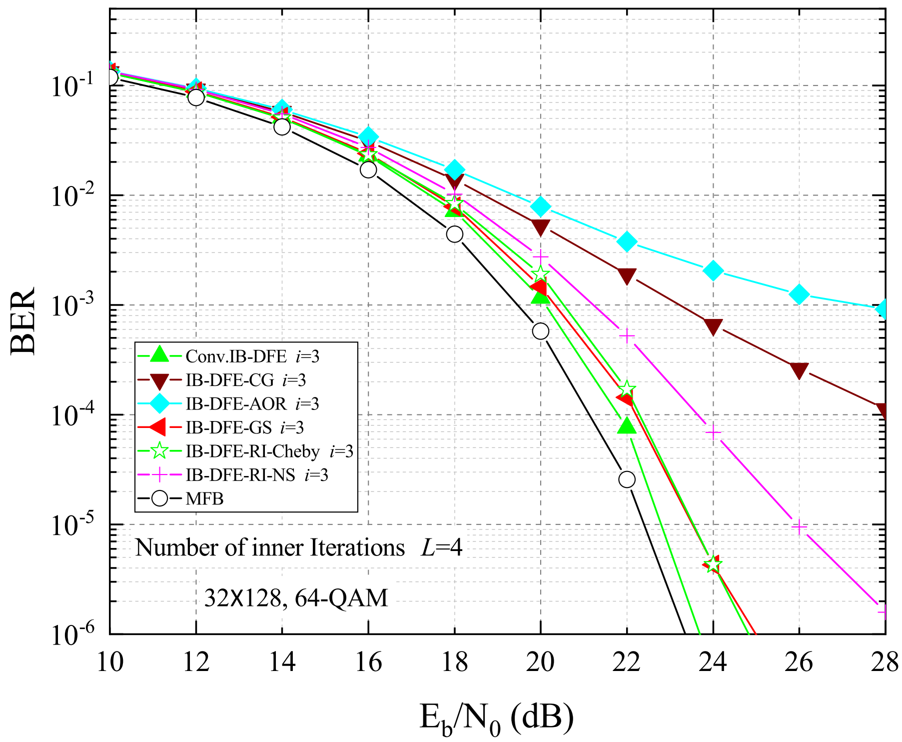


Figure 7.8: BER performance for 32×128 MIMO systems with 64-QAM modulations.

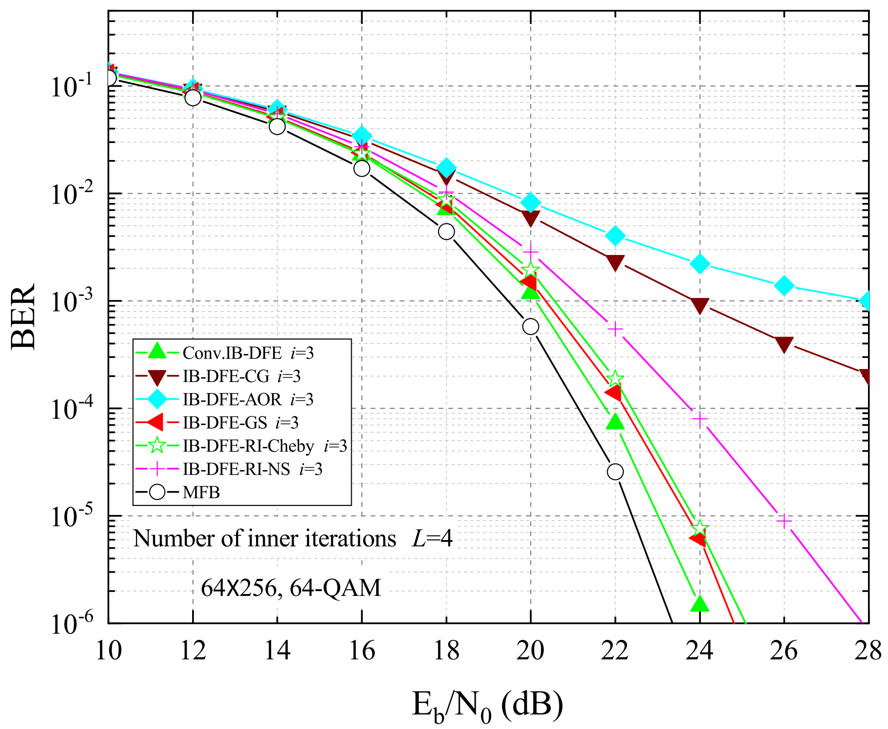


Figure 7.9: BER performance for 64 × 256 MIMO systems with 64-QAM modulations.

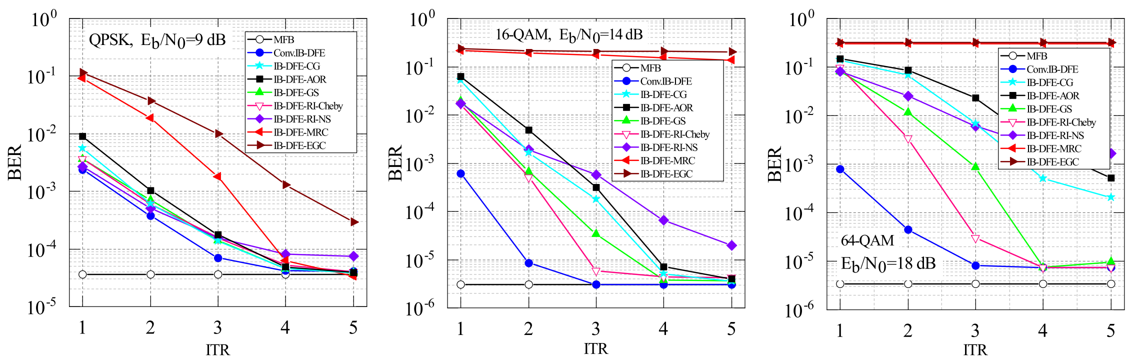


Figure 7.10: Impact of the number of IB-DFE iterations on the BER for 16 × 32 MIMO systems with different constellations and L = 3.

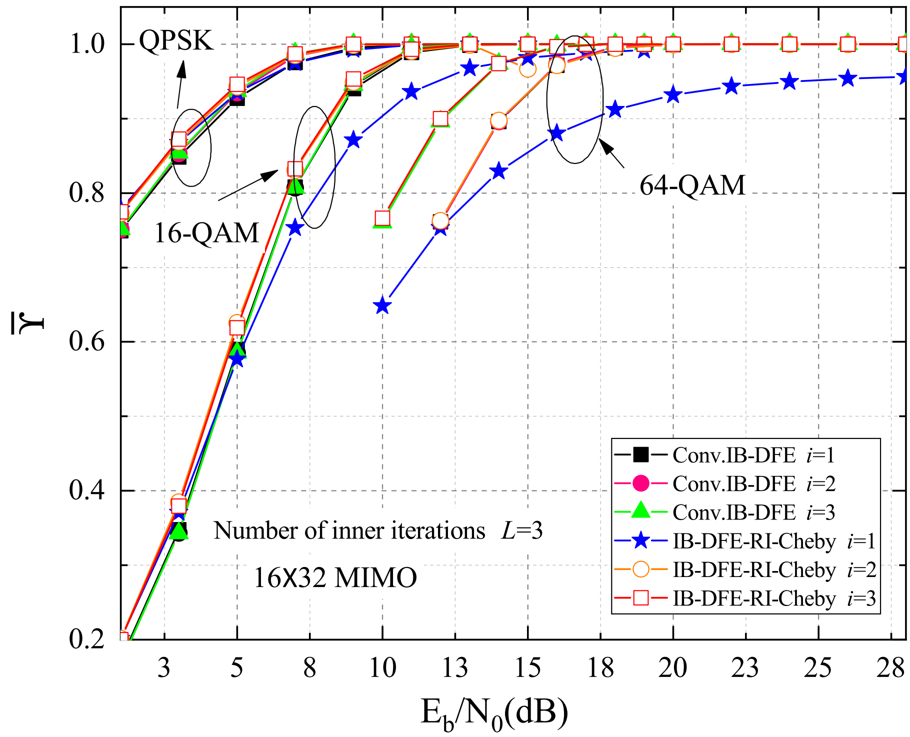


Figure 7.11: Evolution of $\bar{\Upsilon}$ for 16×32 MIMO systems with different constellations and $L = 3$.

As defined in (3.19), $\boldsymbol{\Upsilon}_k^{(i)}$ represents the reliability of the estimates used in the feedback loop, and $\boldsymbol{\Upsilon}_k^{(i)} \approx \mathbf{I}_T$ indicates that the receiver has reached its optimal solution and no further improvement can be done. Therefore, it is important to understand the characteristics of the expected value of $\boldsymbol{\Upsilon}_k^{(i)}$, represented by $\bar{\Upsilon} = \frac{1}{NT}\sum_{k=1}^{N}\sum_{t=1}^{T}[\boldsymbol{\Upsilon}_k^{(i)}]_{t,t}$ with a variation of E_b/N_0. Figs. 7.11 and 7.12, show the variation in the expected correlation value (defined by $\bar{\Upsilon}$) for different IB-DFE receivers with variations of E_b/N_0. The antenna configuration for Figs. 7.11 and 7.12 is set to 16×32 and 64×256, respectively. As expected, for both antenna configurations with QPSK constellation, $\bar{\Upsilon}$ reaches 1 (corresponding to no errors in the feedback estimates) at a much lower SNR, and the proposed receiver has a similar behavior (because for QPSK constellations, we have $\bar{\Upsilon} \approx 1 - 2 \times BER$, differences at low BER values have a minor impact on $\bar{\Upsilon}$ [174]). For 16-QAM modulation, as depicted in Fig. 7.11, there are some differences between the proposed and conventional receivers, especially during the first IB-DFE iteration. In some cases, as shown in Fig. 7.12, the disparity between the proposed and conventional receivers is noticeable in the initial iterations.

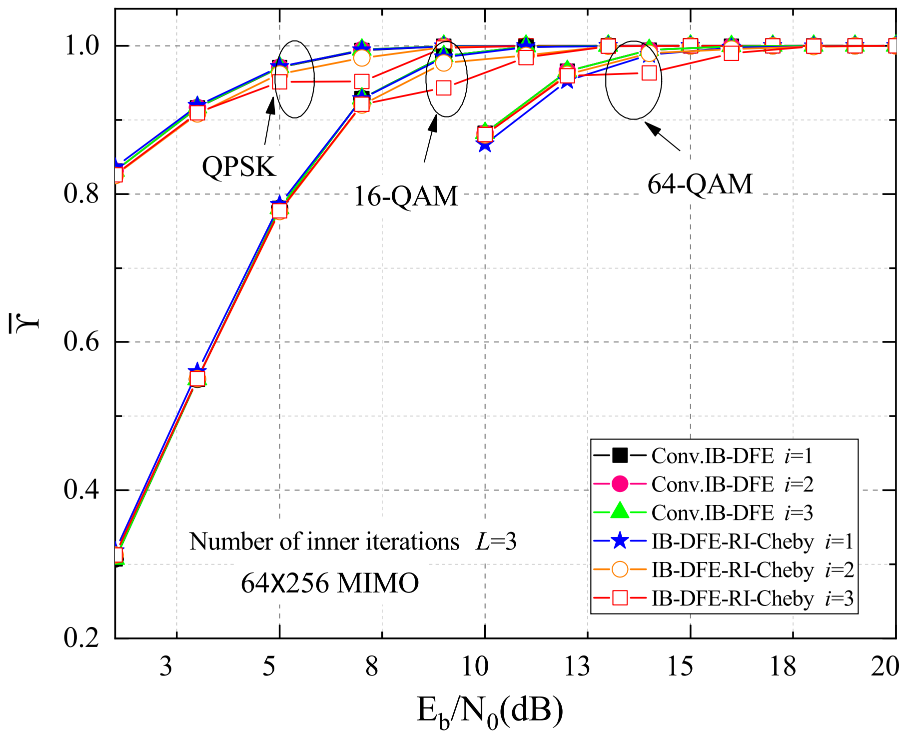


Figure 7.12: Evolution of $\bar{\Upsilon}$ for 64×256 MIMO systems with different constellations and $L = 3$.

For 64-QAM modulation, we can notice from Fig. 7.11 that the differences between the proposed and conventional receivers are much higher for the first IB-DFE iterations, where $\bar{\Upsilon}$ of IB-DFE-RI-Cheby unable to reach 1 in the first iteration. However, with three iterations, the proposed receiver becomes similar to the conventional IB-DFE receiver, indicating excellent convergence of the proposed receiver. Although the $\bar{\Upsilon}$ values of the conventional and proposed receivers look identical, a minor deviation can be observed for the proposed method at $i = 3$. The reason is that we add a termination condition to the proposed IB-DFE method. When the value of $\bar{\Upsilon}$ reaches closer to one, the algorithm terminates and skips the remaining outer iterations and retains the previous iteration $\bar{\Upsilon}$ value.

Table 7.2 displays the required E_b/N_0 to obtain the BER of 10^{-3}, as well as the number of FLOPs for different IB-DFE receivers (IB-DFE-MRC and IB-DFE-EGC receivers are not presented, as they cannot achieve this BER). For simulation, we have considered 16×32 and 64×256 MIMO configurations, with 64-QAM constellations. As expected, conventional IB-DFE requires the lowest E_b/N_0 and the highest number of FLOPs. The proposed IB-DFE-RI-Cheby and IB-DFE-GS receivers offer the best trade-off between complexity and performance. Table 7.3 shows the the number of FLOPS for each outer iteration (after the first) on a $T \times R$ MIMO system with 64-QAM constellations. It can be seen that the complexity per iteration of our proposed IB-DFE-RI-Cheby decreases by a factor around 2 to 5 when compared to the conventional receiver, with larger gains for

Table 7.2 Number of FLOPs and E_b/N_0 required to achieve a BER of 10^{-3}.

Method	16×32 ($i = L = 3$)		64×256 ($i = 3, L = 4$)	
	FLOPs ($\times 10^5$)	E_b/N_0 (dB)	FLOPs ($\times 10^6$)	E_b/N_0 (dB)
Conv.IB-DFE [176]	1.43	17.20	13.23	20.32
IB-DFE-CG	1.08	23.03	9.36	23.91
IB-DFE-AOR	1.41	40.38	10.05	27.98
IB-DFE-GS	1.14	17.96	9.47	20.75
IB-DFE-RI-Cheby	1.05	17.64	9.27	21.06
IB-DFE-RI	1.07	44.23	9.33	21.60

Table 7.3 Number of additional FLOPs for each outer iteration.

Method	16×32 ($L = 3$)	64×256 ($L = 4$)
	FLOPs($\times 10^4$)	FLOPs($\times 10^5$)
Conv.IB-DFE [176]	2.46	11.8
IB-DFE-CG	1.31	2.69
IB-DFE-AOR	2.40	4.99
IB-DFE-GS	1.51	3.04
IB-DFE-RI-Cheby	1.19	2.22
IB-DFE-RI	1.28	2.45

larger MIMO systems.

The impact of different numbers of antennas at the base station on the BER performance is shown in Fig. 7.13. For simulation, we consider the following parameters: $T = 32$, $E_b/N_0 = 18$, and $L = 4$, and $i = 3$. From Fig. 7.13, we can see that the conventional IB-DFE receiver has a stable performance with the change in the number of receiving antennas R. However, the other receivers exhibit poor performance if we reduce R too much. As we increase the value of R, the performance gap between the conventional IB-DFE and other IB-DFE receivers decreases. Furthermore, it can also be noticed that among the IB-DFE receivers except conventional IB-DFE, the performance of IB-DFE-GS and IB-DFE-RI-Cheby is superior.

7.4.2 Performance under imperfect channel state information

This section studies the performance of the proposed and conventional IB-DFE receivers under imperfect channel state information (CSI). A simple way of modelling the noisy (and/or aged) channel estimates is to write them as

$$\hat{\mathbf{H}}_k = \beta \mathbf{H}_k + \sqrt{1 - \beta^2} \mathbf{E}_k, \tag{7.31}$$

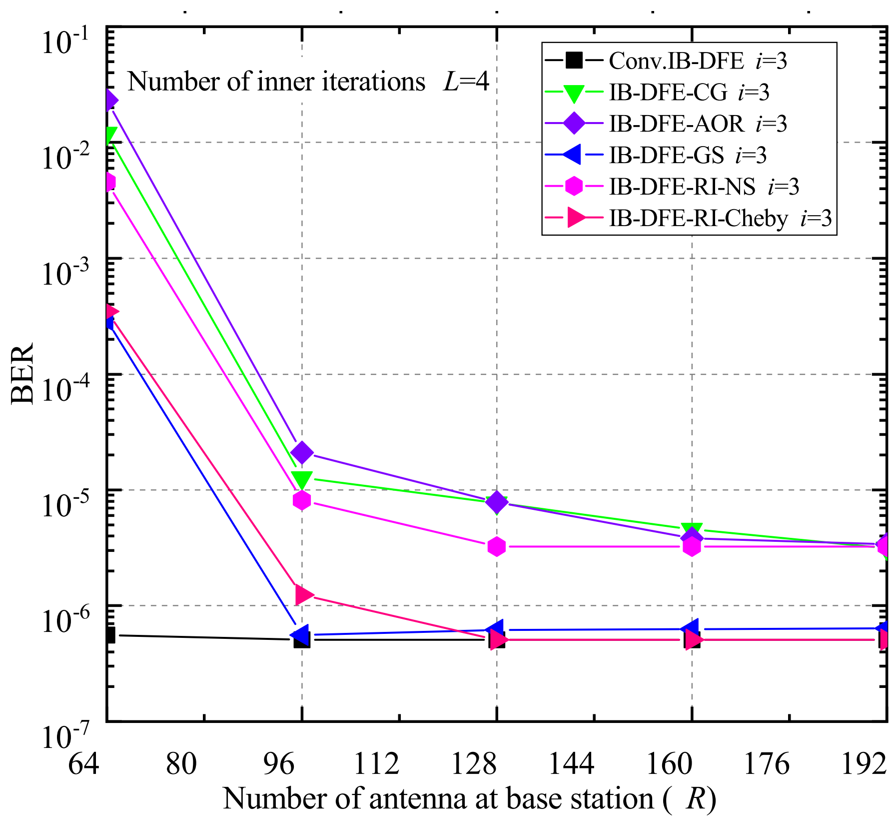


Figure 7.13: BER performance of the proposed method versus the number of antenna at the base station is evaluated under $T = 32$ and $E_b/N_0 = 18$ dB with 64-QAM modulation and $L = 4$.

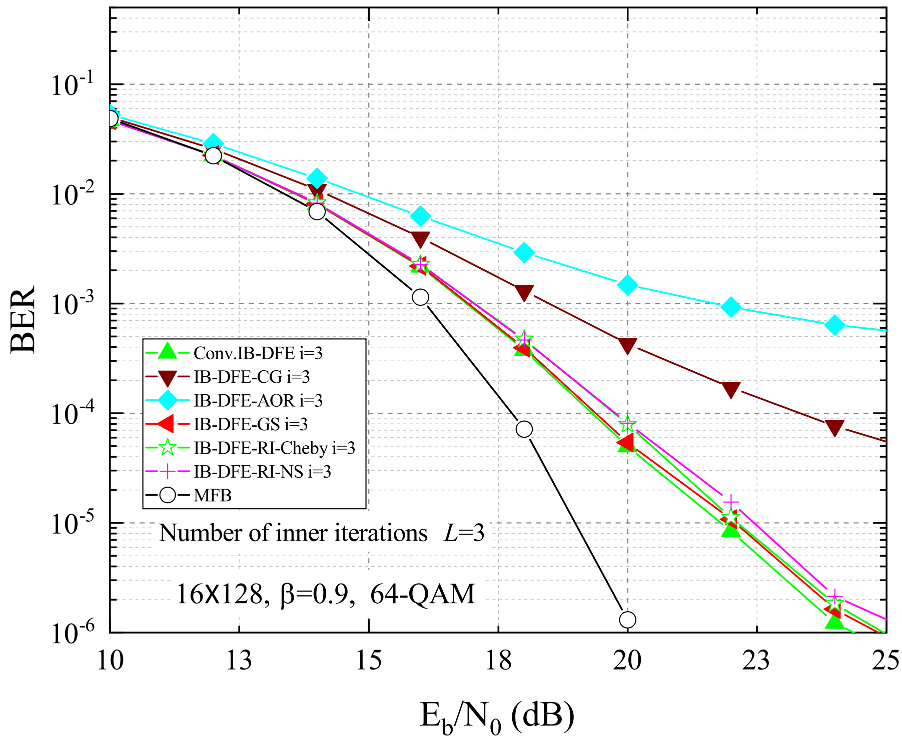


Figure 7.14: BER performance for 16×128 MIMO systems with a 64-QAM modulation and imperfect channel estimation.

where $\mathbf{E}_k$ is the estimation error, and β represents the reliability of the estimate [30]. The value of β lies between 0 and 1, where $\beta = 1$ represents perfect CSI and $\beta = 0$ represents no CSI available at the receiver. The estimation error matrix $\mathbf{E}_k$ can be assumed to be independent of $\mathbf{H}_k$.

Fig. 7.14 shows the BER performance of different IB-DFE receivers with imperfect CSI, considering the antenna configuration 16×128, $\beta = 0.9$, $L = 3$, $i = 3$ and a 64-QAM modulation scheme. Although the performance of all receivers degrades in the presence of imperfect CSI compared with the perfect CSI (see Fig.7.7), the performance of IB-DFE-RI-Cheby, IB-DFE-RI-NS, and IB-DFE-GS are almost close to the conventional IB-DFE.

In summary, the RI has a relatively slow convergence for IB-DFE, but the Chebyshev acceleration scheme significantly improves the convergence rates. As a result, the RI-Cheby method achieves almost the same performance as conventional IB-DFE, even for large constellations such as 64-QAM.

7.5 Conclusions

This chapter considers a massive MIMO system employing SC-FDE modulations with IB-DFE receivers. A novel structure of the IB-DFE receiver is introduced, compatible with conventional iterative methods such as RI, CG, GS, SOR, etc., thus avoiding matrix inversion. A Chebyshev accelerated Richardson iterative method is introduced to replace the matrix inversions in the proposed IB-DFE receiver. Other iterative methods, such as GS, SOR, CG, etc., are also incorporated in the proposed IB-DFE receiver for performance analysis. The complexity analysis shows that the computational complexity of the proposed receiver is much lower than that of the conventional IB-DFE receiver. The simulation results demonstrate that the BER performance of the proposed receiver is close to that of the exact receiver. Furthermore, a performance analysis is presented under imperfect channel state information.

Chapter

8 Conclusions and Future Work

8.1 Conclusions

This thesis primarily focused on developing and analyzing improved detector/receiver techniques for wireless communication systems operating on the uplink massive MIMO system with lower complexity. The main contributions and conclusions of this thesis are summarized below:

An efficient iterative massive MIMO detection algorithm based on the Modified Accelerated Over-Relaxation (MAOR) method was proposed. The speed of convergence for the proposed MAOR-based iterative detection is controlled by two key parameters: acceleration and relaxation. The optimum values for these parameters were derived. Additionally, an eigenvalue-based initial solution was applied instead of the common diagonal matrix approximation, leading to an improved convergence rate. For the acceleration of the MAOR algorithm, the Chebyshev polynomial acceleration method was employed, allowing near-MMSE (Minimum Mean Square Error) performance with MAOR. Error analysis demonstrated that Chebyshev acceleration has a relatively simple structure. Performance results showed that the Chebyshev-accelerated MAOR provides near-MMSE performance with reduced complexity, outperforming state-of-the-art massive MIMO detectors under similar channel conditions.

Low-complexity detection for massive MIMO schemes was considered, and an acceleration technique based on the Chebyshev method was proposed to speed up existing iterative methods. It was shown that this technique can be applied to various methods, outperforming state-of-the-art methods in terms of error rate performance and computational complexity.

Two iterative massive MIMO detectors, namely AC-AOR (Accelerated Chebyshev-based Accelerated Over-Relaxation) and AC-SOR (Accelerated Chebyshev-based Successive Over-Relaxation), were presented, formulated using the accelerated Chebyshev method to enhance the performance of conventional AOR and SOR methods, respectively. Moreover, Deep Unfolding Networks (DUNs) were proposed for the AC-SOR and AC-AOR methods to optimize the algorithm's parameters through training. The proposed DUNs are simple and fast to train, requiring only a few adjustable parameters. Due to the offline training of the proposed methods, the number of trainable variables in the AC-SORNet and AC-AORNet is independent of the number of antennas and is determined only by the number of layers, leading to a speedy and stable training process and reasonable scalability for larger systems. The computational complexity of the AC-AORNet and AC-SORNet is the same as that of AC-AOR and AC-SOR, respectively. Performance results showed that the AC-SORNet and AC-AORNet detectors outperform the AC-SOR and AC-AOR detectors in all settings, indicating that DUN can enhance the AC-SOR and AC-AOR detectors by learning the optimal parameters. Furthermore, experiments were conducted in various channel scenarios, including spatially correlated channel scenarios and realistic propagation environment models, demonstrating that the proposed method outperforms existing detectors.

A massive MIMO system employing Single-Carrier Frequency-Domain Equalization (SC-FDE) modulations with Iterative Block Decision Feedback Equalizer (IB-DFE) receivers was considered. A novel structure of the IB-DFE receiver was introduced, compatible with conventional iterative methods such as Richardson Iteration (RI), Conjugate Gradient (CG), Gauss-Seidel (GS), Successive Over-Relaxation (SOR), etc., thus avoiding matrix inversion. A Chebyshev-accelerated Richardson iterative method was introduced to replace the matrix inversions in the proposed IB-DFE receiver. Other iterative methods, such as GS, SOR, and CG, were also incorporated into the proposed IB-DFE receiver for performance analysis. The complexity analysis showed that the computational complexity of the proposed receiver is much lower than that of the conventional IB-DFE receiver. The simulation results demonstrated that the Bit Error Rate (BER) performance of the proposed receiver is close to that of the exact receiver. Furthermore, a performance analysis was presented under imperfect channel state information, showing that the proposed receiver maintains robust performance.

8.2 Future studies

The work presented in this thesis opens up avenues for further exploration and improvement in wireless communication systems. Some potential directions for future research include:

- Extending the proposed low-complexity detection method to address specific challenges in diverse wireless communication scenarios, such as those encountered in 6G networks.
- Developing low-complexity designs for extra-large MIMO systems in the near-field with spatial non-stationarities. This includes introducing a notion of visibility region (VR) and proposing a VR detection algorithm. The proposed method can then exploit the acquired VR information to design a low-complexity detection scheme for XL-MIMO systems.
- Exploring the integration of the proposed complexity-reduction techniques in receiver SC-FDE into emerging technologies like millimeter-wave and terahertz communication.
- Extending the work to implementation with FPGA cards, as they offer faster processing and lower computational complexity compared to CPUs.

These suggestions represent just a glimpse into the rich landscape of possibilities for future research. Addressing these challenges and expanding on the current work will contribute to the ongoing evolution and improvement of wireless communication systems.

Bibliography

[1] L. J. Vora. “Evolution of mobile generation technology: 1G to 5G and review of upcoming wireless technology 5G”. In: *International journal of modern trends in engineering and research* 2.10 (2015), pp. 281–290 (cit. on p. 9).

[2] E.-K. Hong et al. “6G R&D vision: Requirements and candidate technologies”. In: *Journal of Communications and Networks* 24.2 (2022), pp. 232–245 (cit. on p. 9).

[3] C. Ranaweera et al. “4G to 6G: disruptions and drivers for optical access”. In: *Journal of Optical Communications and Networking* 14.2 (2022), A143–A153 (cit. on p. 10).

[4] C. R. Shah. “Performance and comparative analysis of SISO, SIMO, MISO, MIMO”. In: *International journal of wireless communication and simulation* 9.1 (2017), pp. 1–14 (cit. on p. 10).

[5] A. K. Sarangi and A. Datta. “Capacity comparison of SISO, SIMO, MISO & MIMO systems”. In: *2018 Second International Conference on Computing Methodologies and Communication (ICCMC)*. IEEE. 2018, pp. 798–801 (cit. on p. 11).

[6] J. R. Hampton. *Introduction to MIMO communications*. Cambridge university press, 2013 (cit. on p. 12).

[7] E. Khorov et al. “A tutorial on IEEE 802.11 ax high efficiency WLANs”. In: *IEEE Communications Surveys & Tutorials* 21.1 (2018), pp. 197–216 (cit. on p. 12).

[8] U.-K. Kwon, G.-H. Im, and J.-B. Lim. "MIMO spatial multiplexing technique with transmit diversity". In: *IEEE Signal Processing Letters* 16.7 (2009), pp. 620–623 (cit. on p. 12).

[9] L. Zheng and D. N. C. Tse. "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels". In: *IEEE Transactions on information theory* 49.5 (2003), pp. 1073–1096 (cit. on p. 13).

[10] R. W. Heath and A. J. Paulraj. "Switching between diversity and multiplexing in MIMO systems". In: *IEEE transactions on Communications* 53.6 (2005), pp. 962–968 (cit. on pp. 13, 15).

[11] G. Caire, P. Elia, and K. R. Kumar. "Space-time coding: An overview". In: *Journal of Communications Software and Systems* 2.3 (2006), pp. 212–227 (cit. on p. 13).

[12] S. Rohilla, D. K. Patidar, and N. K. Soni. "Comparative analysis of maximum ratio combining and equal gain combining diversity techniques for WCDMA: a survey". In: *International Journal of Engineering Inventions* 3.1 (2013), pp. 72–77 (cit. on p. 13).

[13] D. Brennan. "Linear diversity combining techniques". In: *Proceedings of the IEEE* 91.2 (2003), pp. 331–356 (cit. on p. 13).

[14] D. Tse and P. Viswanath. *Fundamentals of wireless communication*. Cambridge university press, 2005 (cit. on p. 13).

[15] S. M. Alamouti. "A simple transmit diversity technique for wireless communications". In: *IEEE Journal on selected areas in communications* 16.8 (1998), pp. 1451–1458 (cit. on p. 14).

[16] J. G. Proakis and M. Salehi. *Digital communications*. McGraw-hill, 2008 (cit. on p. 14).

[17] G. J. Foschini et al. "Simplified processing for high spectral efficiency wireless communication employing multi-element arrays". In: *IEEE Journal on Selected areas in communications* 17.11 (1999), pp. 1841–1852 (cit. on p. 15).

[18] E. Telatar. "Capacity of multi-antenna Gaussian channels". In: *European transactions on telecommunications* 10.6 (1999), pp. 585–595 (cit. on p. 15).

[19] G. J. Foschini. "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas". In: *Bell labs technical journal* 1.2 (1996), pp. 41–59 (cit. on p. 15).

[20] Z. Wang and G. B. Giannakis. "Wireless multicarrier communications where fourier meets shannon, department of ece". In: *University of Minnesota, Minneapolis MN* (2000), pp. 1–21 (cit. on p. 16).

[21] D. Falconer et al. "Frequency domain equalization for single-carrier broadband wireless systems". In: *IEEE Communications Magazine* 40.4 (2002), pp. 58–66 (cit. on p. 16).

[22] J. Coon et al. "A comparison of MIMO-OFDM and MIMO-SCFDE in wlan environments". In: *GLOBECOM'03. IEEE Global Telecommunications Conference (IEEE Cat. No. 03CH37489)*. Vol. 6. IEEE. 2003, pp. 3296–3301 (cit. on p. 16).

[23] E. Biglieri et al. *MIMO wireless communications*. Cambridge university press, 2007 (cit. on p. 16).

[24] T. L. Marzetta. "Massive MIMO: An introduction". In: *Bell Labs Technical Journal* 20 (2015), pp. 11–22 (cit. on pp. 16, 19).

[25] Q. H. Spencer et al. "An introduction to the multi-user MIMO downlink". In: *IEEE communications Magazine* 42.10 (2004), pp. 60–67 (cit. on p. 16).

[26] J. Kim, W. Choi, and H. Park. "Beamforming for full-duplex multiuser MIMO systems". In: *IEEE Transactions on Vehicular Technology* 66.3 (2016), pp. 2423–2432 (cit. on p. 16).

[27] T. L. Marzetta. "Noncooperative cellular wireless with unlimited numbers of base station antennas". In: *IEEE transactions on wireless communications* 9.11 (2010), pp. 3590–3600 (cit. on pp. 16, 20, 22).

[28] K. Zheng et al. "Survey of large-scale MIMO systems". In: *IEEE Communications Surveys & Tutorials* 17.3 (2015), pp. 1738–1760 (cit. on p. 16).

[29] E. G. Larsson et al. "Massive MIMO for next generation wireless systems". In: *IEEE communications magazine* 52.2 (2014), pp. 186–195 (cit. on pp. 16, 20).

[30] F. Rusek et al. "Scaling up MIMO: Opportunities and challenges with very large arrays". In: *IEEE signal processing magazine* 30.1 (2012), pp. 40–60 (cit. on pp. 16, 19, 28, 34, 36, 59, 65, 121–123, 140).

[31] Ericsson. *Mobile Traffic Forecast - Ericsson Mobility Report*. Accessed: 2024-12-11. 2024. URL: `https://www.ericsson.com/en/reports-and-papers/mobility-report/dataforecasts/mobile-traffic-forecast` (cit. on pp. 17, 18).

[32] L. Lu et al. "An overview of massive MIMO: Benefits and challenges". In: *IEEE journal of selected topics in signal processing* 8.5 (2014), pp. 742–758 (cit. on pp. 18, 19, 34, 59, 91, 92).

[33] J. Navarro-Ortiz et al. "A survey on 5G usage scenarios and traffic models". In: *IEEE Communications Surveys & Tutorials* 22.2 (2020), pp. 905–929 (cit. on p. 18).

[34] H. Tataria et al. "6G wireless systems: Vision, requirements, challenges, insights, and opportunities". In: *Proceedings of the IEEE* 109.7 (2021), pp. 1166–1199 (cit. on pp. 18, 19).

[35] M. H. Alsharif and R. Nordin. "Evolution towards fifth generation (5G) wireless networks: Current trends and challenges in the deployment of millimetre wave, massive MIMO, and small cells". In: *Telecommunication Systems* 64 (2017), pp. 617–637 (cit. on p. 19).

[36] Y. Niu et al. "A survey of millimeter wave communications (mmWave) for 5G: Opportunities and challenges". In: *Wireless networks* 21 (2015), pp. 2657–2676 (cit. on p. 19).

[37] E. Björnson, E. G. Larsson, and T. L. Marzetta. "Massive MIMO: Ten myths and one critical question". In: *IEEE Communications Magazine* 54.2 (2016), pp. 114–123 (cit. on p. 19).

[38] M. A. Areqi, A. T. Zahary, and M. N. Ali. "State-of-the-art device-to-device communication solutions". In: *IEEE Access* (2023) (cit. on p. 19).

[39] M. A. Adedoyin and O. E. Falowo. "Combination of ultra-dense networks and other 5G enabling technologies: A survey". In: *IEEE Access* 8 (2020), pp. 22893–22932 (cit. on p. 19).

[40] A. N. Uwaechia and N. M. Mahyuddin. "A comprehensive survey on millimeter wave communications for fifth-generation wireless networks: Feasibility and challenges". In: *IEEE Access* 8 (2020), pp. 62367–62414 (cit. on p. 19).

[41] A. A. Laghari et al. "A review and state of art of Internet of Things (IoT)". In: *Archives of Computational Methods in Engineering* (2021), pp. 1–19 (cit. on p. 19).

[42] F. A. P. de Figueiredo. "An Overview of Massive MIMO for 5G and 6G". In: *IEEE Latin America Transactions* 20.6 (2022), pp. 931–940 (cit. on p. 19).

[43] S. Elhoushy, M. Ibrahim, and W. Hamouda. "Cell-free massive MIMO: A survey". In: *IEEE Communications Surveys & Tutorials* 24.1 (2021), pp. 492–523 (cit. on p. 19).

[44] S. A. Busari et al. "Terahertz massive MIMO for beyond-5G wireless communication". In: *ICC 2019-2019 IEEE International Conference on Communications (ICC)*. IEEE. 2019, pp. 1–6 (cit. on p. 19).

[45] E. Nayebi et al. “Cell-free massive MIMO systems”. In: *2015 49th Asilomar Conference on Signals, Systems and Computers*. IEEE. 2015, pp. 695–699 (cit. on p. 19).

[46] S. K. Ibrahim et al. “Design, challenges and developments for 5G massive MIMO antenna systems at sub 6-GHz band: a review”. In: *Nanomaterials* 13.3 (2023), p. 520 (cit. on p. 19).

[47] E. Basar. “Index modulation techniques for 5G wireless networks”. In: *IEEE Communications Magazine* 54.7 (2016), pp. 168–175 (cit. on p. 20).

[48] X. Luo. “Multiuser massive MIMO performance with calibration errors”. In: *IEEE Transactions on Wireless Communications* 15.7 (2016), pp. 4521–4534 (cit. on p. 20).

[49] D. A. Basnayaka and H. Haas. “Spatial modulation for massive MIMO”. In: *2015 IEEE International Conference on Communications (ICC)*. IEEE. 2015, pp. 1945–1950 (cit. on p. 20).

[50] E. Björnson, J. Hoydis, and L. Sanguinetti. “Massive MIMO has unlimited capacity”. In: *IEEE Transactions on Wireless Communications* 17.1 (2017), pp. 574–590 (cit. on p. 20).

[51] E. Björnson, E. G. Larsson, and M. Debbah. “Massive MIMO for maximal spectral efficiency: How many users and pilots should be allocated?” In: *IEEE Transactions on Wireless Communications* 15.2 (2015), pp. 1293–1308 (cit. on p. 20).

[52] J. Chen et al. “Spectral-energy efficiency tradeoff in relay-aided massive MIMO cellular networks with pilot contamination”. In: *IEEE Access* 4 (2016), pp. 5234–5242 (cit. on p. 20).

[53] T. T. Do et al. “Jamming-resistant receivers for the massive MIMO uplink”. In: *IEEE Transactions on Information Forensics and Security* 13.1 (2017), pp. 210–223 (cit. on p. 20).

[54] X. Wu, N. C. Beaulieu, and D. Liu. “On favorable propagation in massive MIMO systems and different antenna configurations”. In: *IEEE Access* 5 (2017), pp. 5578–5593 (cit. on p. 20).

[55] U. Demirhan and A. Alkhateeb. “Cell-Free ISAC MIMO Systems: Joint Sensing and Communication Beamforming”. In: *arXiv preprint arXiv:2301.11328* (2023) (cit. on pp. 20, 21).

[56] L. You et al. “Spectral efficiency and energy efficiency tradeoff in massive MIMO downlink transmission with statistical CSIT”. In: *IEEE Transactions on Signal Processing* 68 (2020), pp. 2645–2659 (cit. on p. 22).

[57] L. Fan et al. “Uplink achievable rate for massive MIMO systems with low-resolution ADC”. In: *IEEE Communications Letters* 19.12 (2015), pp. 2186–2189 (cit. on p. 22).

[58] V. Arya and K. Appaiah. “Kalman filter based tracking for channel aging in massive MIMO systems”. In: *2018 International Conference on Signal Processing and Communications (SPCOM)*. IEEE. 2018, pp. 362–366 (cit. on p. 22).

[59] L. D. Nguyen et al. “Beamforming and power allocation for energy-efficient massive MIMO”. In: *2017 22nd International Conference on Digital Signal Processing (DSP)*. IEEE. 2017, pp. 1–5 (cit. on p. 22).

[60] M. Nguyen. “Massive MIMO: a survey of benefits and challenges”. In: *ICSES Trans. Comput. Hardw. Electr. Eng* 4 (2018), pp. 1–4 (cit. on p. 22).

[61] J. Hoydis et al. “Making smart use of excess antennas: Massive MIMO, small cells, and TDD”. In: *Bell Labs Technical Journal* 18.2 (2013), pp. 5–21 (cit. on p. 22).

[62] V. Jungnickel et al. “The role of small cells, coordinated multipoint, and massive MIMO in 5G”. In: *IEEE communications magazine* 52.5 (2014), pp. 44–51 (cit. on p. 22).

[63] P. Popovski et al. “Wireless access in ultra-reliable low-latency communication (URLLC)”. In: *IEEE Transactions on Communications* 67.8 (2019), pp. 5783–5801 (cit. on p. 22).

[64] J. Hoydis, S. Ten Brink, and M. Debbah. “Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?” In: *IEEE Journal on selected Areas in Communications* 31.2 (2013), pp. 160–171 (cit. on p. 22).

[65] C. Ma et al. “Massive MIMO Empowered Wireless Powered Sensor Networks: An Optimal Design With Statistical CSI”. In: *IEEE Wireless Communications Letters* 11.10 (2022), pp. 2105–2109. DOI: `10.1109/LWC.2022.3194002` (cit. on p. 21).

[66] R. Zhang et al. “Integrated Sensing and Communication with Massive MIMO: A Unified Tensor Approach for Channel and Target Parameter Estimation”. In: *IEEE Transactions on Wireless Communications* (2024) (cit. on p. 21).

[67] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta. “Energy and spectral efficiency of very large multiuser MIMO systems”. In: *IEEE Transactions on Communications* 61.4 (2013), pp. 1436–1449 (cit. on pp. 28, 59, 119, 121).

[68] B. E. Godana and T. Ekman. “Parametrization based limited feedback design for correlated MIMO channels using new statistical models”. In: *IEEE transactions on wireless communications* 12.10 (2013), pp. 5172–5184 (cit. on pp. 29, 112).

[69] J.-P. Kermoal et al. “A stochastic MIMO radio channel model with experimental validation”. In: *IEEE Journal on selected areas in Communications* 20.6 (2002), pp. 1211–1226 (cit. on p. 29).

[70] T. L. Marzetta. “How much training is required for multiuser MIMO?” In: *2006 fortieth asilomar conference on signals, systems and computers*. IEEE. 2006, pp. 359–363 (cit. on p. 29).

[71] A. Zaib et al. “Distributed channel estimation and pilot contamination analysis for massive MIMO-OFDM systems”. In: *IEEE Transactions on Communications* 64.11 (2016), pp. 4607–4621 (cit. on p. 29).

[72] H. Yin et al. “A coordinated approach to channel estimation in large-scale multiple-antenna systems”. In: *IEEE Journal on selected areas in communications* 31.2 (2013), pp. 264–273 (cit. on p. 29).

[73] Y. Liang et al. “Channel compensation for reciprocal TDD massive MIMO-OFDM with IQ imbalance”. In: *IEEE Wireless Communications Letters* 6.6 (2017), pp. 778–781 (cit. on p. 30).

[74] A.-A. Lu et al. “Low complexity polynomial expansion detector with deterministic equivalents of the moments of channel Gram matrix for massive MIMO uplink”. In: *IEEE Transactions on Communications* 64.2 (2015), pp. 586–600 (cit. on p. 30).

[75] X. Zhu and R. D. Murch. “Performance analysis of maximum likelihood detection in a MIMO antenna system”. In: *IEEE Transactions on Communications* 50.2 (2002), pp. 187–191 (cit. on p. 30).

[76] C. Studer and H. Bölcskei. “Soft–Input Soft–Output single Tree-Search Sphere decoding”. In: *IEEE Transactions on Information Theory* 56.10 (2010), pp. 4827–4842 (cit. on p. 30).

[77] P. W. Wolniansky et al. “V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel”. In: *1998 URSI international symposium on signals, systems, and electronics. Conference proceedings (Cat. No. 98EX167)*. IEEE. 1998, pp. 295–300 (cit. on p. 31).

[78] M.-S. Baek, Y.-H. You, and H.-K. Song. “Combined QRD-M and DFE detection technique for simple and efficient signal detection in MIMO-OFDM systems”. In: *IEEE Transactions on Wireless Communications* 8.4 (2009), pp. 1632–1638 (cit. on p. 31).

[79] B. Yin et al. “Implementation trade-offs for linear detection in large-scale MIMO systems”. In: *2013 IEEE international conference on acoustics, speech and signal processing*. IEEE. 2013, pp. 2679–2683 (cit. on p. 31).

[80] A. Benzin et al. “Low-complexity truncated polynomial expansion DL precoders and UL receivers for massive MIMO in correlated channels”. In: *IEEE Transactions on Wireless Communications* 18.2 (2019), pp. 1069–1084 (cit. on p. 31).

[81] C. Tang et al. “High precision low complexity matrix inversion based on Newton iteration for data detection in the massive MIMO”. In: *IEEE Communications Letters* 20.3 (2016), pp. 490–493 (cit. on pp. 31, 118).

[82] C. Zhang et al. “A low-complexity massive MIMO precoding algorithm based on Chebyshev iteration”. In: *IEEE Access* 5 (2017), pp. 22545–22551 (cit. on pp. 31, 39, 61, 118).

[83] A. Kammoun et al. “Linear precoding based on polynomial expansion: Large-scale multi-cell MIMO systems”. In: *IEEE Journal of Selected Topics in Signal Processing* 8.5 (2014), pp. 861–875 (cit. on p. 31).

[84] M. A. Albreem, M. Juntti, and S. Shahabuddin. “Massive MIMO detection techniques: A survey”. In: *IEEE Communications Surveys & Tutorials* 21.4 (2019), pp. 3109–3132 (cit. on pp. 31, 37, 81, 82, 85, 118).

[85] Y. Lee. “Decision-aided Jacobi iteration for signal detection in massive MIMO systems”. In: *Electronics Letters* 53.23 (2017), pp. 1552–1554 (cit. on pp. 31, 93).

[86] W.-S. Lee et al. “An efficient modified Gauss Seidel precoder for downlink massive MIMO systems”. In: *IEEE Access* 8 (2020), pp. 202164–202173 (cit. on p. 31).

[87] A. Yu et al. “Efficient successive over relaxation detectors for massive MIMO”. In: *IEEE Transactions on Circuits and Systems I: Regular Papers* 67.6 (2020), pp. 2128–2139 (cit. on pp. 31, 32, 37, 89, 92, 111, 118).

[88] S. Berra et al. “A Low-Complexity Soft-Output Signal Data Detection Algorithm for UL Massive MIMO Systems”. In: *2021 International Conference on Computer, Information and Telecommunication Systems (CITS)*. IEEE. 2021, pp. 1–6 (cit. on pp. 31, 93, 118, 128, 129).

[89] J. Tu et al. “An efficient massive MIMO detector based on second-order Richardson iteration: From algorithm to flexible architecture”. In: *IEEE Transactions on Circuits and Systems I: Regular Papers* 67.11 (2020), pp. 4015–4028 (cit. on pp. 31, 32, 118).

[90] L. Liu et al. “Energy and area efficient recursive Conjugate Gradient-based MMSE detector for massive MIMO systems”. In: *IEEE Transactions on Signal Processing* 68 (2020), pp. 573–588 (cit. on pp. 31, 32, 118, 128, 129).

[91] 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”. In: *IEEE Communications Letters* 20.2 (2015), pp. 276–279 (cit. on p. 31).

[92] S. Chakraborty, N. B. Sinha, and M. Mitra. “Likelihood ascent search-aided low complexity improved performance massive MIMO detection in perfect and imperfect channel state information”. In: *International Journal of Communication Systems* 35.8 (2022), e5113. DOI: https://doi.org/10.1002/dac.5113 (cit. on pp. 32, 123, 128, 129).

[93] S. Berra, M. A. Albreem, and M. S. Abed. “A low complexity linear precoding method for massive MIMO”. In: *2020 International Conference on UK-China Emerging Technologies (UCET)*. IEEE. 2020, pp. 1–4 (cit. on pp. 32, 89, 92, 93, 111, 118).

[94] J. Minango and C. De Almeida. “Optimum and quasi-optimum relaxation parameters for low-complexity massive MIMO detector based on Richardson method”. In: *Electronics Letters* 53.16 (2017), pp. 1114–1115 (cit. on pp. 32, 89, 122, 123).

[95] D. M. Young. *Iterative solution of large linear systems*. Elsevier, 2014 (cit. on pp. 32, 60, 67, 81, 83, 89, 92, 94–96, 118, 122, 123).

[96] S. Berra et al. “Efficient iterative massive MIMO detection using Chebyshev acceleration”. In: *Physical Communication* 52 (2022), p. 101651 (cit. on pp. 32, 84, 89, 97, 105, 118, 125).

[97] S. Berra, R. Dinis, and S. Shahabuddin. “Fast matrix inversion based on Chebyshev acceleration for linear detection in massive MIMO systems”. In: *Electronics Letters* 58.11 (2022), pp. 451–453 (cit. on pp. 32, 89, 118, 122).

[98] M. Shao, W.-K. Ma, and J. Liu. “An Explanation of Deep MIMO Detection from a Perspective of Homotopy Optimization”. In: *IEEE Open Journal of Signal Processing* (2023) (cit. on pp. 32, 89).

[99] A. Balatsoukas-Stimming and C. Studer. “Deep Unfolding for communications systems: A survey and some new directions”. In: *2019 IEEE International Workshop on Signal Processing Systems (SiPS)*. IEEE. 2019, pp. 266–271 (cit. on pp. 32, 89).

[100] M. Khani et al. “Adaptive neural signal detection for massive MIMO”. In: *IEEE Transactions on Wireless Communications* 19.8 (2020), pp. 5635–5648 (cit. on pp. 32, 40, 89).

[101] H. He et al. “A Model-Driven Deep Learning network for MIMO detection”. In: *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE. 2018, pp. 584–588 (cit. on pp. 32, 89).

[102] J. Liao et al. “A Model-Driven Deep Learning method for massive MIMO detection”. In: *IEEE Communications Letters* 24.8 (2020), pp. 1724–1728 (cit. on pp. 32, 90).

[103] S. Takabe et al. “Trainable projected gradient detector for massive overloaded MIMO channels: Data-driven tuning approach”. In: *IEEE Access* 7 (2019), pp. 93326–93338 (cit. on pp. 32, 90).

[104] H. Chen, G. Yao, and J. Hu. “Algorithm parameters selection method with Deep Learning for EP MIMO detector”. In: *IEEE Transactions on Vehicular Technology* 70.10 (2021), pp. 10146–10156 (cit. on pp. 32, 90).

[105] S. Li et al. “Expectation Propagation aided Model Driven Learning for OTFS Signal Detection”. In: *IEEE Transactions on Vehicular Technology* (2023) (cit. on pp. 32, 90).

[106] S. Shahabuddin, M. Juntti, and C. Studer. “ADMM-based infinity norm detection for large MU-MIMO: Algorithm and VLSI architecture”. In: *IEEE International Symposium on Circuits and Systems*. May 2017, pp. 1–4. DOI: 10.1109/ISCAS.2017.8050311 (cit. on p. 34).

[107] S. Shahabuddin et al. “ADMM-Based Infinity-Norm Detection for Massive MIMO: Algorithm and VLSI Architecture”. In: *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* (Feb. 2021), pp. 1–13. DOI: 10.1109/TVLSI.2021.3056946 (cit. on p. 34).

[108] V. A. Marchenko and L. A. Pastur. “Distribution of eigenvalues for some sets of random matrices”. In: *Matematicheskii Sbornik* 114.4 (1967), pp. 507–536 (cit. on p. 35).

[109] R. Couillet and M. Debbah. *Random matrix methods for wireless communications*. Cambridge University Press, 2011 (cit. on p. 35).

[110] X. Gao et al. “Low-complexity MMSE signal detection based on Richardson method for large-scale MIMO systems”. In: *2014 IEEE 80th Vehicular Technology Conference (VTC2014-Fall)*. IEEE. 2014, pp. 1–5 (cit. on pp. 36, 37, 60).

[111] Å. Björck. *Numerical methods for least squares problems*. SIAM, 1996 (cit. on pp. 36, 122).

[112] L. Dai et al. “Low-complexity soft-output signal detection based on Gauss–Seidel method for uplink multiuser large-scale MIMO systems”. In: *IEEE Transactions on Vehicular Technology* 64.10 (2014), pp. 4839–4845 (cit. on pp. 36, 82, 103, 111, 118, 128, 129).

[113] J. Minango, C. de Almeida, and C. D. Altamirano. “Low-complexity MMSE detector for massive MIMO systems based on Damped Jacobi method”. In: *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*. IEEE. 2017, pp. 1–5 (cit. on pp. 37, 118).

[114] Z. Wu et al. “Efficient architecture for soft-output massive MIMO detection with Gauss-Seidel method”. In: *2016 IEEE international symposium on circuits and systems (ISCAS)*. IEEE. 2016, pp. 1886–1889 (cit. on p. 37).

[115] J. Zeng, J. Lin, and Z. Wang. “An improved Gauss-Seidel algorithm and its efficient architecture for massive MIMO systems”. In: *IEEE Transactions on Circuits and Systems II: Express Briefs* 65.9 (2018), pp. 1194–1198 (cit. on pp. 37, 60).

[116] A. Yu et al. “Efficient SOR-based detection and architecture for large-scale MIMO uplink”. In: *2016 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS)*. IEEE. 2016, pp. 402–405 (cit. on p. 37).

[117] J. Tu et al. “An Efficient Massive MIMO Detector Based on Second-Order Richardson Iteration: From Algorithm to Flexible Architecture”. In: *IEEE Transactions on Circuits and Systems I: Regular Papers* 67.11 (2020), pp. 4015–4028. DOI: 10.1109/TCSI.2020.3010890 (cit. on p. 37).

[118] B. Yin et al. “A 3.8 Gb/s large-scale MIMO detector for 3GPP LTE-Advanced”. In: *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2014, pp. 3879–3883 (cit. on p. 38).

[119] M. Wu et al. “Large-scale MIMO detection for 3GPP LTE: Algorithms and FPGA implementations”. In: *IEEE Journal of Selected Topics in Signal Processing* 8.5 (2014), pp. 916–929 (cit. on pp. 38, 118).

[120] L. Shao and Y. Zu. “Joint Newton Iteration and Neumann Series Method of Convergence-Accelerating Matrix Inversion Approximation in Linear Precoding for Massive MIMO Systems”. In: *Mathematical Problems in Engineering* 2016.1 (2016), p. 1745808 (cit. on p. 38).

[121] B. Kang, J.-H. Yoon, and J. Park. "Low complexity massive MIMO detection architecture based on Neumann method". In: *2015 International SoC Design Conference (ISOCC)*. IEEE. 2015, pp. 293–294 (cit. on p. 38).

[122] M. A. Albreem. "Approximate matrix inversion methods for massive MIMO detectors". In: *2019 IEEE 23rd International Symposium on Consumer Technologies (ISCT)*. IEEE. 2019, pp. 87–92 (cit. on p. 38).

[123] B. Nagy, M. Elsabrouty, and S. Elramly. "Fast converging weighted neumann series precoding for massive MIMO systems". In: *IEEE Wireless Communications Letters* 7.2 (2017), pp. 154–157 (cit. on p. 38).

[124] F. Jin et al. "A low complexity signal detection scheme based on improved Newton iteration for massive MIMO systems". In: *IEEE Communications Letters* 23.4 (2019), pp. 748–751 (cit. on pp. 38, 118).

[125] J. Minango and C. de Almeida. "Low complexity zero forcing detector based on Newton-Schultz iterative algorithm for massive MIMO systems". In: *IEEE Transactions on Vehicular Technology* 67.12 (2018), pp. 11759–11766 (cit. on pp. 38, 118, 123).

[126] H. Wang et al. "Expectation propagation detector for extra-large scale massive MIMO". In: *IEEE Transactions on Wireless Communications* 19.3 (2020), pp. 2036–2051 (cit. on p. 39).

[127] O. Alamu, B. Iyaomolere, and A. Abdulrahman. "An overview of massive MIMO localization techniques in wireless cellular networks: Recent advances and outlook". In: *Ad Hoc Networks* 111 (2021), p. 102353 (cit. on p. 39).

[128] S. Mondal, K. N. Salama, and W. H. Ali. "A novel approach for K-Best MIMO detection and its VLSI implementation". In: *2008 IEEE International Symposium on Circuits and Systems*. IEEE. 2008, pp. 936–939 (cit. on p. 39).

[129] A. Dabah et al. "Performance/complexity trade-offs of the sphere decoder algorithm for massive MIMO systems". In: *arXiv preprint arXiv:2002.09561* (2020) (cit. on p. 39).

[130] S. Chakraborty, N. B. Sinha, and M. Mitra. "Iteration optimized layered tabu search for large scale MIMO detection". In: *2021 10th international conference on internet of everything, microwave engineering, communication and networks (IEMECON)*. IEEE. 2021, pp. 1–5 (cit. on p. 39).

[131] E. Ohlmer and G. Fettweis. "Linear and non-linear detection for MIMO-OFDM systems with linear precoding and spatial correlation". In: *2010 IEEE Wireless Communication and Networking Conference*. IEEE. 2010, pp. 1–6 (cit. on p. 39).

[132] G. Berardinelli et al. “Turbo receivers for single user MIMO LTE-A uplink”. In: *VTC Spring 2009-IEEE 69th Vehicular Technology Conference*. IEEE. 2009, pp. 1–5 (cit. on p. 39).

[133] A. Krizhevsky, I. Sutskever, and G. E. Hinton. “Imagenet classification with deep convolutional neural networks”. In: *Advances in neural information processing systems* 25 (2012) (cit. on p. 39).

[134] G. Hinton et al. “Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups”. In: *IEEE Signal processing magazine* 29.6 (2012), pp. 82–97 (cit. on p. 39).

[135] Z. Qin et al. “Deep learning in physical layer communications”. In: *IEEE Wireless Communications* 26.2 (2019), pp. 93–99 (cit. on p. 39).

[136] Y. Yang et al. “Deep learning-based downlink channel prediction for FDD massive MIMO system”. In: *IEEE Communications Letters* 23.11 (2019), pp. 1994–1998 (cit. on p. 39).

[137] H. Ye et al. “Deep learning-based end-to-end wireless communication systems with conditional GANs as unknown channels”. In: *IEEE Transactions on Wireless Communications* 19.5 (2020), pp. 3133–3143 (cit. on p. 39).

[138] T. O’shea and J. Hoydis. “An introduction to deep learning for the physical layer”. In: *IEEE Transactions on Cognitive Communications and Networking* 3.4 (2017), pp. 563–575 (cit. on p. 40).

[139] H. Ye, G. Y. Li, and B.-H. F. Juang. “Deep reinforcement learning based resource allocation for V2V communications”. In: *IEEE Transactions on Vehicular Technology* 68.4 (2019), pp. 3163–3173 (cit. on p. 40).

[140] L. Liang et al. “Deep-learning-based wireless resource allocation with application to vehicular networks”. In: *Proceedings of the IEEE* 108.2 (2019), pp. 341–356 (cit. on p. 40).

[141] H. Ye, G. Y. Li, and B.-H. Juang. “Power of deep learning for channel estimation and signal detection in OFDM systems”. In: *IEEE Wireless Communications Letters* 7.1 (2017), pp. 114–117 (cit. on p. 40).

[142] N. Farsad and A. Goldsmith. “Neural network detection of data sequences in communication systems”. In: *IEEE Transactions on Signal Processing* 66.21 (2018), pp. 5663–5678 (cit. on p. 40).

[143] N. Shlezinger, R. Fu, and Y. C. Eldar. “DeepSIC: Deep soft interference cancellation for multiuser MIMO detection”. In: *IEEE Transactions on Wireless Communications* 20.2 (2020), pp. 1349–1362 (cit. on p. 40).

[144] H. He et al. “Model-Driven Deep Learning for physical layer communications”. In: *IEEE Wireless Communications* 26.5 (2019), pp. 77–83 (cit. on p. 40).

[145] H. He et al. “Model-driven deep learning for MIMO detection”. In: *IEEE Transactions on Signal Processing* 68 (2020), pp. 1702–1715 (cit. on p. 40).

[146] M. Goutay, F. A. Aoudia, and J. Hoydis. “Deep hypernetwork-based MIMO detection”. In: *2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. IEEE. 2020, pp. 1–5 (cit. on p. 40).

[147] V. Monga, Y. Li, and Y. C. Eldar. “Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing”. In: *IEEE Signal Processing Magazine* 38.2 (2021), pp. 18–44 (cit. on p. 40).

[148] G. Gao, C. Dong, and K. Niu. “Sparsely connected neural network for massive MIMO detection”. In: *2018 IEEE 4th International Conference on Computer and Communications (ICCC)*. IEEE. 2018, pp. 397–402 (cit. on p. 41).

[149] Y. Yu, J. Wang, and L. Guo. “Multisegment mapping network for massive MIMO detection”. In: *International Journal of Antennas and Propagation* 2021 (2021), pp. 1–7 (cit. on p. 41).

[150] Y. LeCun, Y. Bengio, and G. Hinton. “Deep learning”. In: *nature* 521.7553 (2015), pp. 436–444 (cit. on p. 41).

[151] Q. Hu et al. “Understanding deep MIMO detection”. In: *IEEE Transactions on Wireless Communications* (2023) (cit. on p. 41).

[152] S. Kusaladharma and C. Tellambura. “Massive MIMO based underlay networks with power control”. In: *2016 IEEE International Conference on Communications (ICC)*. IEEE. 2016, pp. 1–6 (cit. on p. 41).

[153] V.-D. Nguyen et al. “An efficient precoder design for multiuser MIMO cognitive radio networks with interference constraints”. In: *IEEE Transactions on Vehicular Technology* 66.5 (2016), pp. 3991–4004 (cit. on p. 42).

[154] X. Zhang and Q. Zhu. “AI-enabled network-functions virtualization and software-defined architectures for customized statistical QoS over 6G massive MIMO mobile wireless networks”. In: *IEEE Network* 37.2 (2023), pp. 30–37 (cit. on p. 42).

[155] S. Fortunati et al. “Massive MIMO radar for target detection”. In: *IEEE Transactions on Signal Processing* 68 (2020), pp. 859–871 (cit. on p. 42).

[156] S. H. Younus et al. “Massive MIMO for indoor VLC systems”. In: *2020 22nd International Conference on Transparent Optical Networks (ICTON)*. IEEE. 2020, pp. 1–6 (cit. on p. 42).

[157] L. Kansal et al. “Performance analysis of massive MIMO-OFDM system incorporated with various transforms for image communication in 5G systems”. In: *Electronics* 11.4 (2022), p. 621 (cit. on pp. 42, 117).

[158] H. Yang et al. “Joint precoder and equalizer design for multi-user multi-cell MIMO VLC systems”. In: *IEEE Transactions on Vehicular Technology* 67.12 (2018), pp. 11354–11364 (cit. on p. 42).

[159] H. Li. “Principle of OFDM and multi-carrier modulations”. In: *Encyclopedia of Wireless Networks*. Springer, 2020, pp. 1093–1097 (cit. on p. 45).

[160] J. Li, Y. Du, and Y. Liu. “Comparison of Spectral Efficiency for OFDM and SC-FDE under IEEE 802.16 Scenario”. In: *11th IEEE Symposium on Computers and Communications (ISCC’06)*. IEEE. 2006, pp. 467–471 (cit. on p. 46).

[161] N. Soltani. “Comparison of single-carrier FDMA vs. OFDMA as 3gpp long-term evolution uplink”. In: *EE359 Project* (2009) (cit. on p. 46).

[162] S. Weinstein and P. Ebert. “Data transmission by frequency-division multiplexing using the discrete Fourier transform”. In: *IEEE transactions on Communication Technology* 19.5 (1971), pp. 628–634. DOI: `10.1109/TCOM.1971.1090705` (cit. on p. 46).

[163] H. S. Stone. “R66-50 an algorithm for the machine calculation of complex fourier series”. In: *IEEE Transactions on Electronic Computers* 4 (1966), pp. 680–681. DOI: `10.1109/PGEC.1966.264428` (cit. on p. 46).

[164] E.-M. Amhoud et al. “OFDM with Index Modulation in Orbital Angular Momentum Multiplexed Free Space Optical Links”. In: *2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring)*. IEEE. 2021, pp. 1–5 (cit. on p. 46).

[165] T. Hwang et al. “OFDM and its wireless applications: A survey”. In: *IEEE transactions on Vehicular Technology* 58.4 (2008), pp. 1673–1694 (cit. on p. 46).

[166] Z. Mokhtari, M. Sabbaghian, and R. Dinis. “Massive MIMO downlink based on single carrier frequency domain processing”. In: *IEEE Transactions on Communications* 66.3 (2016), pp. 1164–1175 (cit. on p. 47).

[167] “ORTHOGONAL FREQUENCY DIVISION MULTIPLEXING (OFDM)”. In: *Digital Communication For Practicing Engineers*. John Wiley & Sons, Ltd, 2019. Chap. 10, pp. 453–504. DOI: `10.1002/9781119598138.ch10` (cit. on p. 48).

[168] M. M. Nasralla et al. “A Comparative Performance Evaluation of the HEVC Standard with its Predecessor H.264/AVC for Medical videos over 4G and beyond Wireless Networks”. In: *2018 8th International Conference on Computer Science and Information Technology (CSIT)*. 2018, pp. 50–54. DOI: 10.1109/CSIT.2018.8486153 (cit. on p. 48).

[169] H. Wang et al. “Performance analysis of frequency domain equalization in SC-FDMA systems”. In: *2008 IEEE International Conference on Communications*. IEEE. 2008, pp. 4342–4347 (cit. on p. 48).

[170] J. Zhang et al. “Advances in cooperative single-carrier FDMA communications: Beyond LTE-advanced”. In: *IEEE Communications Surveys & Tutorials* 17.2 (2015), pp. 730–756 (cit. on p. 48).

[171] N. Benvenuto, S. Tomasin, et al. “Block iterative DFE for single carrier modulation”. In: *Electronics Letters* 38.19 (2002), pp. 1144–1145 (cit. on p. 50).

[172] R. Dinis et al. “Iterative layered space-time receivers for single-carrier transmission over severe time-dispersive channels”. In: *IEEE Communications Letters* 8.9 (2004), pp. 579–581 (cit. on pp. 50, 117).

[173] R. Dinis et al. “On the design of turbo equalizers for SC-FDE schemes with different error protections”. In: *2010 IEEE 72nd Vehicular Technology Conference-Fall*. IEEE. 2010, pp. 1–5 (cit. on pp. 51, 54).

[174] R. Dinis et al. “Iterative frequency-domain equalization for general constellations”. In: *2010 IEEE Sarnoff Symposium*. IEEE. 2010, pp. 1–5 (cit. on pp. 51, 53, 54, 136).

[175] N. Benvenuto et al. “Single carrier modulation with nonlinear frequency domain equalization: An idea whose time has come—Again”. In: *Proceedings of the IEEE* 98.1 (2009), pp. 69–96 (cit. on pp. 51, 117).

[176] P. Bento et al. “Low-complexity equalisers for offset constellations in massive MIMO schemes”. In: *IEEE Access* 7 (2019), pp. 94373–94390 (cit. on pp. 52, 128, 138).

[177] F. C. Ribeiro et al. “Receiver design for the uplink of base station cooperation systems employing SC-FDE modulations”. In: *EURASIP Journal on Wireless Communications and Networking* 2015.1 (2015), pp. 1–15 (cit. on pp. 53, 117).

[178] O. Shental and J. Hoydis. “"Machine LLRning": Learning to softly demodulate,” 2019 IEEE Globecom Workshops (GC Wkshps)”. In: *IEEE, Dec.* Vol. 9. 2019 (cit. on p. 54).

[179] R. Dinis, P. Carvalho, and D. Borges. “Low Complexity MRC and EGC Based Receivers For SC-FDE Modulation With Massive MIMO Schemes”. In: *IEEE Global Conf. on Signal and Information Processing-Global SIP*. Vol. 1. 1. 2016, pp. 1–4 (cit. on pp. 55, 56, 117, 119, 130).

[180] R. Dinis and P. Montezuma. “Iterative receiver based on the EGC for massive MIMO schemes using SC-FDE modulations”. In: *Electronics letters* 52.11 (2016), pp. 972–974 (cit. on pp. 56, 117, 119, 129).

[181] D. Borges et al. “Two low complexity MRC and EGC based receivers for SC-FDE modulations with massive MIMO schemes”. In: *Journal of Signal Processing Systems* 90.10 (2018), pp. 1357–1367 (cit. on pp. 56, 117).

[182] K. Kuchi and A. B. Ayyar. “Performance analysis of ML detection in MIMO systems with co-channel interference”. In: *IEEE communications letters* 15.8 (2011), pp. 786–788 (cit. on p. 59).

[183] L. G. Barbero and J. S. Thompson. “Fixing the complexity of the sphere decoder for MIMO detection”. In: *IEEE Transactions on Wireless communications* 7.6 (2008), pp. 2131–2142 (cit. on p. 59).

[184] T. Datta et al. “Random-restart reactive tabu search algorithm for detection in large-MIMO systems”. In: *IEEE Communications Letters* 14.12 (2010), pp. 1107–1109 (cit. on p. 59).

[185] W. Fukuda et al. “Low-complexity detection based on belief propagation in a massive MIMO system”. In: *2013 IEEE 77th Vehicular Technology Conference (VTC Spring)*. IEEE. 2013, pp. 1–5 (cit. on p. 59).

[186] O. Gustafsson et al. “Approximate Neumann series or exact matrix inversion for massive MIMO?” In: *2017 IEEE 24th Symposium on Computer Arithmetic (ARITH)*. IEEE. 2017, pp. 62–63 (cit. on pp. 60, 71, 72).

[187] Q. Deng et al. “High-throughput signal detection based on fast matrix inversion updates for uplink massive multiuser multiple-input multi-output systems”. In: *IET Communications* 11.14 (2017), pp. 2228–2235 (cit. on p. 60).

[188] M. Mandloi and V. Bhatia. “Error recovery based low-complexity detection for uplink massive MIMO systems”. In: *IEEE Wireless Communications Letters* 6.3 (2017), pp. 302–305 (cit. on p. 60).

[189] B. Yin et al. “Conjugate gradient-based soft-output detection and precoding in massive MIMO systems”. In: *2014 IEEE Global Communications Conference*. IEEE. 2014, pp. 3696–3701 (cit. on p. 60).

[190] M. A. Albreem, M. Juntti, and S. Shahabuddin. "Massive MIMO Detection Techniques: A Survey". In: *IEEE Commun. Surveys Tuts.* 21.4 (2019), pp. 3109–3132 (cit. on p. 60).

[191] J. Zeng, J. Lin, and Z. Wang. "Low complexity message passing detection algorithm for large-scale MIMO systems". In: *IEEE Wireless Communications Letters* 7.5 (2018), pp. 708–711 (cit. on p. 60).

[192] C. Zhang et al. "A low-complexity massive MIMO precoding algorithm based on Chebyshev iteration". In: *IEEE Access* 5 (2017), pp. 22545–22551 (cit. on p. 60).

[193] Y. Man et al. "Massive MIMO pre-coding algorithm based on improved Newton iteration". In: *2017 IEEE 85th Vehicular Technology Conference (VTC Spring)*. IEEE. 2017, pp. 1–5 (cit. on p. 60).

[194] A. Yu et al. "Efficient SOR Based Massive MIMO Detection Using Chebyshev Acceleration". In: *2018 IEEE 23rd International Conference on Digital Signal Processing (DSP)*. IEEE. 2018, pp. 1–5 (cit. on pp. 60, 67, 71, 72).

[195] Q. Deng et al. "Accelerated widely-linear signal detection by polynomials for overloaded large-scale MIMO systems". In: *IEICE Transactions on Communications* (2017) (cit. on p. 60).

[196] A. Hadjidimos, A. Psimarni, and A. Yeyios. "On the convergence of the modified accelerated overrelaxation (MAOR) method". In: *Applied numerical mathematics* 10.2 (1992), pp. 115–127 (cit. on pp. 60–63, 67).

[197] S. Berra, M. A. M. Albreem, and M. S. Abed. "A Low Complexity Linear Precoding Method for Massive MIMO". In: *2020 International Conference on UK-China Emerging Technologies (UCET)*. 2020, pp. 1–4 (cit. on p. 61).

[198] H. Schneider. "Theorems on M-splittings of a singular M-matrix which depend on graph structure". In: *Linear Algebra and its Applications* 58 (1984), pp. 407–424 (cit. on p. 61).

[199] Y. Saad. *Iterative methods for sparse linear systems*. SIAM, 2003 (cit. on pp. 61, 62, 122).

[200] A. Hadjidimos, M. Lapidakis, and M. Tzoumas. "On Iterative Solution for Linear Complementarity Problem with an H_+-Matrix". In: *SIAM Journal on Matrix Analysis and Applications* 33.1 (2012), pp. 97–110 (cit. on pp. 62, 63).

[201] A. M. Tulino, S. Verdú, and S. Verdu. *Random matrix theory and wireless communications*. Now Publishers Inc, 2004 (cit. on pp. 63, 64, 84).

[202] A. Hadjidimos. "Accelerated overrelaxation method". In: *Mathematics of Computation* 32.141 (1978), pp. 149–157 (cit. on p. 63).

[203] G. Avdelas and A. Hadjidimos. “Optimum accelerated overrelaxation method in a special case”. In: *Mathematics of computation* 36.153 (1981), pp. 183–187 (cit. on pp. 63, 93).

[204] Å. Björck. *Numerical methods in matrix computations*. Vol. 59. Springer, 2015 (cit. on p. 64).

[205] X. Gao et al. “Capacity-approaching linear precoding with low-complexity for large-scale MIMO systems”. In: *2015 IEEE international conference on communications (ICC)*. IEEE. 2015, pp. 1577–1582 (cit. on p. 66).

[206] J. W. Demmel. *Applied numerical linear algebra*. SIAM, 1997 (cit. on pp. 67, 124, 125).

[207] R. S. Varga. *Matrix iterative analysis*. Vol. 27. Springer Science & Business Media, 1999 (cit. on pp. 67, 94).

[208] Q. Lin and W. Peng. “An Acceleration Method for Stationary Iterative Solution to Linear System of Equations”. In: *Advances in Applied Mathematics and Mechanics* 4.4 (2012), pp. 473–482 (cit. on p. 69).

[209] F. Jiang et al. “Stair matrix and its applications to massive MIMO uplink data detection”. In: *IEEE Transactions on Communications* 66.6 (2018), pp. 2437–2455 (cit. on pp. 81–83).

[210] H. S. Najafi and S. Edalatpanah. “Two stage mixed-type splitting iterative methods with applications”. In: *Journal of Taibah University for Science* 7.2 (2013), pp. 35–43. ISSN: 1658-3655. DOI: https://doi.org/10.1016/j.jtusci.2013.03.003. URL: https://www.sciencedirect.com/science/article/pii/S1658365513000149 (cit. on p. 92).

[211] K. C. Jea and D. M. Young. “On the effectiveness of adaptive Chebyshev acceleration for solving systems of linear equations”. In: *Journal of computational and applied mathematics* 24.1-2 (1988), pp. 33–54 (cit. on pp. 94, 96).

[212] H. Wang. “A chebyshev semi-iterative approach for accelerating projective and position-based dynamics”. In: *ACM Transactions on Graphics (TOG)* 34.6 (2015), pp. 1–9 (cit. on pp. 95, 96, 105).

[213] S. Takabe and T. Wadayama. “Convergence acceleration via Chebyshev step: Plausible interpretation of deepUunfolded Gradient Descent”. In: *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences* (2022), 2021EAP1139 (cit. on p. 95).

[214] G. H. Golub and C. F. Van Loan. *Matrix computations*. JHU press, 2013 (cit. on pp. 96, 118).

[215] D. P. Kingma and J. Ba. “ADAM: A method for stochastic optimization”. In: *arXiv preprint arXiv:1412.6980* (2014) (cit. on p. 100).

[216] X. Gao et al. “Matrix inversion-less signal detection using SOR method for uplink large-scale MIMO systems”. In: *2014 IEEE Global Communications Conference*. IEEE. 2014, pp. 3291–3295 (cit. on p. 105).

[217] P. Kyosti. “WINNER II channel models”. In: *IST, Tech. Rep. IST-4-027756 WINNER II D1. 1.2 V1. 2* (2007) (cit. on p. 114).

[218] M. Mounir et al. “A novel hybrid precoding-companding technique for peak-to-average power ratio reduction in 5G and beyond”. In: *Sensors* 21.4 (2021), p. 1410 (cit. on p. 117).

[219] M. Luzio, R. Dinis, and P. Montezuma. “SC-FDE for offset modulations: an efficient transmission technique for broadband wireless systems”. In: *IEEE Transactions on Communications* 60.7 (2012), pp. 1851–1861 (cit. on p. 117).

[220] K. Izadinasab, A. W. Shaban, and O. Damen. “Low-complexity detectors for uplink massive MIMO systems leveraging truncated polynomial expansion”. In: *IEEE Access* 10 (2022), pp. 91610–91621 (cit. on p. 118).

[221] S. Berra et al. “Deep Unfolding of Chebyshev Accelerated Iterative method for Massive MIMO Detection”. In: *IEEE Access* (2023) (cit. on p. 118).

[222] T. Xie et al. “Low-complexity SSOR-based precoding for massive MIMO systems”. In: *IEEE Communications Letters* 20.4 (2016), pp. 744–747 (cit. on p. 118).

[223] A. Bereyhi et al. “GLSE precoders for massive MIMO systems: Analysis and applications”. In: *IEEE Transactions on Wireless Communications* 18.9 (2019), pp. 4450–4465 (cit. on p. 118).

[224] M. A. Sedaghat, A. Bereyhi, and R. R. Müller. “Least square error precoders for massive MIMO with signal constraints: Fundamental limits”. In: *IEEE Transactions on Wireless Communications* 17.1 (2017), pp. 667–679 (cit. on p. 118).

[225] W. H. Press. *Numerical recipes 3rd edition: The art of scientific computing*. Cambridge university press, 2007 (cit. on p. 118).

[226] W. Zhang et al. “Widely linear precoding for large-scale MIMO with IQI: Algorithms and performance analysis”. In: *IEEE Transactions on Wireless Communications* 16.5 (2017), pp. 3298–3312 (cit. on p. 118).

[227] J. Jose et al. “Pilot contamination and precoding in multi-cell TDD systems”. In: *IEEE Transactions on Wireless Communications* 10.8 (2011), pp. 2640–2651 (cit. on p. 119).

[228] M. M. da Silva and R. Dinis. “A simplified massive MIMO implemented with pre or post-processing”. In: *Physical Communication* 25 (2017), pp. 355–362 (cit. on p. 119).

[229] A. Pereira et al. “Complexity analysis of FDE receivers for massive MIMO block transmission systems”. In: *IET Communications* 13.12 (2019), pp. 1762–1768 (cit. on pp. 119, 128).

[230] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta. “Aspects of favorable propagation in massive MIMO”. In: *2014 22nd European Signal Processing Conference (EUSIPCO)*. IEEE. 2014, pp. 76–80 (cit. on p. 121).

[231] G. Opfer and G. Schober. “Richardson’s iteration for nonsymmetric matrices”. In: *Linear algebra and its applications* 58 (1984), pp. 343–361 (cit. on p. 122).

[232] J. C. Mason and D. C. Handscomb. *Chebyshev polynomials*. Chapman and Hall/CRC, 2002 (cit. on p. 124).

[233] W. Hackbusch. *Iterative solution of large sparse systems of equations*. Vol. 95. Springer, 1994 (cit. on p. 125).

[234] R. Hunger. *Floating point operations in matrix-vector calculus*. Vol. 2019. Munich University of Technology, Inst. for Circuit Theory and Signal . . ., 2005 (cit. on p. 127).

[235] S. S. Haykin. *Adaptive filter theory*. Pearson Education India, 2008 (cit. on p. 170).

[236] J. Sherman. “Adjustment of an inverse matrix corresponding to changes in the elements of a given column or a given row of the original matrix”. In: *Annals of mathematical statistics* 20.4 (1949), p. 621 (cit. on p. 171).

APPENDIX

A Chapter 4

A.1 Chebyshev accleration method

To express the Chebyshev polynomial method to purposed method, we rewrite the equation in (4.3) of $\mathbf{y}^{(m)}$, which also satisfies a three-term-recurrence as

$$
\begin{aligned}
\mathbf{y}^{(m+1)} - \mathbf{x} &= p_m(\mathbf{B}_{MAOR})(\mathbf{x}^{(0)} - \mathbf{x}) \\
&= \mu_m T_m(\frac{\mathbf{B}_{MAOR}}{\rho})(\mathbf{x}^{(\mathbf{0})} - \mathbf{x}) \\
&= \mu_m \Big[2.\frac{\mathbf{B}_{MAOR}}{\rho}.T_m(\frac{\mathbf{B}_{MAOR}}{\rho})(\mathbf{x}^{(\mathbf{0})} - \mathbf{x}) \\
&- T_{m-1}(\frac{\mathbf{B}_{MAOR}}{\rho})(\mathbf{x}^{(\mathbf{0})} - \mathbf{x})\Big] \\
&= \mu_m \Big[2.\frac{\mathbf{B}_{MAOR}}{\rho}.\frac{p_m(\frac{\mathbf{B}_{MAOR}}{\rho})(\mathbf{x}^{(\mathbf{0})} - \mathbf{x})}{\mu_m} \\
&- \frac{p_{m-1}(\frac{\mathbf{B}_{MAOR}}{\rho}))(\mathbf{x}^{(\mathbf{0})} - \mathbf{x})}{\mu_{m-1}}\Big] \\
&= \mu_m \Big[2.\frac{\mathbf{B}_{MAOR}}{\rho}.\frac{\mathbf{y}^{(m)} - \mathbf{x}}{\mu_m} - \frac{\mathbf{y}^{(m-1)} - \mathbf{x}}{\mu_{m-1}}\Big] \\
&= 2\frac{\mu_m}{\rho\mu_{m+1}}\mathbf{B}_{MAOR}\mathbf{y}^{(m)} - \frac{\mu_{m-1}}{\mu_{m+1}}\mathbf{y}^{(m-1)} + \mathbf{d}_m.
\end{aligned}
\tag{B.1}
$$

From (4.42), the rest of properties can be defined as

$$\begin{aligned}
\mathbf{d}_m &= \mathbf{x} - 2\frac{\mu_m}{\mu_{m+1}}\frac{\mathbf{B}_{MAOR}}{\rho}\mathbf{x} - \frac{\mu_{m-1}}{\mu_{m+1}}\mathbf{x} \\
&= \mathbf{x} - 2\frac{\mu_m}{\mu_{m+1}}\frac{\mathbf{x} - \mathbf{c}}{\rho}\mathbf{x} - \frac{\mu_{m-1}}{\mu_{m+1}}\mathbf{x} \\
&= \mu_m\Big(\frac{1}{\mu_{m+1}} - \frac{2}{\rho\mu_m + \frac{1}{\mu_{m-1}}}\Big)\mathbf{x} + 2\frac{\mu_m}{\rho\mu_{m+1}}\mathbf{c} \\
&= 2\frac{\mu_m}{\rho\mu_{m+1}}\mathbf{c}.
\end{aligned} \tag{B.2}$$

Appendix

B

Chapter 7

B.1 Feedforward and feedback matrices

The feedforward and feedback matrices $\mathbf{F}_k^{(i)}$ and $\mathbf{B}_k^{(i)}$, respectively, are selected to minimize the MSE $\Theta_k^{(i)}$ given by[1]

$$\begin{aligned}\Theta &= \mathrm{Tr}\left\{\mathbb{E}\left[||\mathbf{FY}-\mathbf{B}\hat{\mathbf{X}}-\mathbf{X}||^2\right]\right\},\\ &= \mathrm{Tr}\left\{\mathbb{E}\Big[(\mathbf{FY}-\mathbf{B}\hat{\mathbf{X}}-\mathbf{X})(\mathbf{FY}-\mathbf{B}\hat{\mathbf{X}}-\mathbf{X})^H\Big]\right\}.\end{aligned} \tag{A.1}$$

The expression of Θ can further expressed as

$$\begin{aligned}\Theta =\ & \mathrm{Tr}(\mathbf{F}\mathbf{R}_{\mathbf{YY}^H}(\mathbf{F})^H) - \mathrm{Tr}(\mathbf{F}\mathbf{R}_{\mathbf{Y}\hat{\mathbf{X}}^H}(\mathbf{B})^H)\\ & - \mathrm{Tr}(\mathbf{F}\mathbf{R}_{\mathbf{YX}^H}) - \mathrm{Tr}(\mathbf{B}\mathbf{R}_{\hat{\mathbf{X}}\mathbf{Y}^H}(\mathbf{F})^H)\\ & + \mathrm{Tr}(\mathbf{B}\mathbf{R}_{\hat{\mathbf{X}}\hat{\mathbf{X}}^H}(\mathbf{B})^H) + \mathrm{Tr}(\mathbf{B}\mathbf{R}_{\hat{\mathbf{X}}\mathbf{X}^H})\\ & - \mathrm{Tr}(\mathbf{R}_{\mathbf{XY}^H}(\mathbf{F})^H + \mathbf{R}_{\mathbf{X}\hat{\mathbf{X}}^H}(\mathbf{B})^H) + \mathrm{Tr}(\mathbf{R}_{\mathbf{X}}),\end{aligned} \tag{A.2}$$

where the second equality uses the matrix definitions

$$\mathbf{R}_{\mathbf{XX}^H} = \mathbb{E}[\mathbf{XX}^H] = \mathbf{R}_{\mathbf{X}}, \tag{A.3}$$

$$\mathbf{R}_{\mathbf{YY}^H} = \mathbf{H}\mathbf{R}_{\mathbf{X}_k}\mathbf{H}^H + \mathbf{R}_W, \tag{A.4}$$

$$\begin{aligned}\mathbf{R}_{\mathbf{X}\hat{\mathbf{X}}^H} &= \mathbf{R}_{\hat{\mathbf{X}}\mathbf{X}^H} = \mathbf{R}_{\hat{\mathbf{X}}\hat{\mathbf{X}}^H} = \mathbb{E}[\mathbf{X}\hat{\mathbf{X}}^H]\\ &= \mathbb{E}[\hat{\mathbf{X}}\mathbf{X}^H] = \mathbb{E}[\hat{\mathbf{X}}\hat{\mathbf{X}}^H] = \Upsilon^2\mathbf{R}_{\mathbf{X}},\end{aligned} \tag{A.5}$$

$$\mathbf{R}_{\mathbf{XY}^H} = \mathbf{R}_{\mathbf{X}}\mathbf{H}^H, \tag{A.6}$$

[1] For simplicity of notation, in this appendix we omit the iteration index $'i'$ and sub-carrier $'k'$ dependence in the different matrices and vectors.

$$\mathbf{R}_{\mathbf{Y}\mathbf{X}^H} = \mathbf{H}\mathbf{R}_{\mathbf{X}}, \tag{A.7}$$

$$\mathbf{R}_{\hat{\mathbf{X}}\mathbf{Y}^H} = \boldsymbol{\Upsilon}^2\mathbf{R}_{\mathbf{X}}\mathbf{H}^H, \tag{A.8}$$

$$\mathbf{R}_{\mathbf{Y}\hat{\mathbf{X}}^H} = \mathbf{H}\mathbf{R}_{\mathbf{X}}\boldsymbol{\Upsilon}^{2H}, \tag{A.9}$$

where $\boldsymbol{\Upsilon}$ is defined as in (3.19). To minimize the MSE for normalized signals in the different data streams, we have the optimization problem (3.16), which can be solved using Lagrange multipliers [235] through the Lagrangian function

$$\begin{aligned}\mathcal{L}(\mathbf{F}, \mathbf{B}, \gamma) = \Theta \\ - \gamma\left(\frac{1}{N}\sum_{k=0}^{N-1}\mathbf{F}_k\mathbf{H}_k - \mathbf{I}_T\right),\end{aligned} \tag{A.10}$$

where γ denotes the Lagrange multiplier. The optimum $\mathbf{F}$ and $\mathbf{B}$ satisfy the following conditions:

$$\begin{aligned}\nabla_{\mathbf{F}}\mathcal{L}(\mathbf{F}, \mathbf{B}, \gamma) = 0,\\ \nabla_{\mathbf{B}}\mathcal{L}(\mathbf{F}, \mathbf{B}, \gamma) = 0,\\ \nabla_{\gamma}\mathcal{L}(\mathbf{F}, \mathbf{B}, \gamma) = 0.\end{aligned} \tag{A.11}$$

From the equations (A.1) and (A.11), we have

$$\begin{aligned}\nabla_{\mathbf{F}}\mathcal{L}(\mathbf{F}, \mathbf{B}, \gamma) = 2\mathbf{F}\mathbf{R}_{\mathbf{Y}\mathbf{Y}^H} - 2\mathbf{B}\mathbf{R}_{\hat{\mathbf{X}}\mathbf{Y}^H} \\ - 2\mathbf{R}_{\mathbf{X}\mathbf{Y}^H} - \gamma\mathbf{H}^H,\end{aligned} \tag{A.12}$$

$$\begin{aligned}\nabla_{\mathbf{B}}\mathcal{L}(\mathbf{F}, \mathbf{B}, \gamma) = -2\mathbf{F}\mathbf{R}_{\mathbf{Y}\hat{\mathbf{X}}^H} + 2\mathbf{B}\mathbf{R}_{\hat{\mathbf{X}}\hat{\mathbf{X}}^H} \\ + 2\mathbf{R}_{\mathbf{X}\hat{\mathbf{X}}^H},\end{aligned} \tag{A.13}$$

$$\nabla_{\gamma}\mathcal{L}(\mathbf{F}, \mathbf{B}, \gamma) = \frac{1}{N}\sum_{k=0}^{N-1}(\mathbf{F}_k\mathbf{H}_k) - \mathbf{I}_T, \tag{A.14}$$

where $\mathbf{R}_{\mathbf{Y}\mathbf{Y}^H} = (\mathbf{R}_{\mathbf{Y}\mathbf{Y}^H})^H$, $\mathbf{R}_{\mathbf{Y}\hat{\mathbf{X}}^H} = \mathbf{R}_{\hat{\mathbf{X}}^H\mathbf{Y}}$, and $\mathbf{R}_{\mathbf{Y}\mathbf{X}^H} = \mathbf{R}_{\mathbf{X}^H\mathbf{Y}}$. This leads to:

$$2\mathbf{F}\mathbf{R}_{\mathbf{Y}\mathbf{Y}^H} - 2\mathbf{B}\mathbf{R}_{\hat{\mathbf{X}}\mathbf{Y}^H} - 2\mathbf{R}_{\mathbf{X}\mathbf{Y}^H} - \gamma\mathbf{H}^H = 0, \tag{A.15}$$

$$-2\mathbf{F}\mathbf{R}_{\mathbf{Y}\hat{\mathbf{X}}^H} + 2\mathbf{B}\mathbf{R}_{\hat{\mathbf{X}}\hat{\mathbf{X}}^H} + 2\mathbf{R}_{\mathbf{X}\hat{\mathbf{X}}^H} = 0, \tag{A.16}$$

$$\frac{1}{N}\sum_{k=0}^{N-1}(\mathbf{F}_k\mathbf{H}_k) - \mathbf{I}_T = 0. \tag{A.17}$$

From the equation (A.16), (A.5), and (A.9), the optimal value of $\mathbf{B}$ can be rewritten as

$$\begin{aligned}\mathbf{B} &= (\mathbf{F}\mathbf{R}_{\mathbf{Y}\hat{\mathbf{X}}^H} - \mathbf{R}_{\mathbf{X}\hat{\mathbf{X}}^H})(\mathbf{R}_{\hat{\mathbf{X}}\hat{\mathbf{X}}^H})^{-1} \\ &= (\mathbf{F}\mathbf{H}\boldsymbol{\Upsilon}^{2H}\mathbf{R}_{\mathbf{X}} - \boldsymbol{\Upsilon}^2\mathbf{R}_{\mathbf{X}})(\boldsymbol{\Upsilon}^2\mathbf{R}_{\mathbf{X}})^{-1} \\ &= \mathbf{F}\mathbf{H}\boldsymbol{\Upsilon}^{2H}\mathbf{R}_{\mathbf{X}}(\boldsymbol{\Upsilon}^2\mathbf{R}_{\mathbf{X}})^{-1} - \boldsymbol{\Upsilon}^2\mathbf{R}_{\mathbf{X}}(\boldsymbol{\Upsilon}^2\mathbf{R}_{\mathbf{X}})^{-1} \\ &= \mathbf{F}\mathbf{H} - \mathbf{I}_T.\end{aligned} \tag{A.18}$$

From equation (A.15), and (A.18), the optimal value of $\mathbf{F}$ is

$$\mathbf{F} = \boldsymbol{\Gamma}\mathbf{H}^H(\mathbf{H}(\mathbf{I}_T - \boldsymbol{\Upsilon}^2)\mathbf{H}^H + \frac{\sigma_W^2}{\sigma_X^2}\mathbf{I}_R)^{-1}, \tag{A.19}$$

where $\boldsymbol{\Gamma}$ is selected so as to ensure that $\frac{1}{N}\sum_{k=0}^{N-1}(\mathbf{F}_k\mathbf{H}_k) = \mathbf{I}_T$. By using the matrix inversion lemma [236]

$$\mathbf{Z}(\mathbf{U} + \mathbf{V}\mathbf{Z})^{-1} = (\mathbf{I} + \mathbf{Z}\mathbf{U}^{-1}\mathbf{V})^{-1}\mathbf{Z}\mathbf{U}^{-1}, \tag{A.20}$$

with

$$\begin{aligned} \mathbf{U} &= \frac{\sigma_W^2}{\sigma_X^2}\mathbf{I}_R, \\ \mathbf{V} &= \mathbf{H}(\mathbf{I}_T - \boldsymbol{\Upsilon}^2), \\ \mathbf{Z} &= \mathbf{H}^H. \end{aligned} \tag{A.21}$$

The equation (A.19) takes the equivalent form

$$\mathbf{F} = \boldsymbol{\Gamma}\left(\frac{\sigma_W^2}{\sigma_X^2}\mathbf{I}_T + \mathbf{H}^H\mathbf{H}(\mathbf{I}_T - \boldsymbol{\Upsilon}^2)\right)^{-1}\mathbf{H}^H. \tag{A.22}$$

This latter formula for $\hat{\mathbf{F}}$ is particularly interesting for the case where $R > T$, i.e., the channel matrix $\mathbf{H}$ is "tall".